2023-06-13 Analysis of the water matrix data of Fort Devens Groundwater samples

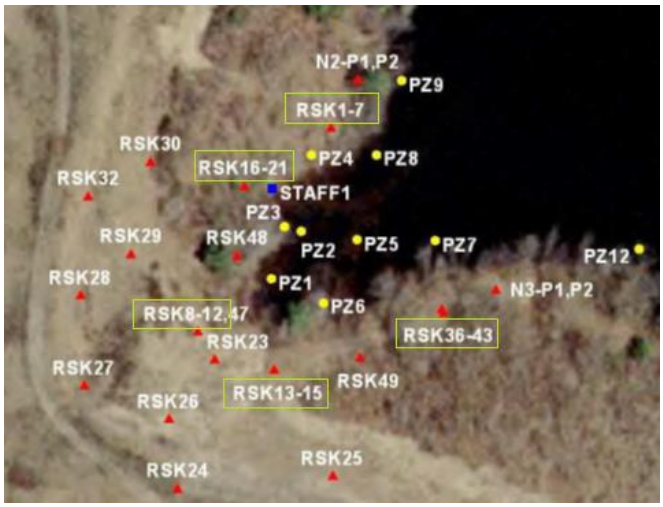

Figure 1. Map of sampled wells

1. Transcribe and organize data

- Sample analysis date: 2006-08-10 to 2021-11-2 from 26 wells, in five groups (RSK 1-7, RSK 8-12, RSK 13-15, RSK 16-19, RSK 36-43)
- Measured chemicals: As, Fe, Mn, Ca, K, Mg, Na, Cl^-, SO_4^{2-}, NH_3, Alkalinity (as $CaCO_3$), all in mg/L. TOC was not included partly based on the finding that it correlated to Fe and NH_3 positively.
- Incomplete sample analyses (missing NH_3, SO_4, or alkalinity) were excluded.

In total, results of 90 sample analyses were collected from 26 wells, some were analyzed more than the others. Most data were from 2006-2008 and 2018-2021.

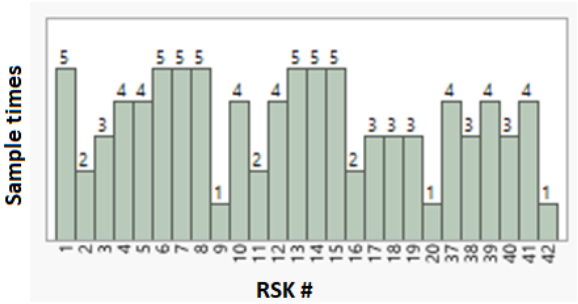

Figure 2. Sample times for the wells in Fort Devens

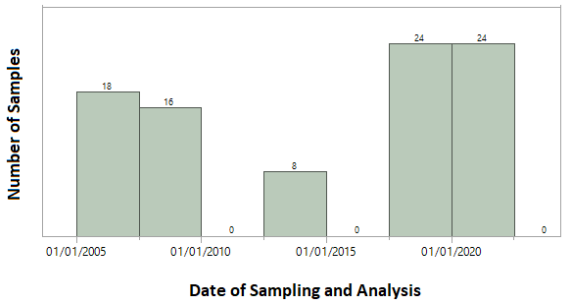

Figure 3. Sample times over the date

2. Correlations among chemicals in the groundwater

Table 1. Correlation among components in water matrix

Row	As	Fe	Mn	Ca	K	Mg	Na	Cl	Sulfate	NH3	Alkalinity
As	1.00	0.70	0.13	0.63	0.75	0.71	0.68	0.61	-0.07	0.61	0.75
Fe		1.00	0.08	0.40	0.68	0.38	0.48	0.47	-0.10	0.72	0.71
Mn			1.00	0.40	0.22	0.34	0.29	0.32	0.06	-0.04	0.28
Ca				1.00	0.75	**0.95**	**0.84**	**0.80**	-0.06	0.37	**0.80**
K					1.00	0.76	0.84	0.79	-0.09	**0.82**	**0.86**
Mg						1.00	0.82	0.75	-0.11	0.39	**0.81**
Na							1.00	**0.93**	0.11	0.51	0.74
Cl								1.00	0.01	0.50	0.68
Sulfate									1.00	-0.17	-0.20
NH3										1.00	0.66
Alkalinity											1.00

Summary:

- Moderate (0-3-0.5) to strong (> 0.5) positive correlations were found among 10 out of 11 chemicals, suggesting the underlying trend of water chemistry could be simpler.
- As correlated to most other inorganics except Mn and Sulfate.
- Sulfate was independent from other variables.
- There were no strong negative correlations.

3. Dimension reduction with principal component analysis (PCA)

Table 2. Eigenvalues and percentage of variances explained

Number	Eigenvalue	Percent	Cum Percent
1	**6.6239**	**60.217**	**60.217**
2	**1.4823**	**13.476**	**73.693**
3	**1.0048**	**9.135**	**82.827**
4	0.7237	6.579	89.406
5	0.4475	4.069	93.475
6	0.2771	2.519	95.994
7	0.2308	2.099	98.093
8	0.0933	0.848	98.941
9	0.0577	0.525	99.466
10	0.0339	0.308	99.774
11	0.0248	0.226	100

Table 3. Eigenvectors and components

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11
As	**0.323**	-0.158	0.091	0.003	0.579	-0.223	0.643	-0.068	0.141	-0.022	-0.197
Fe	0.273	**-0.394**	0.172	0.405	0.228	-0.390	-0.481	0.282	-0.200	0.022	0.147
Mn	0.122	**0.509**	-0.235	0.787	-0.097	0.009	0.194	0.010	-0.004	0.055	0.021
Ca	**0.340**	0.283	-0.148	-0.194	0.114	0.161	-0.287	0.497	0.182	0.097	-0.578
K	**0.367**	-0.119	0.070	0.014	-0.253	0.258	0.128	-0.078	-0.625	-0.489	-0.253
Mg	**0.340**	0.233	-0.185	-0.255	0.264	0.267	0.037	0.242	-0.121	-0.074	0.713
Na	**0.352**	0.195	0.156	-0.214	-0.216	-0.240	0.032	-0.253	-0.337	0.697	0.001
Cl	**0.336**	0.190	0.073	-0.168	-0.417	-0.537	-0.008	-0.033	0.396	-0.426	0.127
Sulfate	-0.037	**0.347**	**0.883**	0.082	0.134	0.230	-0.073	0.001	0.089	-0.083	0.037
NH3	**0.273**	**-0.459**	0.145	0.157	-0.430	0.385	0.241	0.248	0.365	0.261	0.121
Alkalinity	**0.357**	-0.083	-0.129	0.089	0.182	0.295	-0.391	-0.691	0.299	-0.031	-0.020

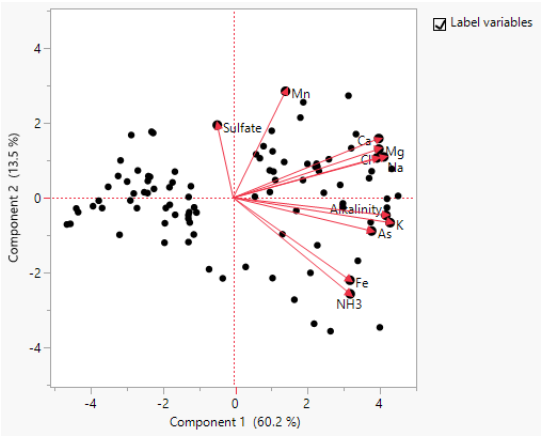

Figure 4. Biplot of the Principal component 1 and 2

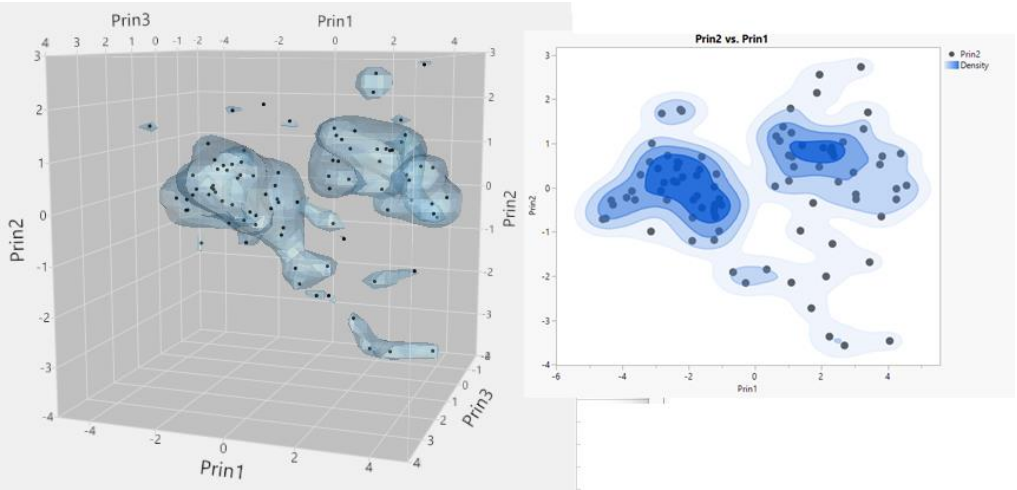

Figure 5 2-D and 3-D plots for principal components

Summary:

- PCA analysis showed that the variables could be reduced to 3 Principal Components. They accounted for 83% of the variance.
- Scattered plots suggested that the 90 samples could be grouped into at least 2 clusters.

4. K-mean clustering

The number of groups: Based on the 11 chemical characteristics, the samples were organized by K-mean clustering. The clustering was carried with group number n = 3 because increasing the number did not result in better organization as indicated by the value of cubic clustering criterion (n = 3, 4, 5, CCC = 0.429, -0.667, and 0.503)

The 90 samples were organized into three clusters in 2 iterations.

Group characteristics:

Table 4. Cluster means for three group of samples

Cluster	Count	As	Fe	Mn	Ca	K	Mg	Na	Cl	Sulfate	NH_3	Alkalinity
1	44	0.28	23.10	2.44	21.22	3.62	2.93	5.61	3.63	14.55	0.97	101.48
2	34	0.70	35.09	3.52	59.64	8.58	7.88	29.37	43.14	14.23	2.90	229.16
3	12	0.76	54.15	1.86	40.74	9.83	5.59	18.51	24.02	6.97	6.51	259.17

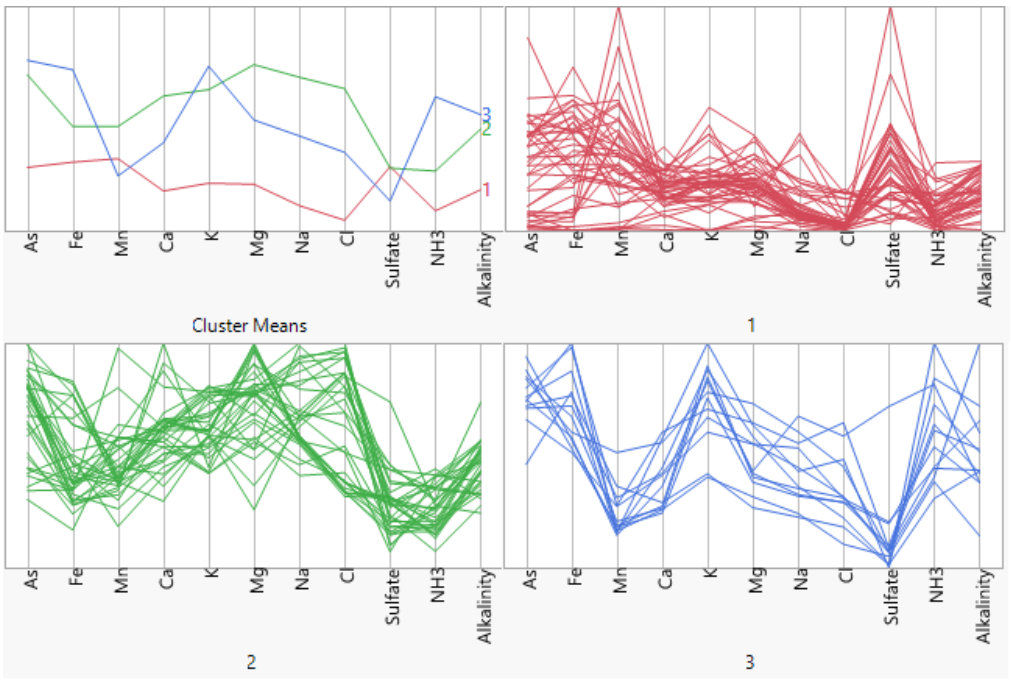

Figure 6. Parallel Coordinate Plots for clusters and their means

Comment:

- Most samples are in Cluster 1 (44) and 2 (34). On 18 samples are in C3.
- Sulfate and Mn did not play important roles in the clustering (similar cluster means, wide range within a cluster)
- Cluster 1 is highly different from Cluster 2 and 3, featuring low As, Fe, Ca, K, Mg, Na, NH_3, and alkalinity.
- Cluster 2 and 3 are both high in As, but C2 is lower in NH_3, Fe.

5. Clustering and Data Interpretation

a. Spatial and temporal variation of water chemistry

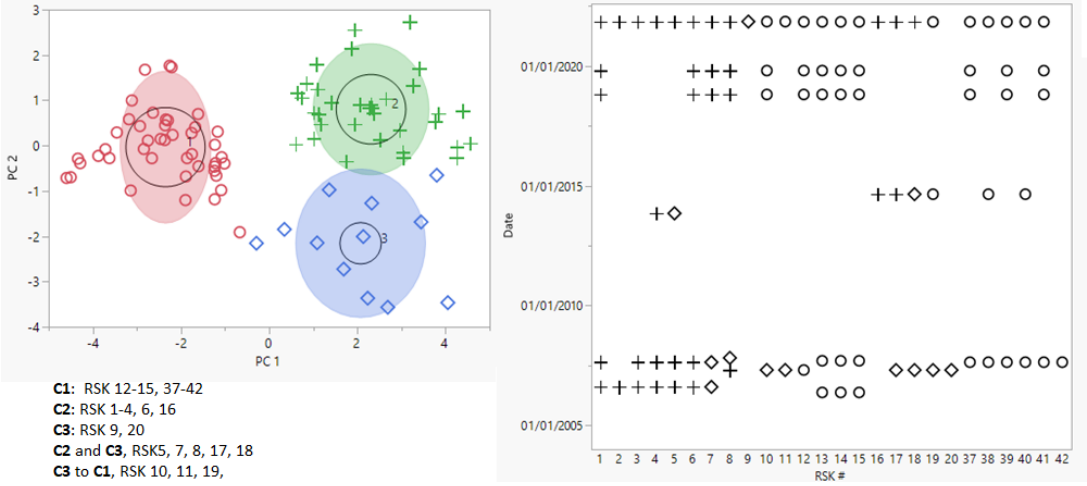

Figure 7. Biplot of the clusters and spatiotemporal variations

Figure 8, Locations of the clustered samples (Red C1, Green C2, Blue C2+C3, or C3 to C1)

Comment

- Wells with consistent C1 (low As, low Fe, Low NH_3) samples are in the south, including RSK12-15, and 37-42
- Wells with Consistent C2 (High As, medium Fe, medium NH_3) samples are in the north, including RSK1-6, and 16.
- Wells on the west generated samples that fell in C3 (High As, High Fe, High NH_3) or C2, and this variation were reversible over time (RSK 7, 8, 17, 18). Water in a few wells (RSK 10, 11, 19) changed from C3 to C1.

- No wells underwent the change from C1 to C3, C1 to C2, or C2 to C1, but some changed between C3 and C2. The observation is consistent with an underlying trend of C2 <=> C3=>C1.

b. Correlation between chemicals in different clusters

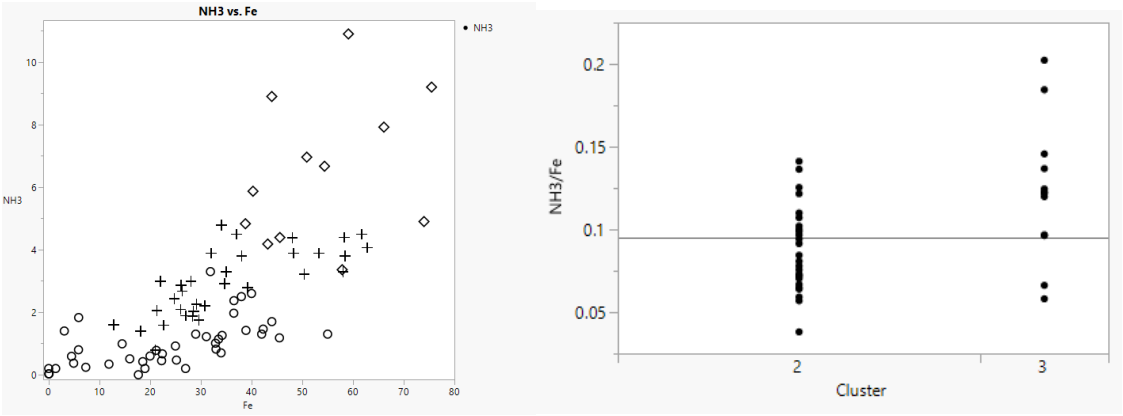

Figure 9 The concentration range of NH_3 and Fe and their correlation

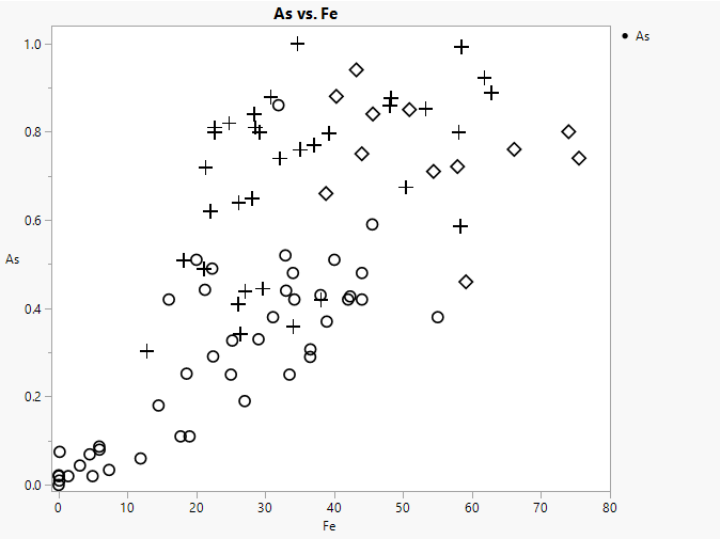

Figure 10. The correlation between As and Fe

Comment

- NH_3 and Fe are positively correlated. The concentrations of Fe and NH_3 were in the order of C3>C2>C1. The correlation between Fe and NH_3 variates in three clusters. The order of NH_3/Fe ratio was C3>C2>C1
- In C2 and C3, As concentration had no apparent correlation to Fe. In C1, it is very likely that As and Fe were positively correlated.

6. Select representative wells for sampling

Table 5. Wells that represent three clusters of the water matrices based on samples collected on Nov21

Chemicals	RSK12 C1	RSK13 C1	RSK15 C1	RSK40 C1	RSK41 C1	Mean C1	RSK 6 C2	Mean C2	RSK9 C2	Mean C3
As	0.11	**0.327**	0.252	0.442	0.486	**0.28**	0.852	**0.7**	0.721	**0.76**
Fe	17.7	**25.20**	18.57	21.23	22.28	**23.1**	53.2	**35.1**	57.87	**54.15**
Mn	1.5	**2.62**	1.56	3.83	2.22	**2.44**	3.4	**3.5**	2.75	**1.86**
Ca	20.3	**28.27**	20.74	17.60	17.57	**21.22**	58.4	**59.6**	31.52	**40.74**
K	3.7	**3.04**	2.71	3.75	3.456	**3.62**	10.7	**8.6**	6.08	**9.83**
Mg	2.4	**3.60**	2.88	2.68	2.839	**2.93**	8.1	**7.9**	3.49	**5.59**
Na	9.2	**3.44**	3.99	4.25	4.731	**5.61**	39.0	**29.4**	10.33	**18.51**
Cl	0.6	**1.45**	0.81	1.87	2.38	**3.63**	63.4	**43.1**	12.70	**24.02**
SO4	24.1	**16.30**	15.70	20.40	21.9	**14.55**	19.5	**14.2**	1.07	**6.97**
NH3	0.0	**0.47**	0.42	0.78	0.454613	**0.97**	3.9	**2.9**	3.36	**6.51**
Alk	90	**125**	94	85	82	**101.48**	241.0	**229.2**	218	**259.17**
Euclidian Distance	2.46	**0.47**	0.64	1.76	1.43	**0**	**3.9**	0	**6.23**	**0**

Three criteria are used to select the wells that represent each cluster:

1. The chemical composition of the latest samples fell into the cluster
2. The consistency of the water from the well
3. Minimal Euclidian Distance in Clustering

Table 6. Summary of the selected wells that represent the three clusters

Cluster	Well	Latest sample	Euclidian Distance	Consistency
1	RSK13	11/4/2021	0.47	2006-2021, 5 tests
2	RSK6	11/2/2021	3.9	2006-2021, 5 tests
3	RSK9	11/8/2021	6.23	1 test in 2021

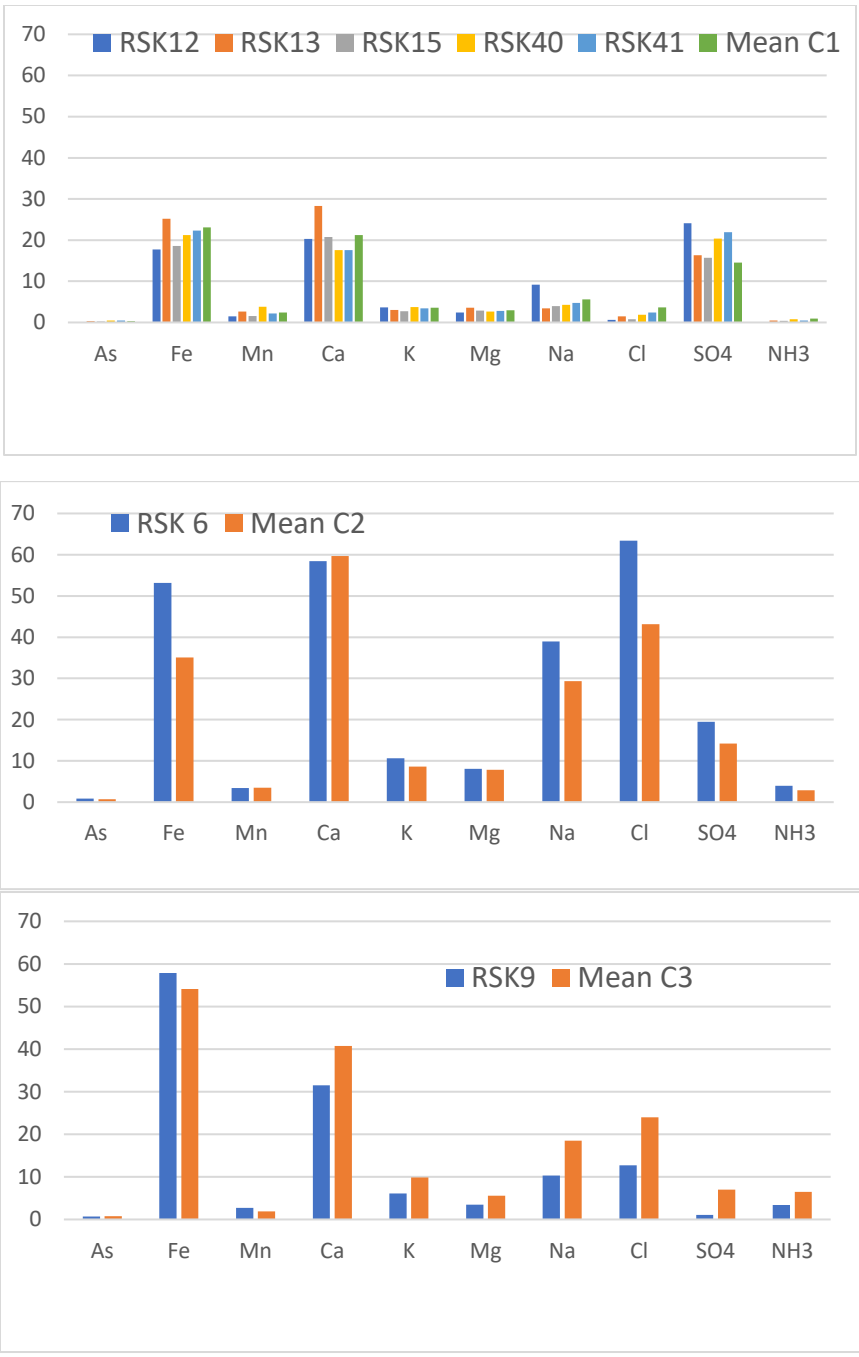

Figure 11. Profiles of chemical composition for samples and the average in a cluster (alkalinity excluded)

Recommended wells: RSK 13 for C1, RSK 6 for C2, and RSK 9 for C3

Table 7. Summary of distribution statistics for three clusters

Cluster		As	Fe	Mn	Ca	K	Mg	Na	Cl	Sulfate	NH3	Alkalinity
1 n = 44	Range	0-0.86	0.02-55	0.01-7.56	5.09-39	1.1-7.64	1-4.95	0.96-19	0.46-13	1.88-50	0-3.3	15-161
	Median	0.299	23.71	2.4	21.5	3.6	2.9	4.26	2.1	13.2	0.81	99
	Mean	0.282	23.1	2.43	21.2	3.62	2.92	5.61	3.62	14.5	0.97	101
2 n = 34	Range	0.304 -1	12.8-62.7	1.4-7.4	32-96	6.1-10.8	3.4-10.3	18-42	21.8-67.9	3.92-37	0.8-4.8	130-363
	Median	0.765	30.13	3.2	60.5	8.3	8.2	28.1	45	14.5	2.9	237
	Mean	0.695	35.09	3.5	59.6	8.6	7.9	29.4	43.1	14.2	2.9	229
3 n = 12	Range	0.46-0.94	38.8-75.5	1.1-3.88	27.2-60	5.9-13	3.49-7.8	10.3-29	7.52-44	0.37-36	3.4-10.9	82-487
	Median	0.755	52.7	1.5	40.3	10.3	5.56	17.4	21.4	4.01	6.27	239
	Mean	0.759	54.1	1.85	40.7	9.8	5.59	18.5	24	6.97	6.5	259