

Appendix A: Results of Principal Components Analysis

Following the transformation of the imputed geochemical data using a logcentred transform, the data were treated with a principal component analysis (PCA). Figure A.1 shows an ordered screeplot of the PCA, where it can be observed that the first eight principal components (PCs) account for a significant amount of the variability of the data. This is demonstrated quantitatively in Table A.1 where the first eight components account for more than 80 % of the variability of the data. Note, that the number of principal components is one less than the number of elements. This is due to the nature of the logcentred transform, the last of which is zero, and reflects the closure property of compositional data.

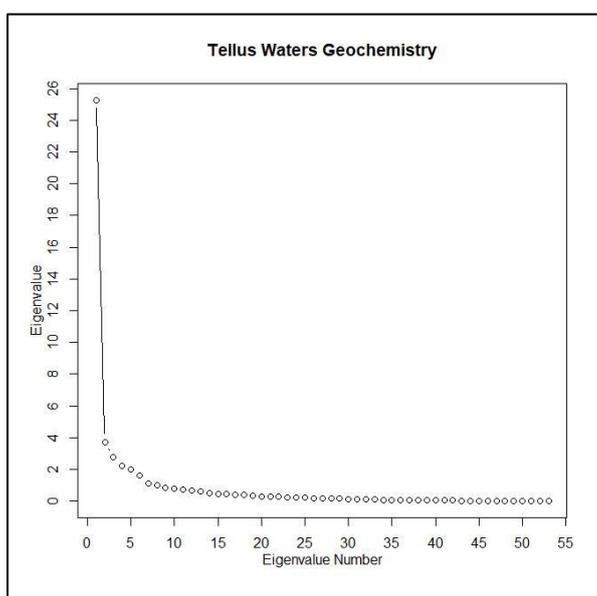


Figure A.1 Ordered screeplot of principal component eigenvalues

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
λ	25.29	3.73	2.77	2.24	2	1.61	1.13	1.02
$\lambda\%$	50.885	7.505	5.5734	4.507	4.024	3.239	2.274	2.052
$\Sigma\lambda\%$	50.885	58.390	63.964	68.471	72.495	75.734	78.008	80.060

Table A.1 Contribution of the first eight Principal Components to data variability. λ is the absolute contribution of each PC, $\lambda\%$ the percentage contribution and $\Sigma\lambda\%$ the cumulative % contribution of the PCs in ascending order.

The loadings (R-scores) of the variables are shown in Table A.2. The red/blue shading represents positive/negative relative relationships of the variables. In PC1 there is an inverse relationship between bicarbonate (HCO_3^-) and Al as demonstrated by the difference in colour. The shading of each cell indicates the relative significance of the element/anion across the entire range of elements and principal components. Generally, the first principal component shows the more intense shading, which decreases with increasing PC number.



	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
HCO3	-1.2985	-0.1796	0.0624	-0.1252	0.2067	-0.1476	0.0873	-0.0319
NPOC	0.2045	0.0448	0.2802	0.1742	0.0435	-0.2667	0.0247	-0.0161
SO4	-0.716	0.2537	-0.2024	0.1017	-0.0197	0.3931	-0.1206	-0.0977
NO3	-1.2259	0.2314	-0.8014	-0.4029	-0.6942	-0.4854	-0.0417	0.3731
Br	-0.2656	0.436	0.1665	-0.2271	0.0261	0.1086	0.0761	0.023
Ca	-1.3952	-0.2958	0.0226	-0.122	0.205	-0.1057	0.0518	-0.0643
Cl	-0.1673	0.5061	0.1264	-0.0898	0.0654	0.0924	0.0275	-0.0594
F	-0.9034	-0.1204	0.1004	0.0814	0.2555	-0.1115	0.0118	0.0144
K	-0.7457	0.0763	-0.146	-0.039	-0.1786	0.2281	0.2111	-0.11
Mg	-0.6534	0.2623	0.0637	-0.1392	0.0479	0.0642	0.0088	-0.0908
Na	-0.146	0.4868	0.1335	-0.0834	0.074	0.0823	0.0106	-0.0599
P	-0.312	-0.3433	0.0673	0.2125	-0.5124	-0.072	0.5046	-0.3511
S	-0.6668	0.2313	-0.1301	0.0933	-0.0191	0.2947	-0.0917	-0.0942
Al	1.0854	0.2724	-0.0189	0.1963	-0.0035	-0.1119	-0.0989	-0.0815
As	-0.1965	-0.1056	0.3291	0.0542	-0.1743	-0.1977	0.0202	0.0058
Ba	-0.6928	-0.0597	0.1162	-0.4649	-0.0241	-0.3142	-0.5248	-0.4638
Be	0.3802	0.0238	0.0227	0.0664	0.2054	-0.058	-0.1721	0.0391
Cd	-0.137	-0.1942	0.0393	0.4202	-0.0539	0.0513	-0.4265	0.2607
Ce	1.1686	-0.0691	-0.1836	-0.075	0.0756	0.0575	0.0823	0.0376
Co	0.1649	-0.3893	0.3242	-0.0809	-0.3165	0.2871	-0.0217	0.1403
Cr	0.0531	0.008	0.0332	0.0269	-0.0721	-0.0769	-0.0114	-0.115
Cs	0.1811	0.8463	0.1569	-0.0873	0.0954	0.132	0.1651	0.26
Cu	-0.232	-0.1902	-0.2323	0.2999	-0.2962	0.23	-0.0436	-0.0746
Dy	0.6879	-0.1708	-0.1706	-0.1166	0.0873	0.0174	-0.0418	-0.018
Er	0.5644	-0.1704	-0.151	-0.1524	0.1006	0.0051	-0.0085	0.0101
Eu	0.7219	-0.1272	-0.1803	-0.0764	0.0162	0.052	-0.0795	-0.0608
Fe	0.8364	-0.2372	0.5904	-0.2463	-0.017	-0.1978	0.068	0.0864
Gd	0.7756	-0.1705	-0.2024	-0.0966	0.061	0.0446	-0.0447	-0.0297
Ho	0.5544	-0.1496	-0.1476	-0.1222	0.0952	0.0073	-0.0312	-0.0012
La	0.9485	-0.059	-0.3027	-0.059	0.0781	0.0663	0.0812	0.0341
Li	-0.2963	0.3199	0.1355	-0.0644	0.0941	0.1515	-0.136	-0.0614
Mn	0.1305	-0.3741	0.6843	-0.5693	-0.2289	0.1824	0.0476	0.3051
Mo	-0.7749	-0.1591	0.0799	0.2704	0.1383	-0.0108	0.0141	0.096
Nb	0.1076	0.1262	-0.0472	0.1258	0.036	-0.0896	0.1785	-0.0666
Nd	0.9874	-0.1276	-0.2825	-0.0571	0.096	0.0383	-0.0084	0.0088
Ni	-0.4723	-0.5037	0.0721	0.2305	-0.2391	0.3202	-0.1295	0.019
Pb	0.9564	0.2403	0.2049	0.3514	-0.2302	-0.3128	-0.0056	0.0442
Pr	0.9752	-0.0911	-0.2813	-0.0571	0.0631	0.0717	0.0355	0.0016
Rb	-0.3248	0.2046	-0.0878	-0.0939	-0.0724	0.3016	0.1666	-0.0652
Sb	-0.5113	0.0067	-0.0521	0.3735	-0.1498	-0.032	-0.0631	-4.00E-04
Se	-0.3608	0.1208	0.0452	0.2958	0.0743	-0.123	-0.1073	0.0886
Si	-0.4378	0.1341	-0.0126	-0.2215	0.0079	0.0122	0.0075	-0.0053
Sm	0.8288	-0.1418	-0.2327	-0.0859	0.0558	0.0572	-0.041	-0.0441
Sn	0.0893	0.44	0.1498	-0.0153	-0.0156	-0.0713	-0.0126	0.0512
Sr	-1.1863	-0.1889	0.0574	-0.0363	0.2382	0.0014	0.0087	-0.0929
Th	0.6475	-0.0242	-0.1056	-0.0013	0.1317	0.0528	0.0696	-0.0139
Ti	0.8982	0.1593	0.0814	0.2178	-0.002	-0.255	0.1422	-0.1266
U	-1.2548	-0.2765	-0.1981	0.2343	0.5731	-0.1277	0.205	0.3412
V	0.1624	0.1352	0.0596	0.1797	0.0236	-0.2733	0.0691	-0.0945
Y	0.5208	-0.1972	-0.1499	-0.1131	0.1176	-0.0248	-0.0265	0.0293
Yb	0.4788	-0.1594	-0.1381	-0.2007	0.0999	0.0046	0.0158	0.0111
Zn	0.2738	0.1296	0.2498	0.3087	-0.2565	0.0933	-0.1941	0.1296
Zr	-0.0105	-0.4204	0.0019	0.0073	0.0875	-0.035	0.092	-0.0193

Table A.2 R-scores for first eight Principal Components

The relative contributions of the elements/anions of the PCs is shown in the Relative Contributions table (Table A.3). This table indicates the contribution of each element across all of the principal components. PC1 accounts for 79.8 % of the variability of HCO₃⁻. PC2 accounts for 63.8 % of the variability of Na.

The absolute contributions table is a measure of how much an element/anion contributes within each principal component. In the case of PC1, Ca and HCO₃⁻ have the



highest contributions (7.7 % & 6.7 % respectively). PC2 shows that 19.2 % of the variability is accounted for by Cs.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
HCO3	79.8065	1.5265	0.1845	0.7419	2.0229	1.0309	0.3608	0.0483
NPOC	11.929	0.5725	22.3937	8.6565	0.5403	20.2893	0.1741	0.074
SO4	50.9997	6.4032	4.0756	1.0294	0.0385	15.3697	1.4461	0.9491
NO3	46.1403	1.6435	19.7176	4.984	14.7953	7.2348	0.0535	4.2735
Br	11.1214	29.9674	4.3713	8.1315	0.1072	1.86	0.9125	0.0835
Ca	86.7777	3.9016	0.0228	0.6633	1.8739	0.4979	0.1197	0.1843
Cl	6.6735	61.0455	3.8102	1.9217	1.0196	2.0352	0.1802	0.8418
F	67.4985	1.1986	0.8339	0.5483	5.399	1.0287	0.0115	0.0171
K	63.451	0.6642	2.4319	0.1737	3.6404	5.9356	5.0846	1.3797
Mg	68.7053	11.0705	0.6536	3.12	0.3695	0.6634	0.0123	1.3266
Na	5.7404	63.8368	4.8002	1.8756	1.4737	1.8242	0.0302	0.9679
P	8.6114	10.427	0.4008	3.996	23.2314	0.4587	22.5276	10.9046
S	57.9831	6.9802	2.2068	1.1344	0.0474	11.3244	1.0959	1.1566
Al	78.1715	4.9224	0.0237	2.5569	8.00E-04	0.8312	0.6489	0.4408
As	7.3256	2.1145	20.5614	0.5574	5.767	7.4176	0.0773	0.0065
Ba	31.8119	0.2363	0.8949	14.3233	0.0384	6.5435	18.2496	14.2552
Be	26.4246	0.1039	0.0945	0.8064	7.7128	0.6158	5.4161	0.28
Cd	2.3532	4.7319	0.1936	22.1455	0.3642	0.3295	22.8143	8.5232
Ce	90.9046	0.318	2.2427	0.3742	0.3809	0.2203	0.4504	0.094
Co	4.3321	24.147	16.7457	1.0435	15.9577	13.1279	0.0752	3.1342
Cr	1.1573	0.0265	0.4516	0.2964	2.1344	2.4254	0.0532	5.4233
Cs	2.5715	56.1695	1.9294	0.5982	0.7136	1.3671	2.138	5.3011
Cu	8.4667	5.6908	8.4918	14.1503	13.8071	8.3253	0.2987	0.8764
Dy	80.8057	4.9797	4.9705	2.3233	1.3019	0.0516	0.2988	0.0554
Er	70.4005	6.4189	5.0387	5.1358	2.2386	0.0057	0.0159	0.0225
Eu	78.0281	2.4215	4.8642	0.8739	0.0391	0.4049	0.9466	0.5539
Fe	48.728	3.9188	24.2794	4.2248	0.0202	2.7249	0.3225	0.5196
Gd	84.1767	4.0664	5.7326	1.3062	0.5206	0.2786	0.2794	0.1237
Ho	71.7625	5.2242	5.0889	3.4848	2.1157	0.0125	0.2275	3.00E-04
La	81.5222	0.3157	8.3022	0.3152	0.5521	0.3981	0.597	0.1053
Li	15.2864	17.8247	3.1975	0.7223	1.5423	3.9957	3.2233	0.6556
Mn	1.281	10.5282	35.2244	24.3764	3.9404	2.5024	0.1706	7.0004
Mo	57.8485	2.4396	0.6156	7.0415	1.842	0.0113	0.019	0.8884
Nb	3.0815	4.2354	0.5932	4.2108	0.3442	2.1343	8.4702	1.1796
Nd	82.3197	1.3749	6.7358	0.2755	0.7784	0.1236	0.006	0.0065
Ni	24.4322	27.7943	0.5695	5.8202	6.2607	11.233	1.8375	0.0397
Pb	60.3474	3.8108	2.7702	8.1465	3.4945	6.4568	0.0021	0.1288
Pr	85.9928	0.7497	7.1533	0.2947	0.3599	0.4644	0.1138	2.00E-04
Rb	21.7662	8.6384	1.5898	1.8183	1.0821	18.7662	5.7251	0.8783
Sb	43.5197	0.0074	0.4528	23.228	3.7375	0.1708	0.662	0
Se	30.4408	3.4122	0.4788	20.4706	1.2915	3.5399	2.6913	1.8355
Si	34.9682	3.2816	0.029	8.9492	0.0114	0.0272	0.0104	0.0051
Sm	84.5355	2.4758	6.6675	0.9077	0.3834	0.4032	0.2071	0.2398
Sn	1.0754	26.0866	3.0231	0.0314	0.0328	0.6842	0.0214	0.3529
Sr	82.4984	2.092	0.1932	0.0773	3.3263	1.00E-04	0.0044	0.5058
Th	63.7725	0.0893	1.6978	3.00E-04	2.6386	0.4233	0.7362	0.0292
Ti	70.8417	2.227	0.5819	4.1634	4.00E-04	5.7099	1.7742	1.4067
U	62.9341	3.056	1.5686	2.1937	13.1291	0.6517	1.6792	4.6537
V	6.1274	4.2476	0.8247	7.5038	0.1292	17.3634	1.1094	2.0742
Y	66.3084	9.5089	5.4944	3.1289	3.379	0.1507	0.1714	0.2093
Yb	57.2677	6.348	4.7618	10.0617	2.4935	0.0053	0.062	0.0305
Zn	9.3312	2.0897	7.7677	11.8581	8.1858	1.083	4.6894	2.0919
Zr	0.024	38.5322	8.00E-04	0.0116	1.6677	0.2677	1.8435	0.0814

Table A.3 Relative contributions of analytes to the first eight Principal Components

The measures of relative contribution and absolute contribution assist in recognizing independent relationships and associations between and within the PCs. Distinct elemental associations may represent mineralogical relationships that are governed by



mineral stoichiometry, adsorption, weathering, groundwater effects due to eH and pH conditions and non-linear gravitational effects.

A graphical way of describing the inter-element relationships within and between the PCs is through the use of a PC biplot. See Grunsky (2010) for a description of PC biplots.

A.1 Principal Components Associated with Rock Type

Figure A.2 shows a PC biplot for PC1 vs. PC2. The scores (loadings) of the elements are coloured according to the Goldschmidt classification and the affinity to lithophile, siderophile and chalcophile. Anions that reflect features related to soil/water conditioning are also coded with a unique colour. The biplot shows a relative increase in NO_3^- , HCO_3^- , Ca, U, Sr, F^- , Mg, SO_4^{2-} , Ba, Mo along the negative portion of the PC1 axis. The bedrock types associated with these elements are primarily Carboniferous limestone as well as pelite, greywacke, quartzite and sandstone that are interbedded with the limestone or occur around the margins of the Carboniferous basin. Along the positive portion of the PC1 axis, there is relative enrichment in lithophile elements (Ce, Nd, La, Al, Ti, Sm, Gd, Eu, Yb, Ho, Er, Y, Th), Fe (siderophile) and Pb (chalcophile). This portion of the PC1-PC2 biplot is dominated by granite. Figures A.3 to A.8 show biplots for PC2-PC3, PC3-PC4, PC4-PC5, PC5-PC6, PC6-PC7 and PC7-PC8, which are the dominant components as indicated in Table A.1. Figures A.9 to A.22 show the PC1-PC2 biplot for each of the rock types. Comparisons of the biplots indicate that there is a significant amount of overlap between the rock types on the PC1-PC2 biplot.

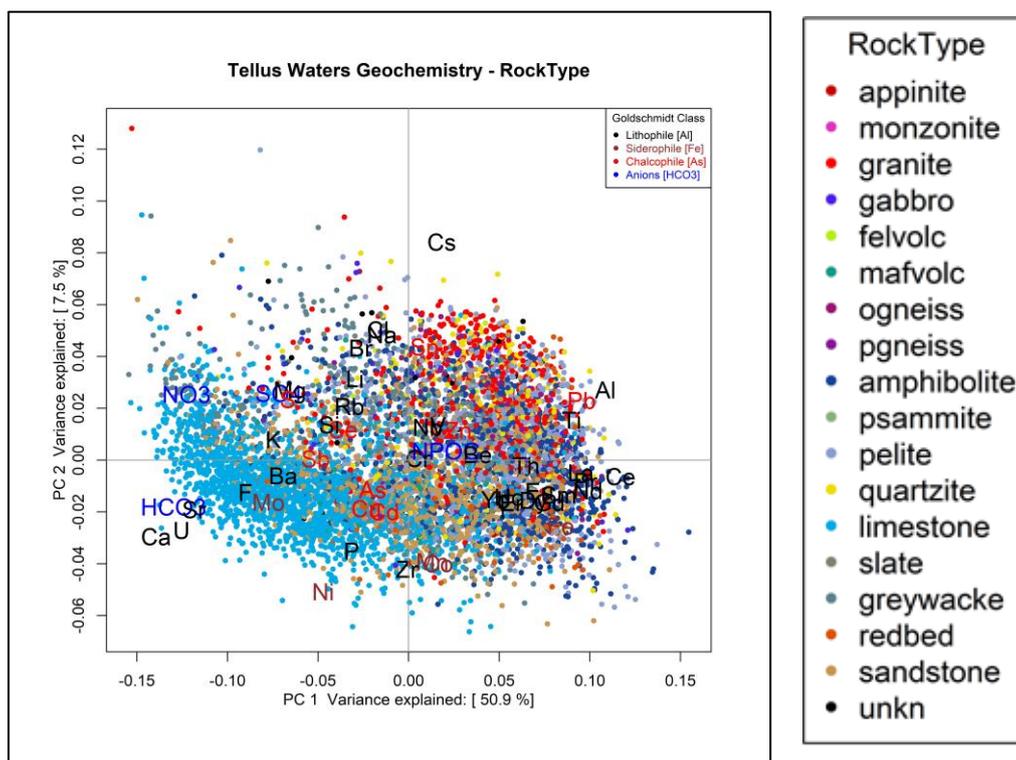


Figure A.2 Principal Component biplot for PC1 v PC2 with legend (right)



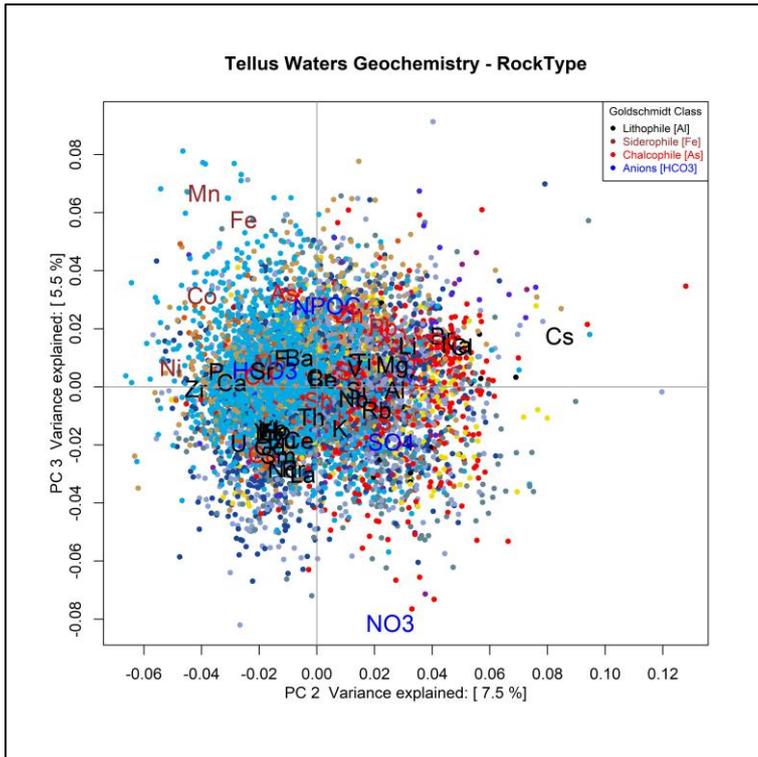


Figure A.3 Principal Component biplot for PC2 v PC3. (legend as for Figure A.2)

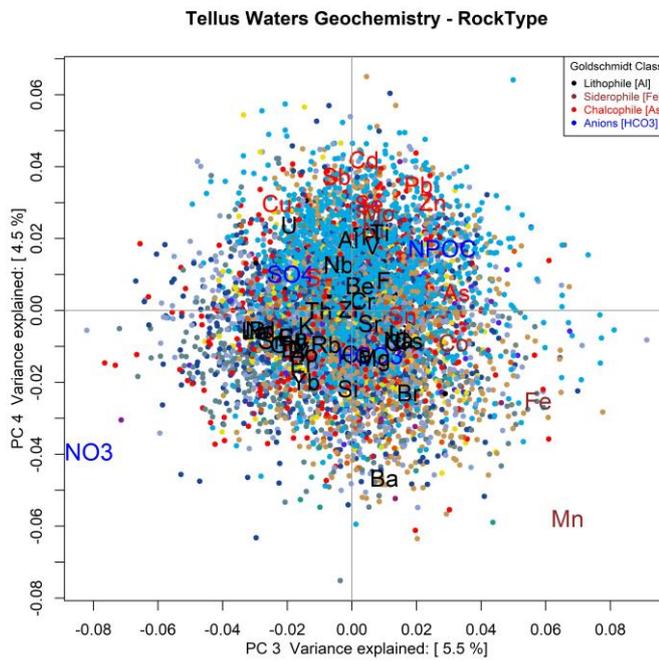


Figure A.4 Principal Component biplot for PC3 v PC4. (legend as for Figure A.2)



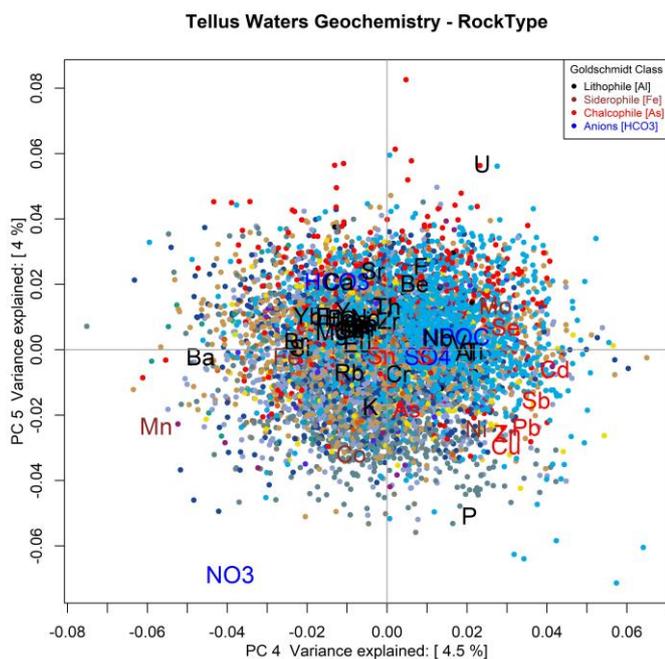


Figure A.5 Principal Component biplot for PC4 v PC5. (legend as for Figure A.2)

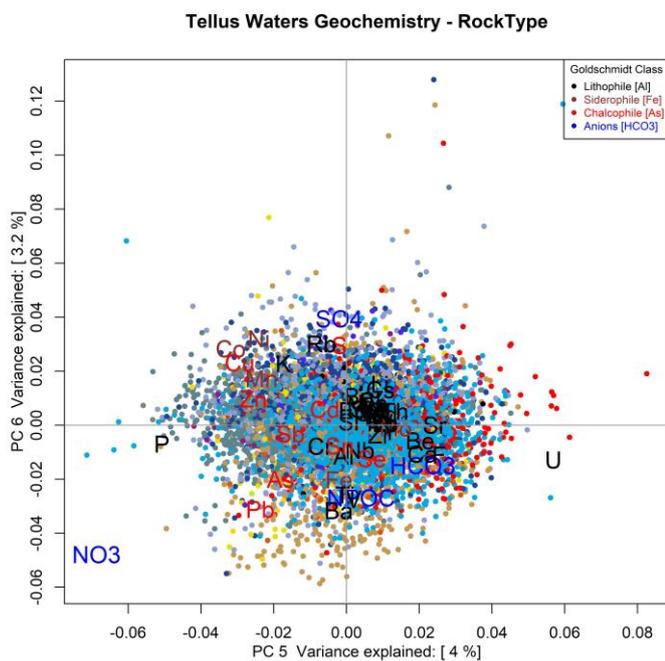


Figure A.6 Principal Component biplot for PC5 v PC6. (legend as for Figure A.2)



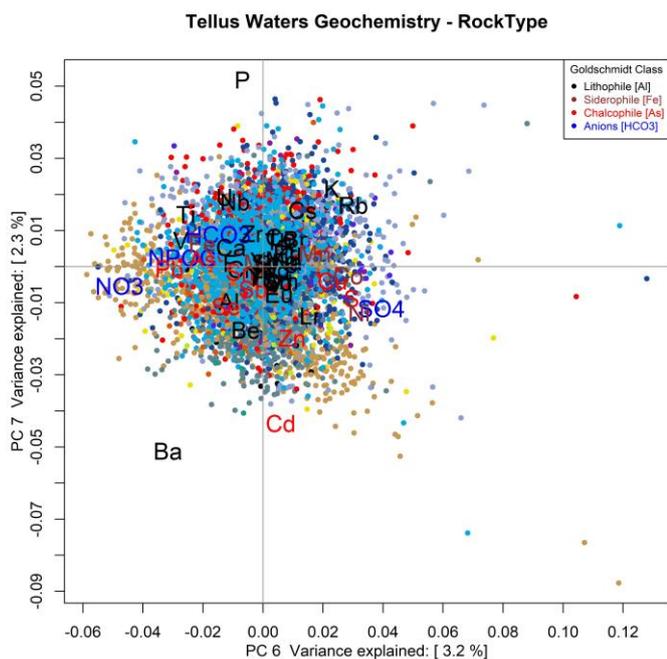


Figure A.7 Principal Component biplot for PC6 v PC7. (legend as for Figure A.2)

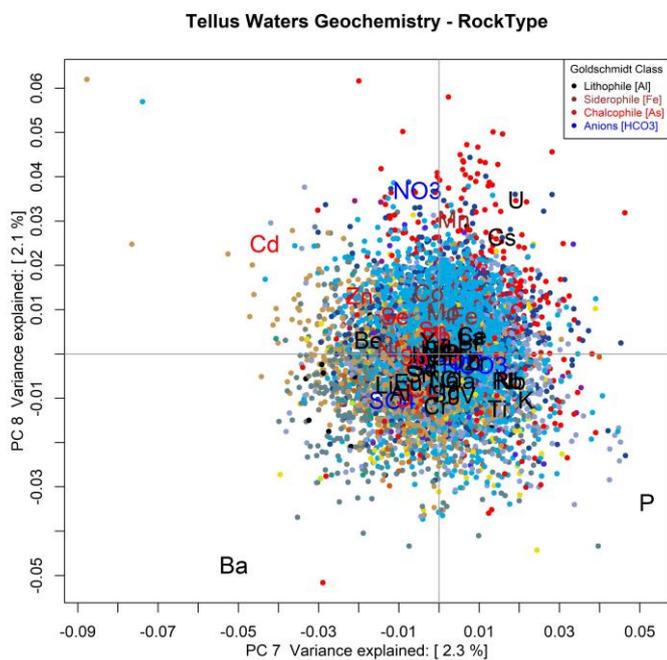


Figure A.8 Principal Component biplot for PC7 v PC8. (legend as for Figure A.2)



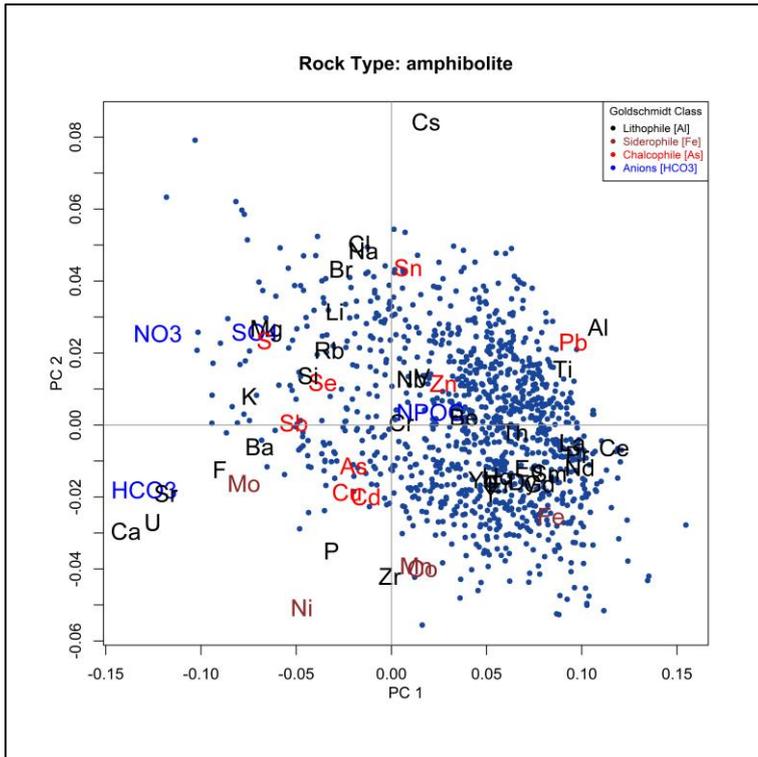


Figure A.9 Principal Component biplot for PC1 v PC2 for amphibolite

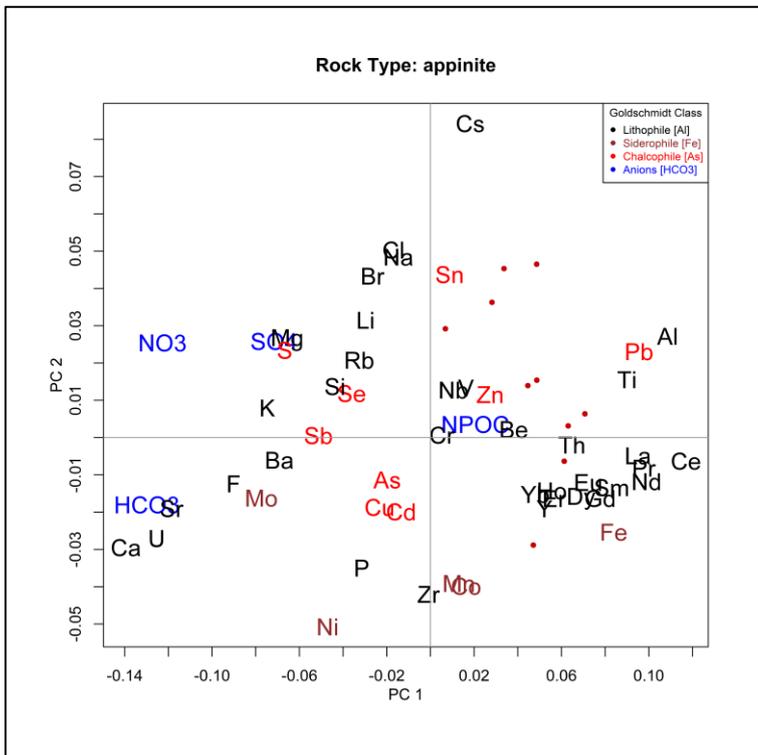


Figure A.10 Principal Component biplot for PC1 v PC2 for appinite



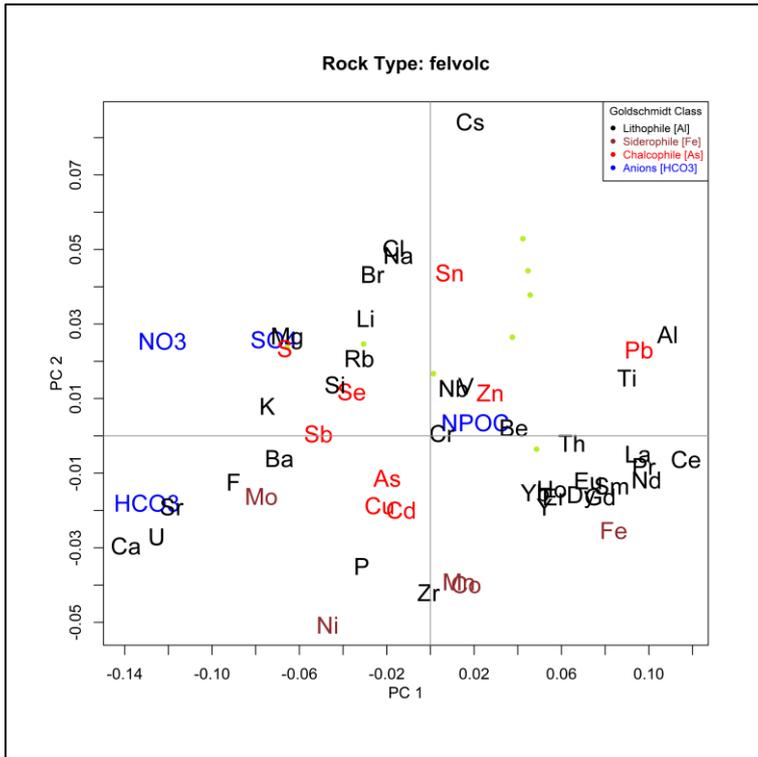


Figure A.11 Principal Component biplot for PC1 v PC2 for felsic volcanics

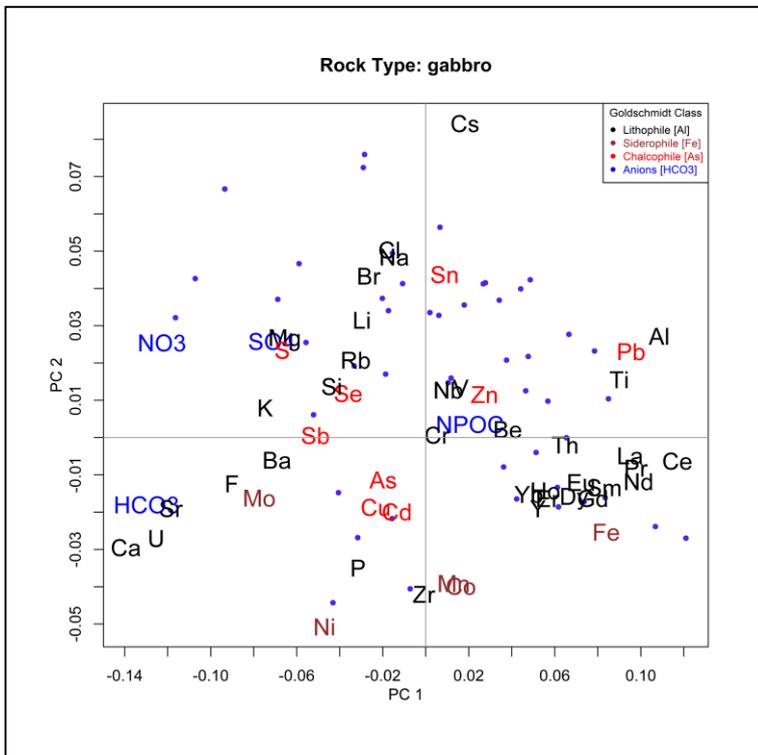


Figure A.12 Principal Component biplot for PC1 v PC2 for gabbro



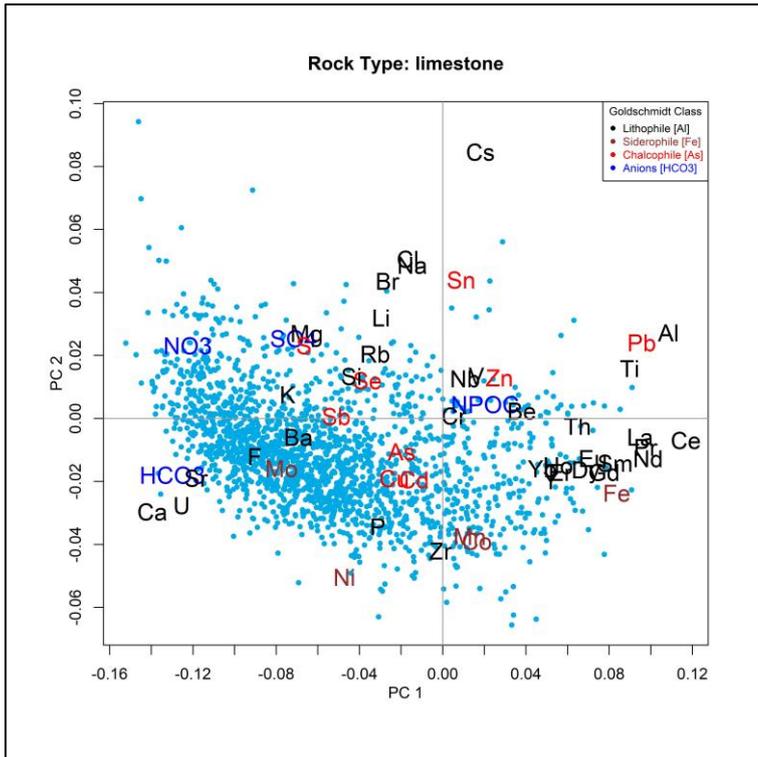


Figure A.15 Principal Component biplot for PC1 v PC2 for limestone

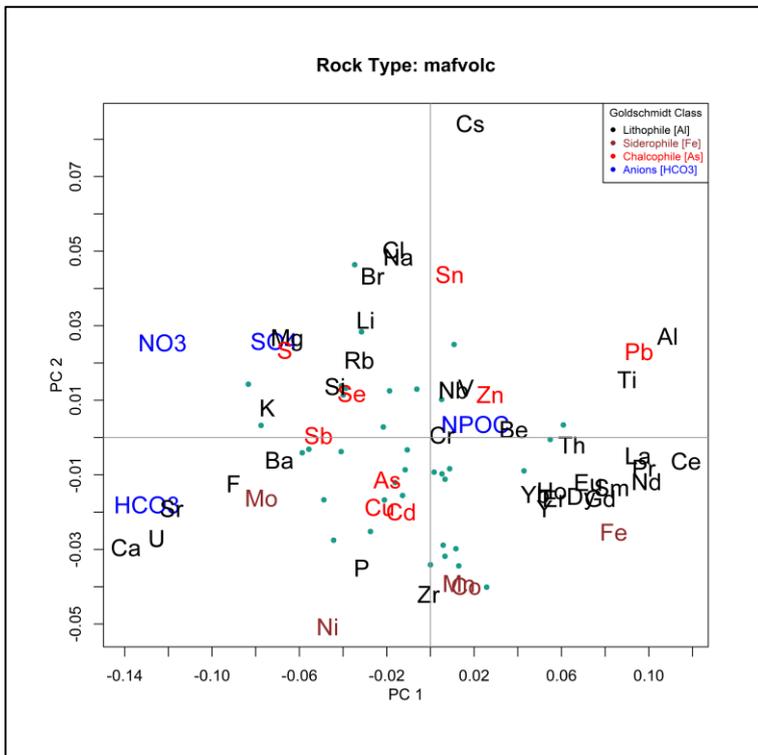


Figure A.16 Principal Component biplot for PC1 v PC2 for mafic volcanic



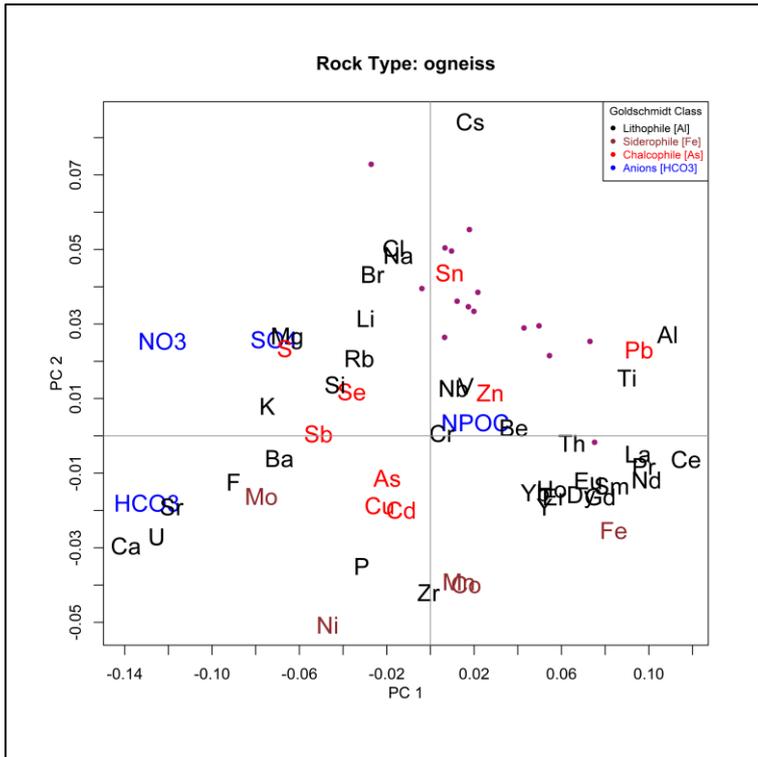


Figure A.17 Principal Component biplot for PC1 v PC2 for orthogneiss

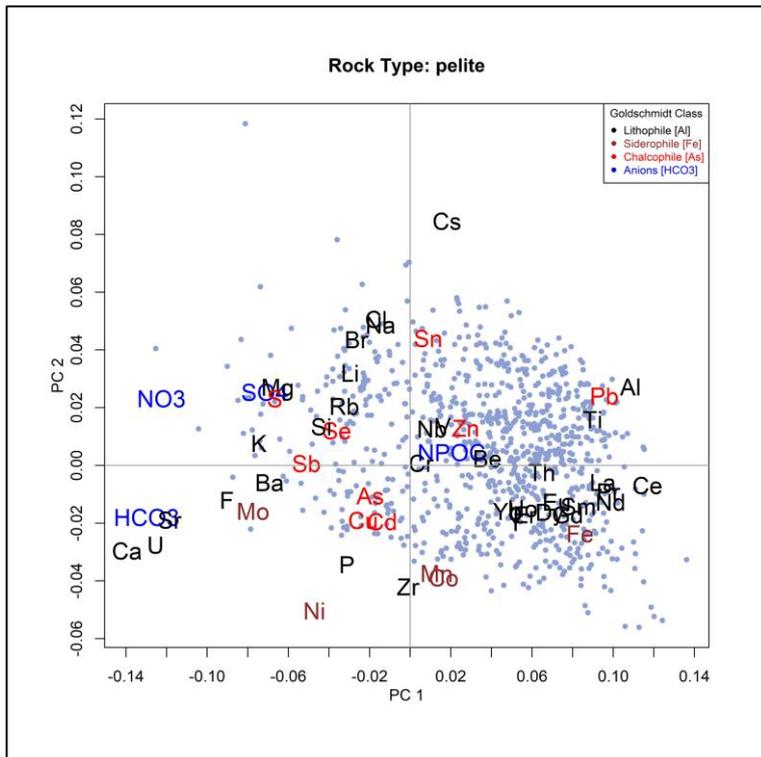


Figure A.18 Principal Component biplot for PC1 v PC2 for pelite



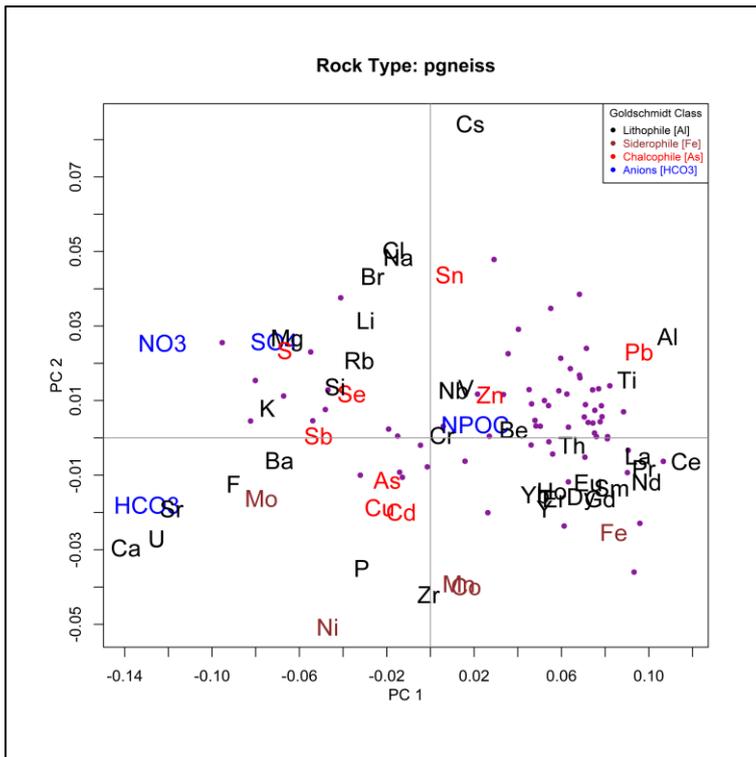


Figure A.19 Principal Component biplot for PC1 v PC2 for paragneiss

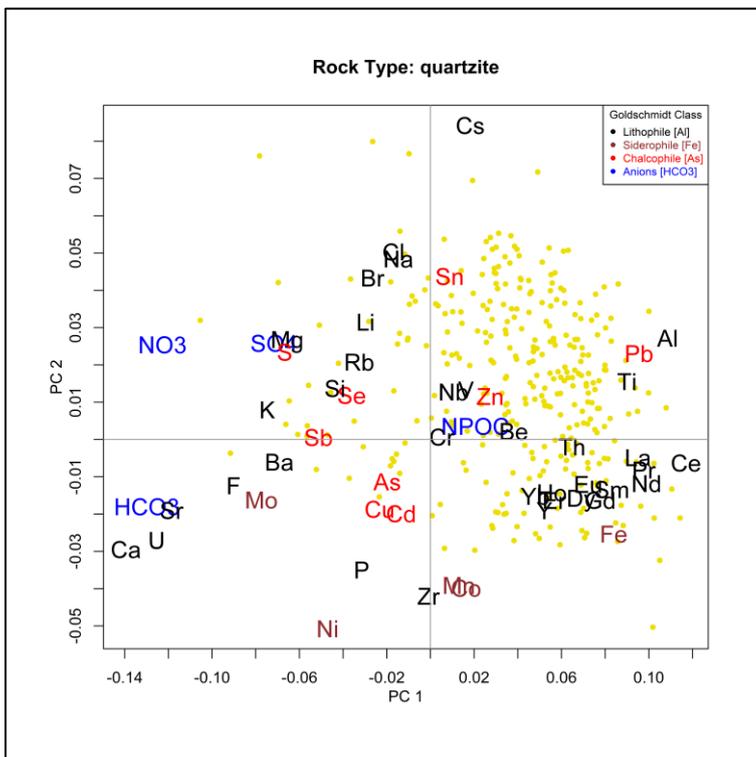


Figure A.20 Principal Component biplot for PC1 v PC2 for quartzite



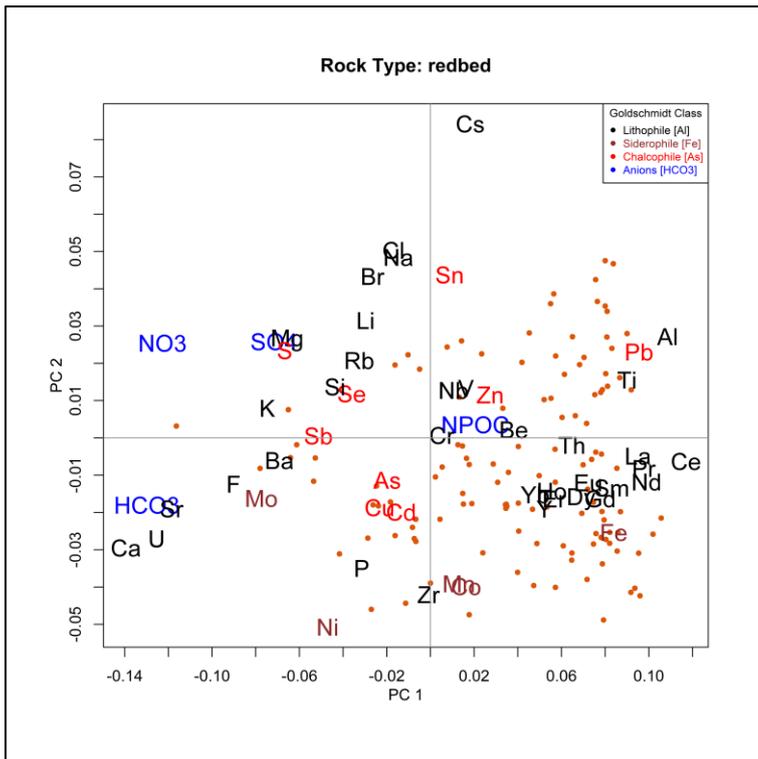


Figure A.21 Principal Component biplot for PC1 v PC2 for red beds

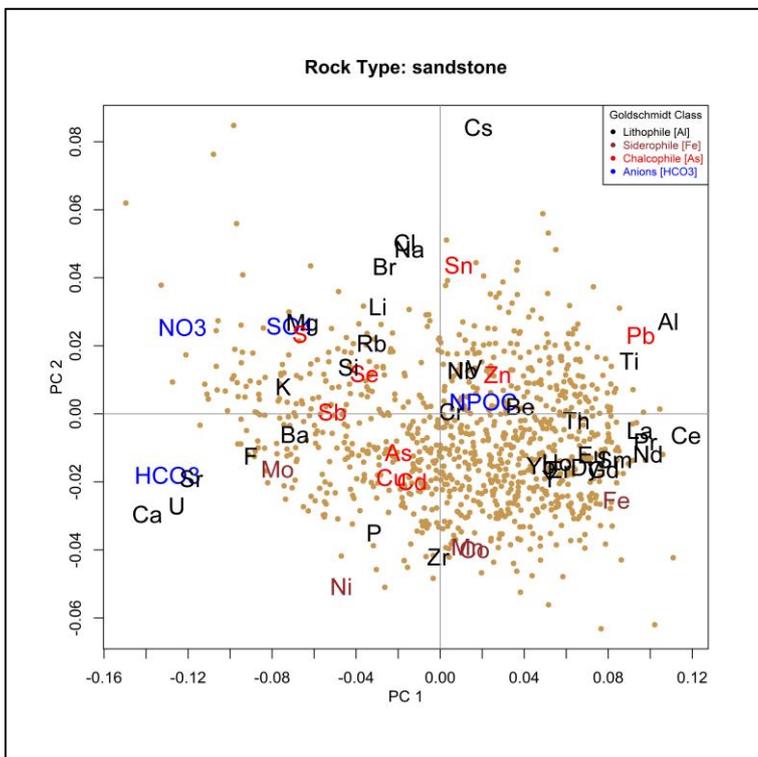


Figure A.22 Principal Component biplot for PC1 v PC2 for sandstone



For many geochemical surveys that are geospatially based, the PCs, based on the inter-relationships of the elements, reveal features related to mineral stoichiometry and hence geological processes, which can reveal patterns in a geospatial context. Figures A.23 to A.30 show kriged images calculated for each of the eight dominant PCs. The kriged image of PC1 (Figure A.23) shows a clear distinction between (1) the Lower Palaeozoic carbonate and clastic rocks and (2) the Caledonian intrusive rocks and Precambrian metamorphic rocks. The map of PC2 (Figure A.24) contrasts the Caledonian and Palaeogene intrusive rocks with the Precambrian and Palaeozoic sedimentary rocks. The map of PC3 (Figure A.25) shows areas of relative Fe-Mn enrichment in stream waters (positive PC3) and areas of nitrate (NO_3^-) enrichment (negative PC3). Fe and Mn are positively correlated with Co, As and Zn (Table A.2). Figure A.26 (PC4) reveals relative enrichment of chalcophile elements (positive PC4 score), such as Zn, Cd, Pb, Sb, Mo and Cu, in areas distinct from those displaying lithophile element (Ba, Fe, Mn) and NO_3^- enrichment (negative PC4 scores) (Table A.2). Areas of positive PC4 scores that have relative enrichment of chalcophile elements occur in proximity to known Zn-Pb mineralization in County Monaghan but also over areas of mixed bedrock where they may be associated with organic-rich content in the underlying soils.

Negative PC5 scores (Figures A.5 and A.27) reflect relative enrichment in NO_3^- and P as well as chalcophile elements (Table A.2), and the mapped areas of low PC5 values in counties Monaghan and Louth reflect known Zn-Pb mineralization and tillage farming. Positive PC5 scores predominate in the western and northwestern part of the area, reflecting relative enrichment in U-F⁻-Sr-Ca-Be-Th-Mo. This combination of elements co-occurs within areas dominated by limestone and granite bedrock (Figures A.5 and A.27).

Positive scores of PC6 reflect relative general SO_4^{2-} enrichment associated with clastic sediments and granitoid rocks (Figure A.28). The negative PC6 scores reflect relative enrichment in NO_3^- , Ba, Ti, NPOC, V and Fe associated primarily with greywacke and sandstone. A small area of very high relative SO_4^{2-} enrichment in north County Clare (Figure A.28) reflects a seawater influence in a sample site close to the coast; other sites along parts of the coastline are also likely to be a consequence of marine influences. The relative enrichment of SO_4^{2-} in stream water in the area west of Lough Allen, southeast of Sligo, appears to be linked to drainage from coal mines. Relative enrichment of NO_3^- and NPOC is concentrated in north County Mayo in an area of peat and clastic bedrock.

The biplot of Figure A.8 shows relative enrichment of Ba and Cd associated with clastic rocks (Namurian sandstone/shale and Lower Palaeozoic greywacke) in PC7 and shows up geospatially in south Mayo and around Lough Allen. Negative PC7 values are also associated with Zn-Pb mineralization at Tynagh in County Galway and in County Monaghan (Figure A.29).

Positive scores of PC8 (Figure A.8) are associated with granitoid rocks and show relative enrichment in NO_3^- along with Cd, Mn, Cs and U. Figure A.30 shows that these positive scores are associated with the Caledonian Barnesmore Granite (County Donegal) and Galway Granite and the Palaeogene felsic intrusion near Carlingford in County Louth. All are relatively radioelement-rich with the Barnesmore Granite notable for its associated uranium mineralization. Negative PC8 scores are associated with relative enrichment in Ba and P (Figure A.8). On the map (Figure A.30) the lowest PC8 scores are most consistently associated with coastal areas, suggesting a marine influence.



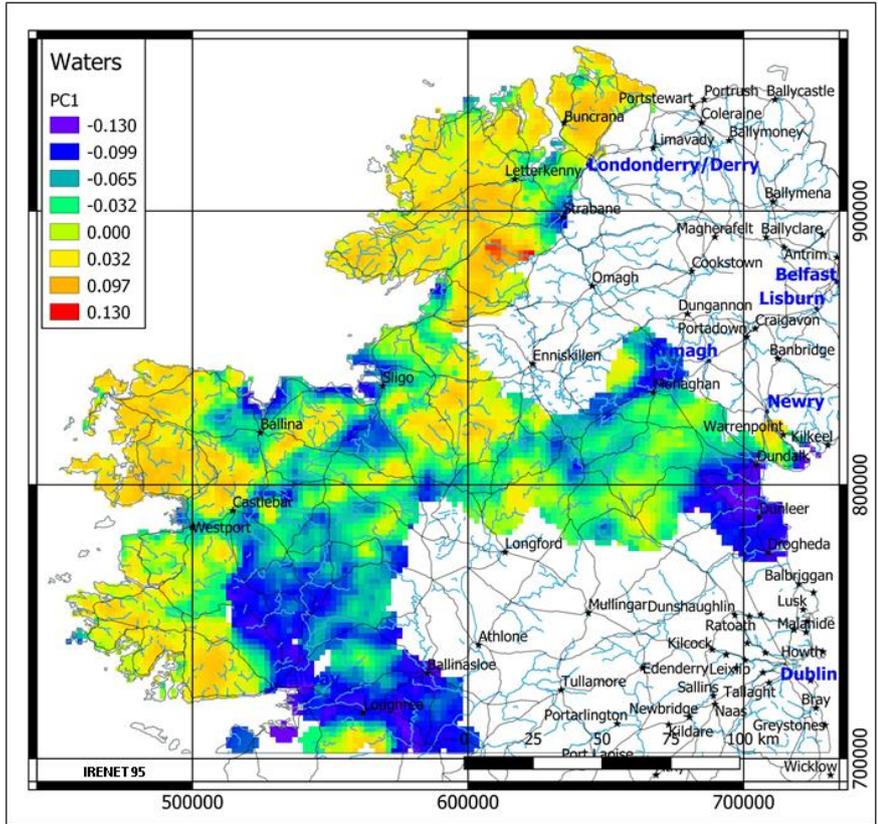


Figure A.23 Kriged image for PC1

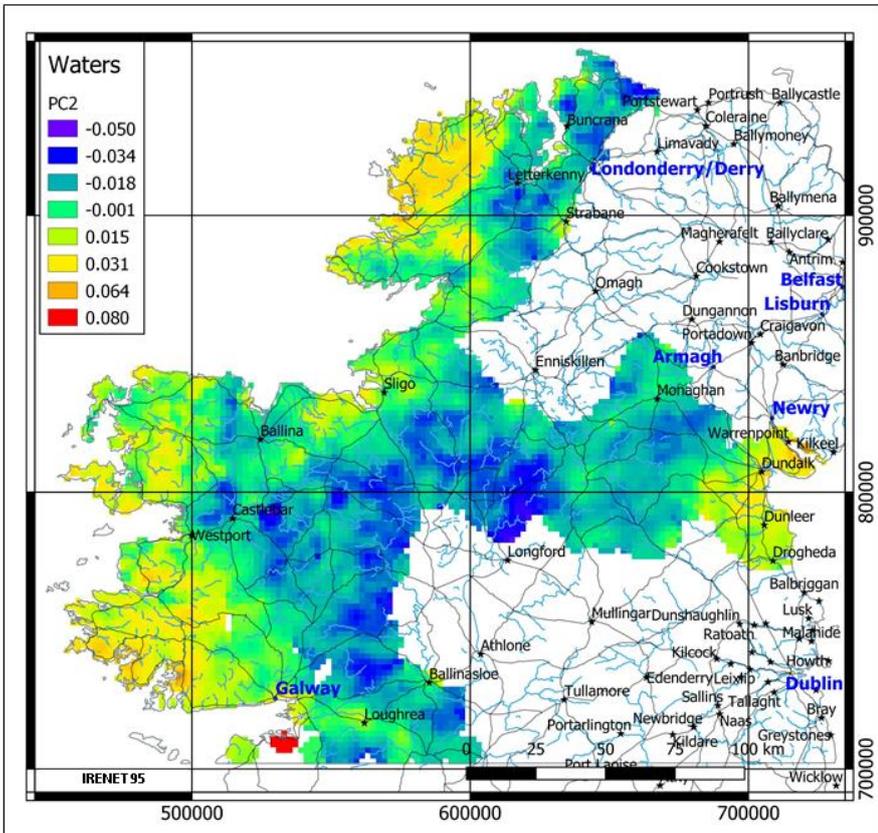


Figure A.24 Kriged image for PC2



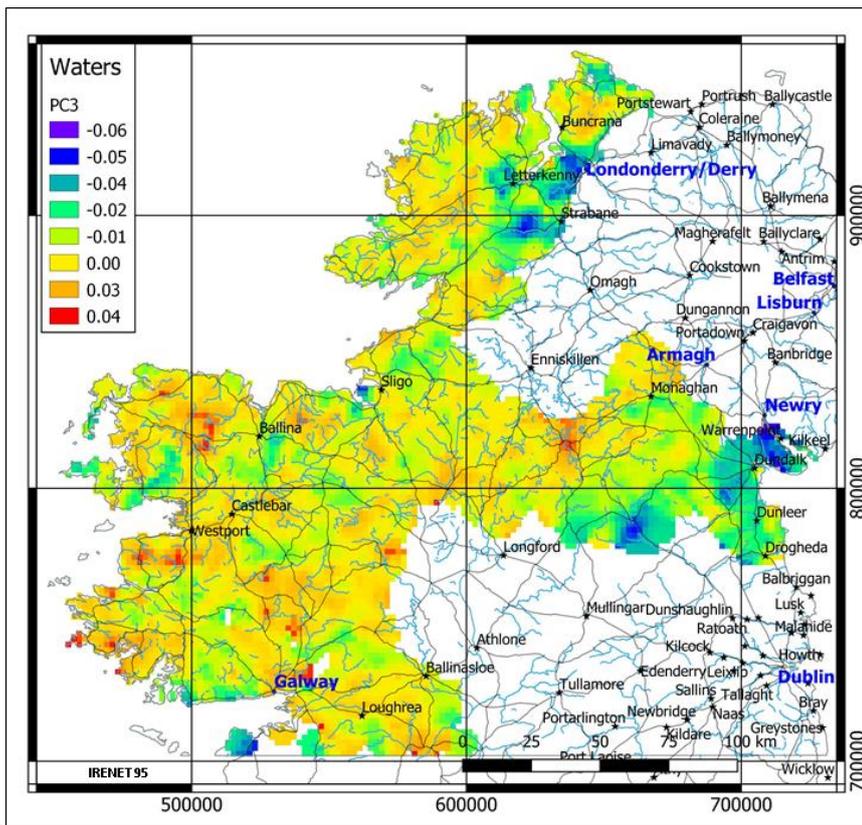


Figure A.25 Kriged image for PC3

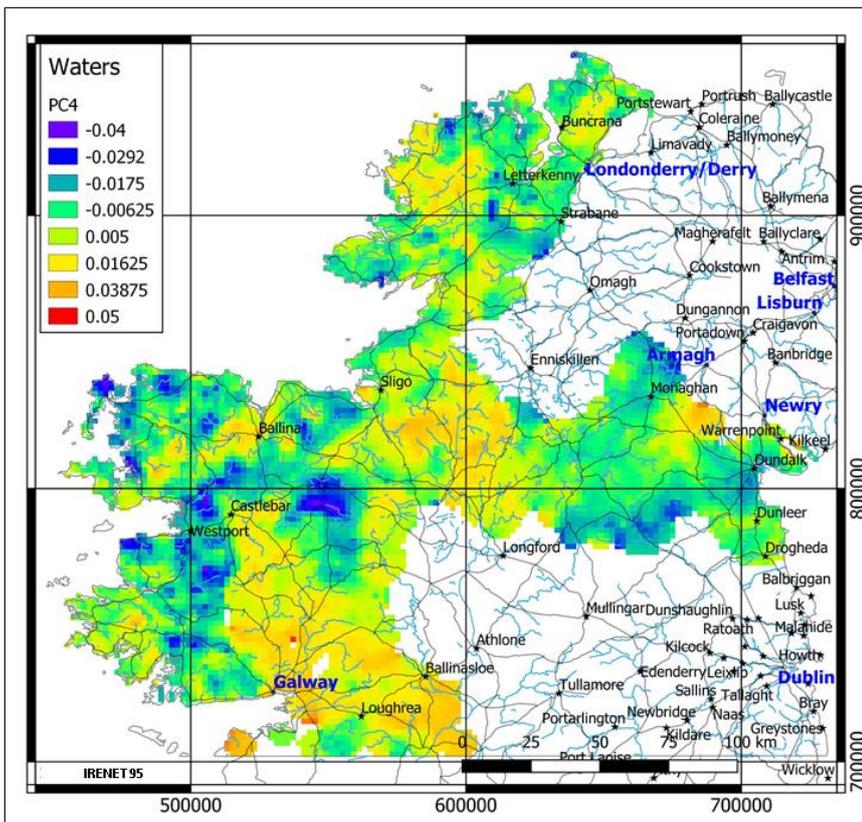


Figure A.26 Kriged image for PC4



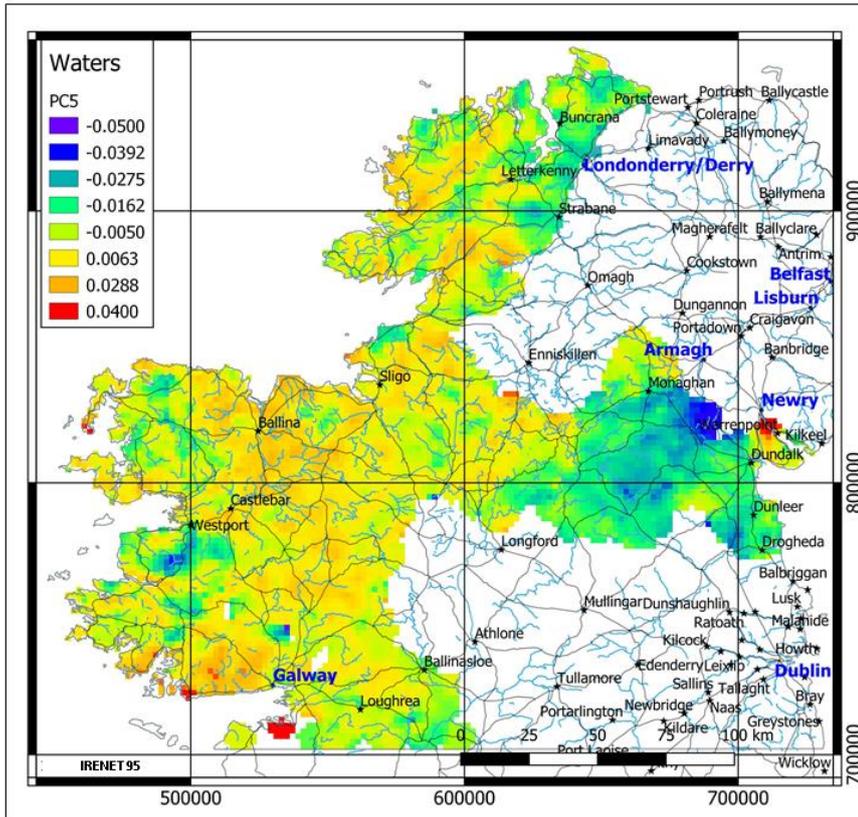


Figure A.27 Gridded image for PC5

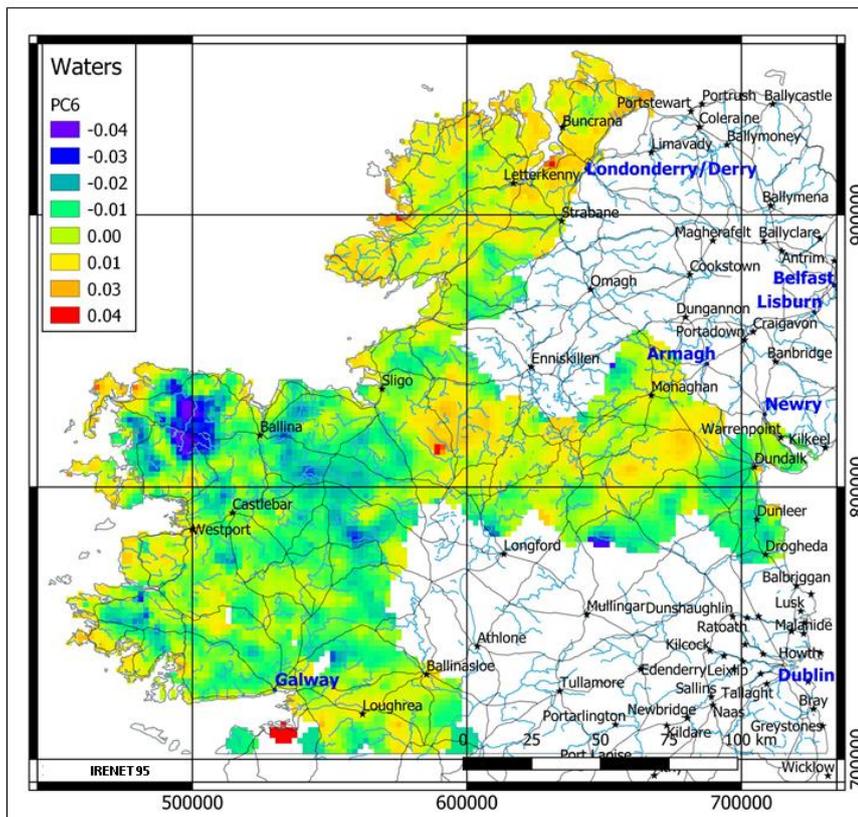


Figure A.28 Gridded image for PC6



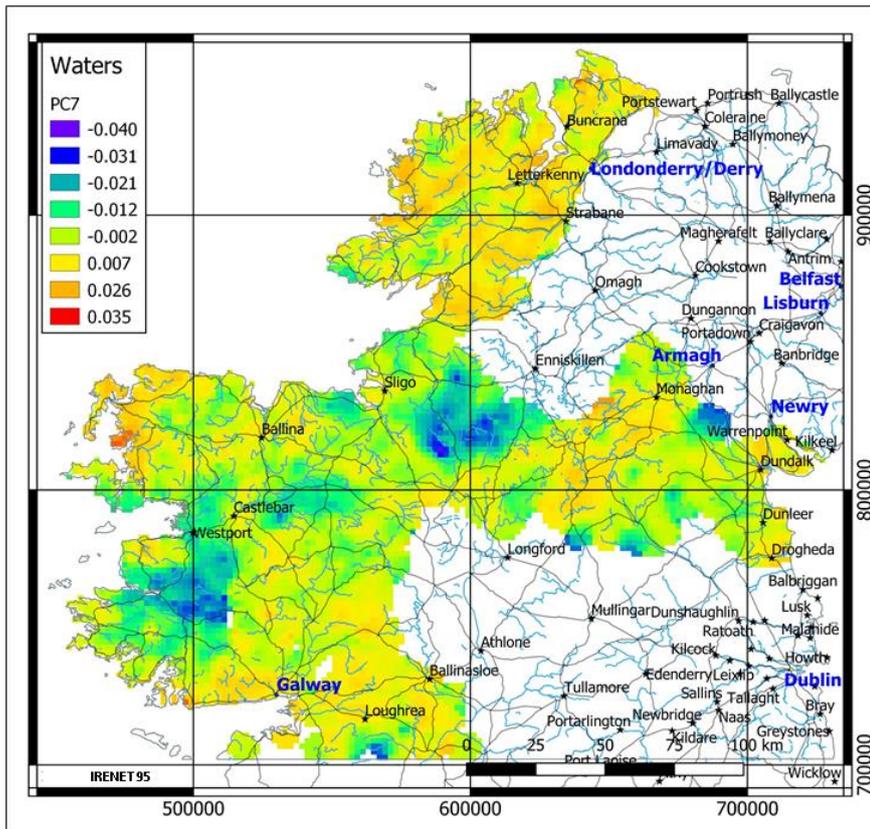


Figure A.29 Kriged image for PC7

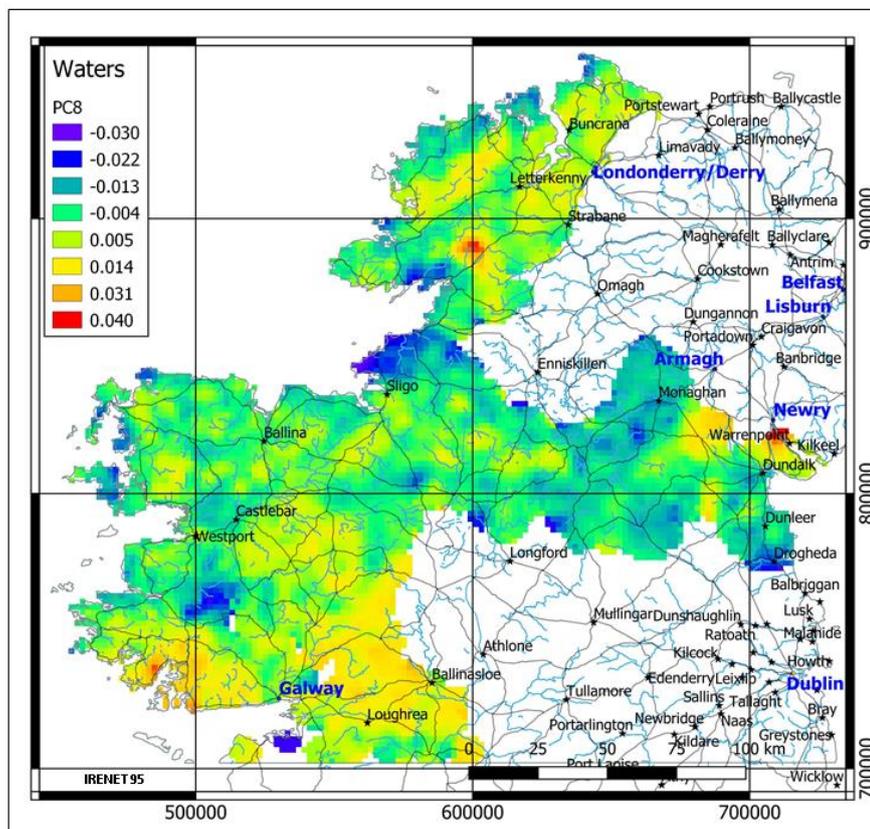


Figure A.30 Kriged image for PC8



A.2 Principal Components Associated with Teagasc Subsoils

Figures A.31 to A.37 show the principal component biplots for PC1-PC2, PC2-PC3, PC3-PC4, PC4-PC5, PC5-PC6, PC6-PC7 and PC7-PC8, respectively, for Tellus stream water data classified by Teagasc subsoil type. The loadings of the elements/anions and the scores of the waters sites are identical to those shown in the biplots of Figures A.2 to A.8 (rock types). The colours and symbols of the scores have been coded according the legend for the Teagasc subsoil types shown in each figure.

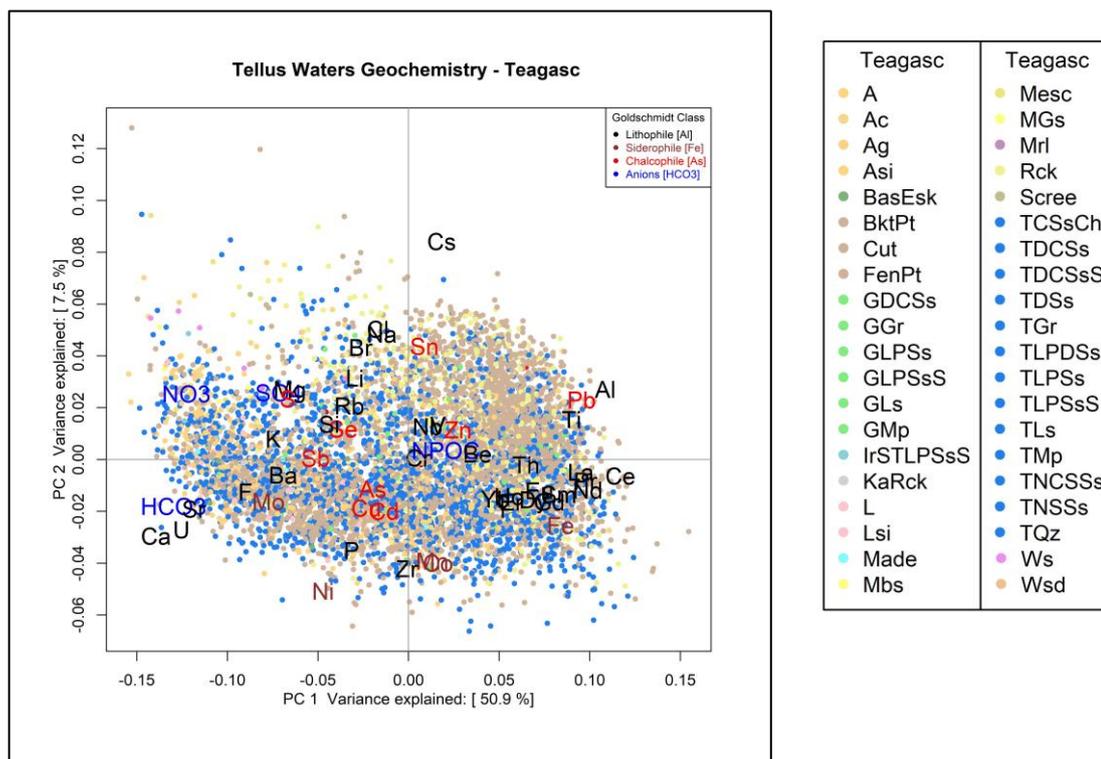


Figure A.31 PC1 v PC2 for Teagasc subsoil classes (with legend)



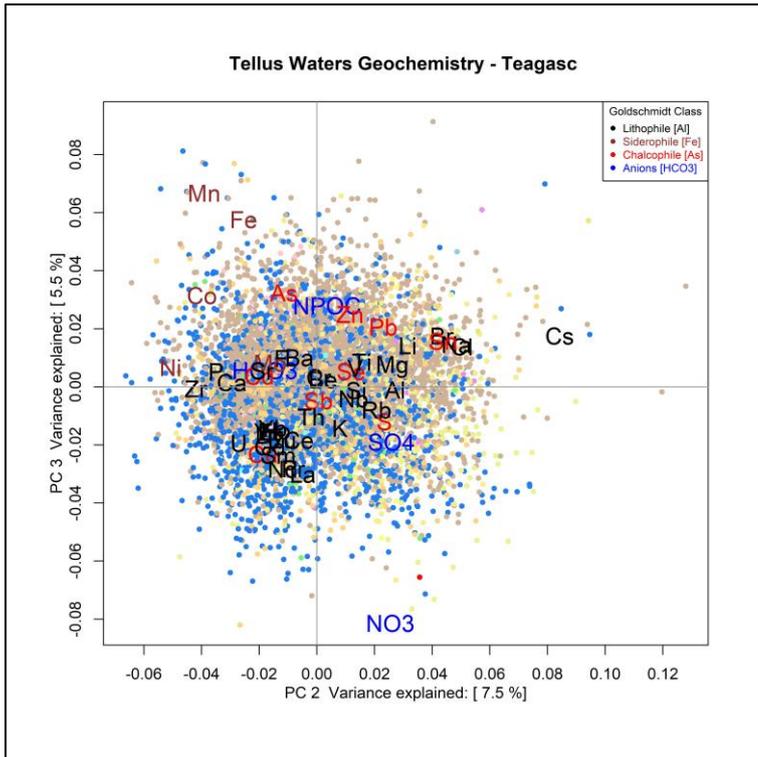


Figure A.32 PC2 v PC3 for Teagasc subsoil classes. (legend as for Figure A.31)

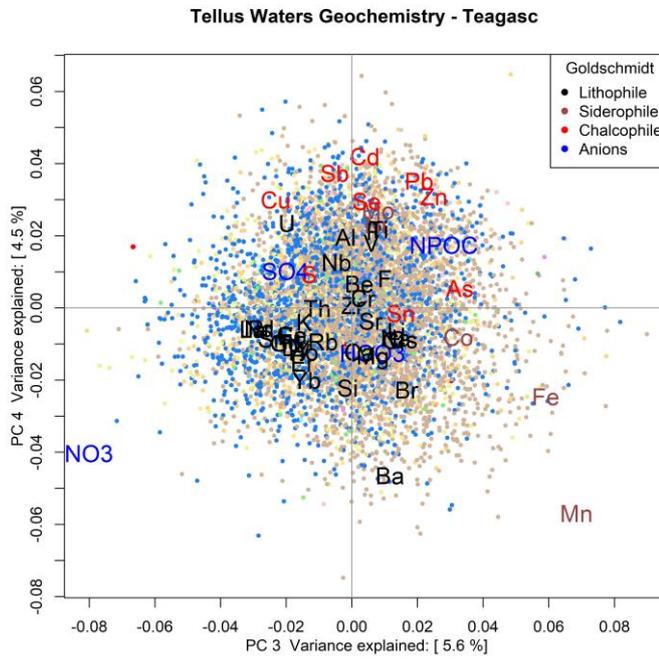


Figure A.33 PC3 v PC4 for Teagasc subsoil classes. (legend as for Figure A.31)



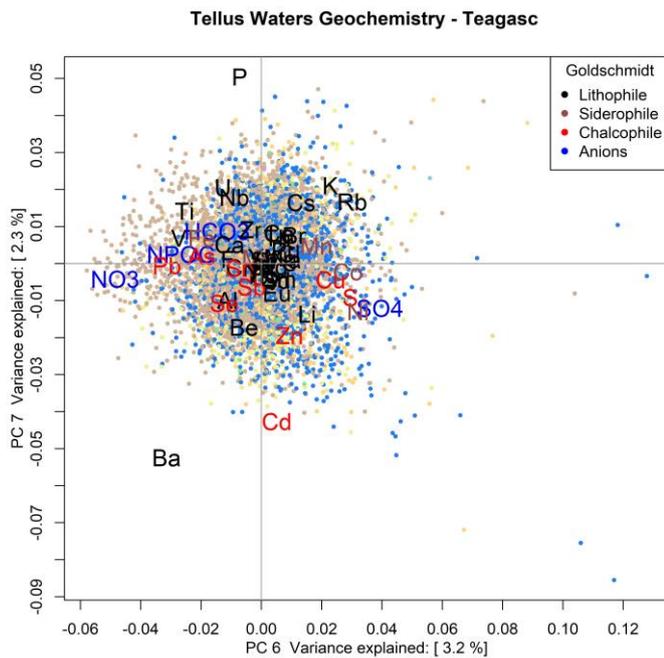


Figure A.36 PC6 v PC7 for Teagasc subsoil classes (legend as for Figure A.31)

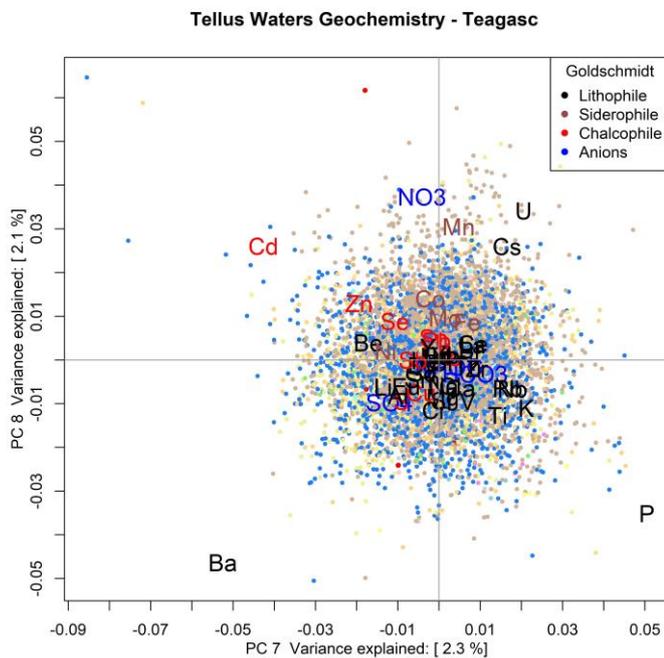


Figure A.37 PC7 v PC8 for Teagasc subsoil classes (legend as for Figure A.31)

The PCA biplot of Figure A.31 shows a distinct contrast between stream waters draining peat and those draining Alluvium (A), till derived from limestone (TLs) and other tills.



Peat displays an association of Cs-Cl-Na-Al-Pb-Ti whereas the tills and alluvium display a Ca-U-Sr- HCO_3^- F^- NO_3^- -Mo-Ba-K-Ni-P-Zr association. Figures A.38 to A.48 show the PC1-PC2 biplot for each of the dominant Teagasc subsoil classes, including alluvium (A), cutover peat (Cut), tills derived from sandstone/shale (TDCSs, TDSs, TLP), limestone (TLs) and others (TMp, TNSSs, TQ). Glaciofluvial sand and gravel subtypes (G) do not feature prominently in the study area.

Figures A.39 (BkPt), A.41 (Rck), A.44 (TGr), A.47 (TMp) and A.48 (TNSSs) display an association along the positive PC1 and positive PC2 axes that reflects relative enrichment with Al-Ti-Fe and several rare earth elements. This indicates an association with silicate minerals. Figures A.38 (A), A.40 (Cut), A.45 (TLPSs) and A.46 (TLs), show an association along the negative PC1 axis that corresponds to relative enrichment in Ca-U-Sr- HCO_3^- F^- NO_3^- - SO_4^{2-} -Mg-Mo-Ba-K-Ni-P-Zr-Mn-Co. This trend is primarily associated with carbonate bearing assemblages along with clastic sediments. The NO_3^- - HCO_3^- - SO_4^{2-} association likely represents agricultural effects.

Waters draining till derived from metamorphic rocks (TMp) and blanket peat (BkPt) are located long the positive PC1 axis of Figure A.31 and map as regions of positive PC1 scores on Figure A.23. Negative PC1 scores are associated with alluvium and till derived from limestone (Figure A. 23).

These two major groups overlap and are not clearly distinguished on the PC1 v PC2 biplot other, indicating that there is a transition between the geochemistry of materials derived from crystalline rocks (TMp) and those derived from sedimentary rocks (TLs).

The map of PC2 (Figure A.24) shows positive scores in the areas dominated by peat (BkPt), till derived from metamorphic rocks (TMp) and bedrock close to surface (Rck). Regions with negative PC2 scores are associated with till derived from limestone and Devonian-Carboniferous sandstones (TLs and TDCSs).

The PC2 v PC3 biplot (Figure A.32) shows NO_3^- plotting strongly on the negative PC3 axis – on the map of PC3 (Figure A.25) areas of negative PC3 coincide with known areas of high stream water nitrate concentrations in counties Louth and east Donegal. Mn and Fe plot strongly on the positive PC3 axis (Figure A.32). Positive PC3 values occur widely in the western and northwestern parts of the area (Figure A.25) and have a particular but not exclusive spatial association with peat. The high PC3 areas may reflect the occurrence of Fe-Mn oxyhydroxide coatings on mineral grains.

It is noted that PC 1 accounts for more than 50 % of the variability of the data and the remaining 52 components account for minor, under-sampled or random processes. Examination of the kriged PC map images can assist in determining the likelihood of identifying additional processes through the discovery of coherent geospatial patterns. Strongly negative PC6 values are observed for NO_3^- on the PC5 v PC6 biplot (Figure A.34) and shown in the map of Figure A.28. The PC6 v PC 7 biplot shows an association of Ba and Cd (negative PC7) (Figure A.36) and they can be observed on the PC 7 map (Figure A.29) where they are spatially associated with (i) former Zn-Pb mines in counties Galway (Tynagh) and Monaghan and (ii) tills derived from sandstones and shales, especially Namurian rocks around Lough Allen. The PC7 v PC 8 biplot shows U with a strong positive PC8 value (Figure A.37) – on the PC8 map (Figure A.30) high positive PC8 values coincide with granitic rocks, including the Barnesmore Granite and Main Donegal Granite in County Donegal, both of which have associated uranium mineralization.



Figures A.38 to A.48 show PC1 v PC2 biplots for stream waters for individual Teagasc subsoil classes. Tills derived from granites (TGr) and tills derived from limestones (TLs) show essentially opposing distributions (Figure A.44, A.46). The TLs subsoils display an association of Ca-U-Sr-HCO₃⁻-F⁻-NO₃⁻-SO₄²⁻-Mo-Ba-K-Ni-P-Zr, consistent primarily with carbonate bearing assemblages but also suggesting the influence of clastic sediments. The anions HCO₃⁻-NO₃⁻-SO₄²⁻ association (negative PC1) is most evident in surface waters classed as Alluvium (A) (Figure A.38), Cutover Peat (Cut) (Figure A.40), Lower Paleozoic sandstone and shale (TLPSSs) (Figure A.45) and Carboniferous limestone (TLs) (Figure A.46). The stream waters draining most of these subsoil types display a range of anion concentrations.

Stream waters draining the TGr subsoils have a strong lithophile association (Mg-K-Si-Rb-Na-Li-Nb-V-Cr-Ti-Al-REE), indicating the influence of silicate mineral assemblages on their chemistry (Figure A.44). This association is shared by subsoil comprising till derived from metamorphic rocks (TMP) (Figure A.47). Stream water draining till derived from Namurian sandstone and shales (TNSSs) (Figure A.48) plots between the limestone till and granite-derived till. Of particular interest is the difference between subsoil comprising blanket bog peat (BPt) (Figure A.41) and cutover (raised bog) peat (Cut) (Figure A.40), with the former resembling a combination of the TGr and TMP distributions and the latter the TLs distribution. This appears to reflect the occurrence of blanket peat in upland areas underlain mainly by metamorphic rocks and granites and of raised bogs in the midlands where the bedrock is dominated by limestone.

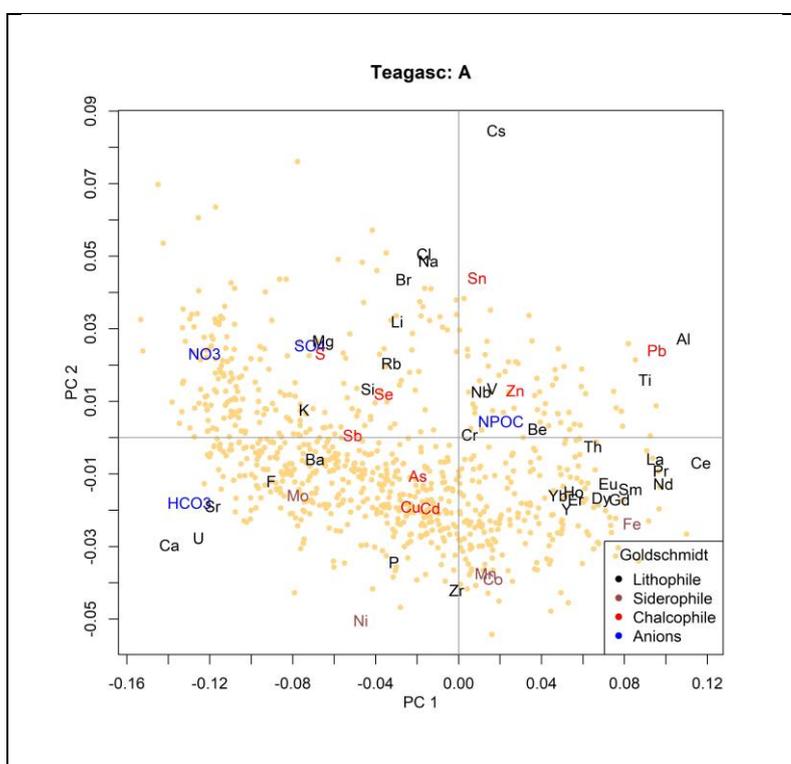


Figure A.38 PCA1 v PCA2 biplot for Teagasc subsoil class Alluvium



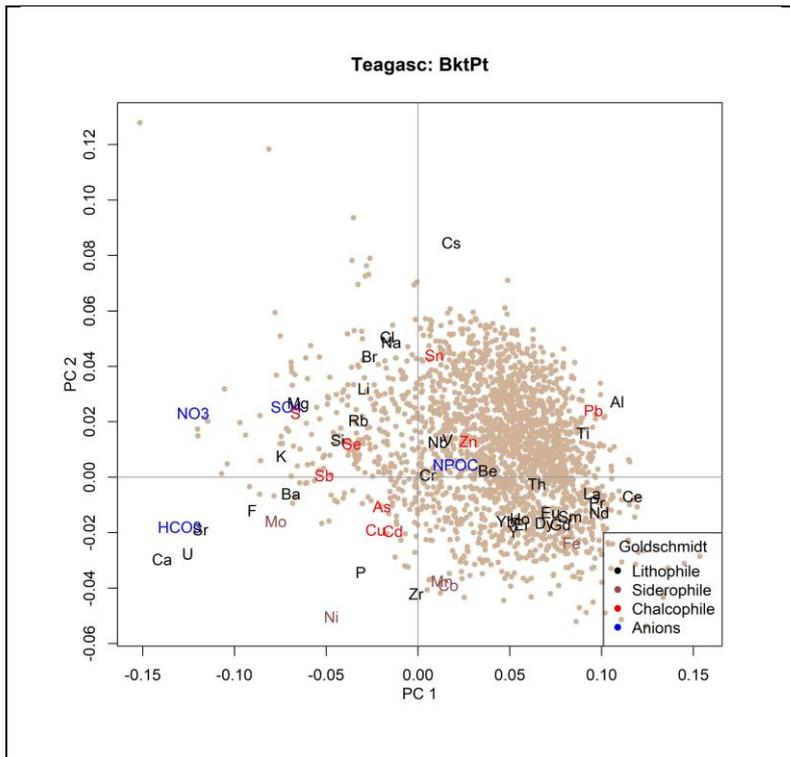


Figure A.39 PCA1 v PCA2 biplot for Teagasc subsoil class Blanket Peat

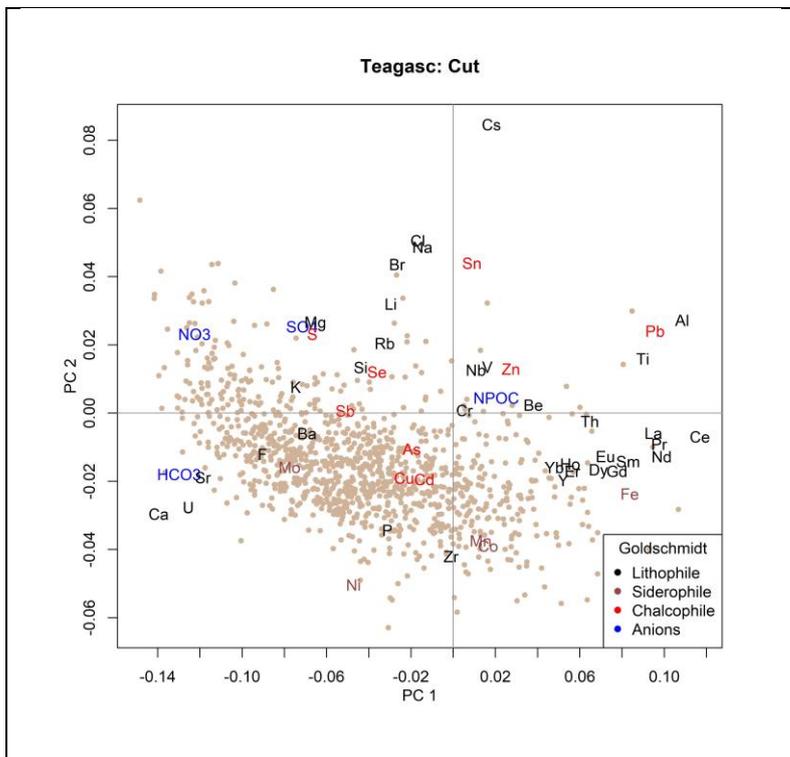


Figure A.40 PCA1 v PCA2 biplot for Teagasc subsoil class Cut over peat



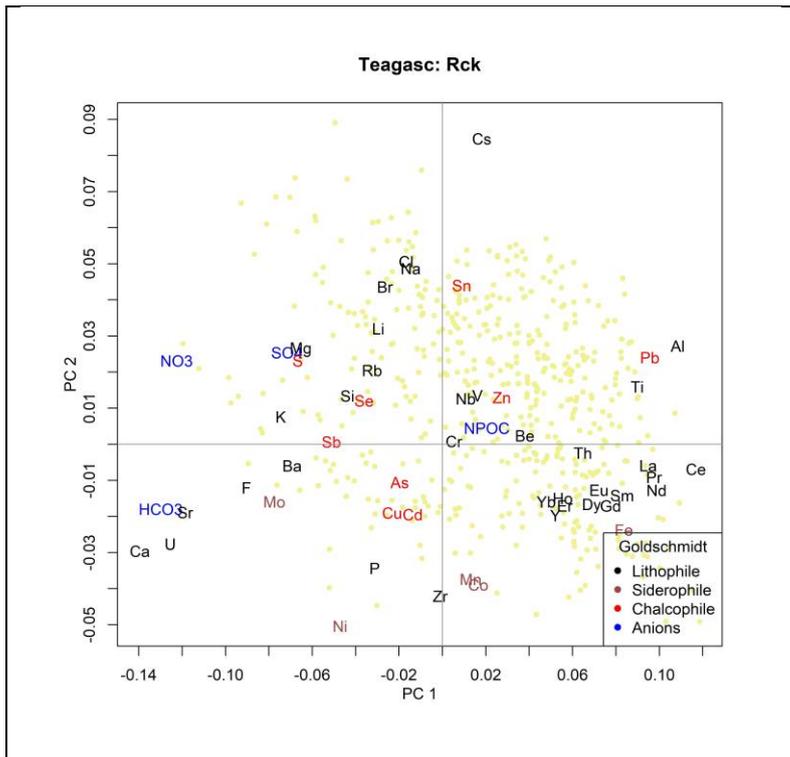


Figure A.41 PCA1 v PCA2 biplot for Teagasc subsoil class Rock

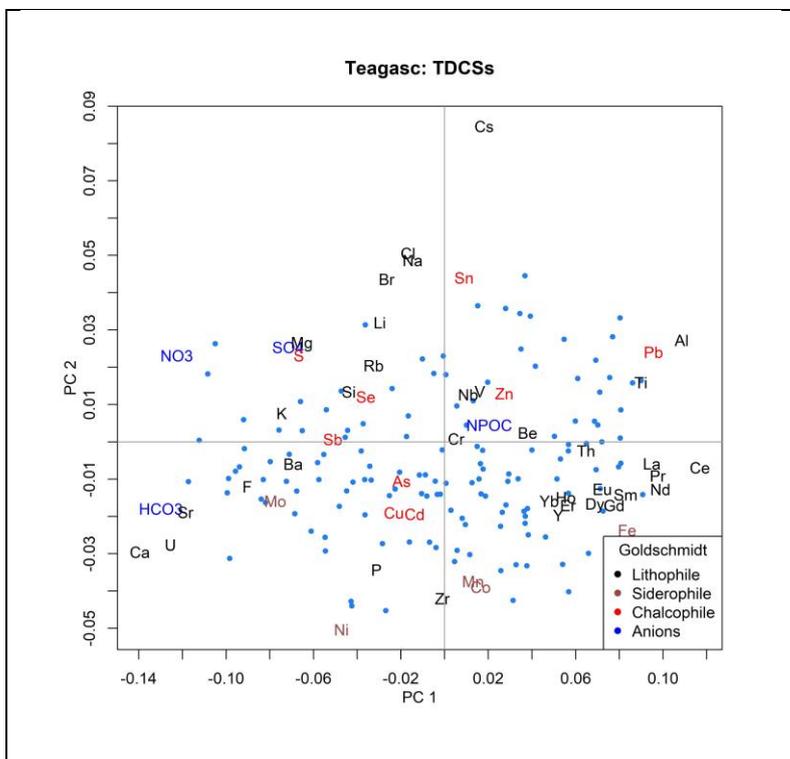


Figure A.42 PCA1 v PCA2 biplot for Teagasc subsoil class Till (Devonian-Carboniferous sandstone)



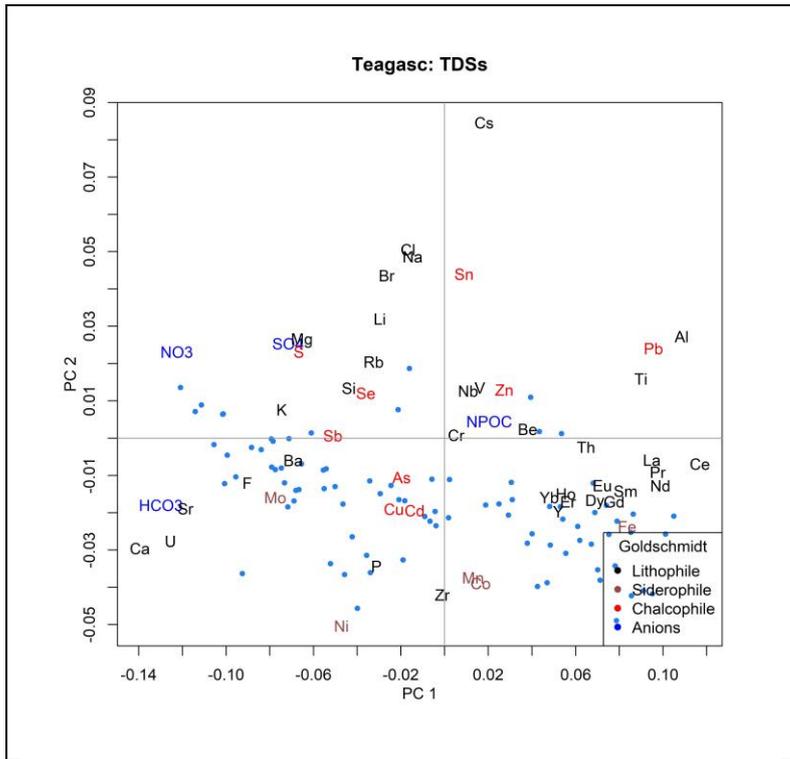


Figure A.43 PCA1 v PCA2 biplot for Teagasc subsoil class Till (Devonian sandstone)

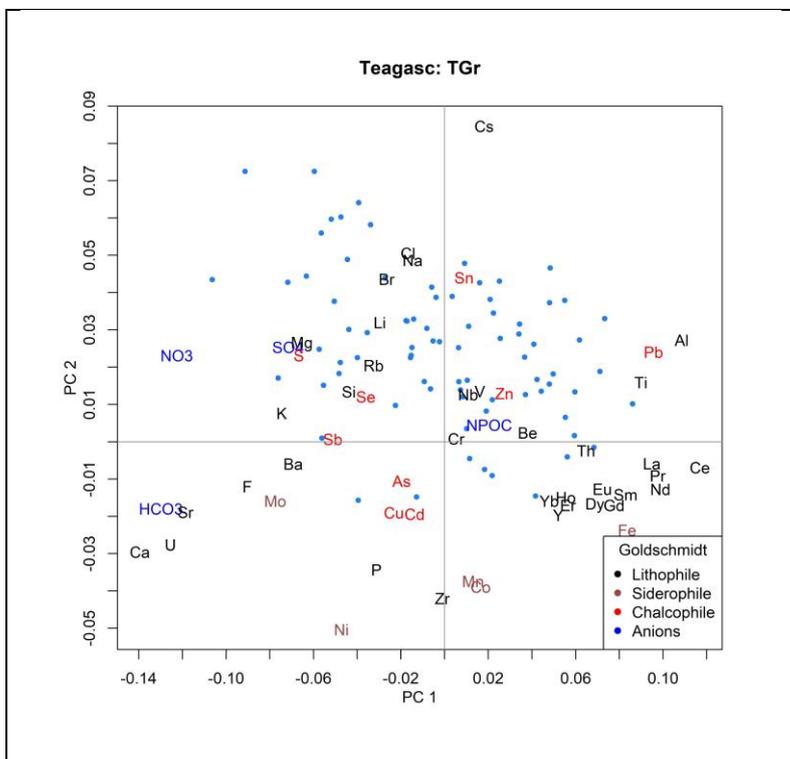


Figure A.44 PCA1 v PCA2 biplot for Teagasc subsoil class Till (granite)



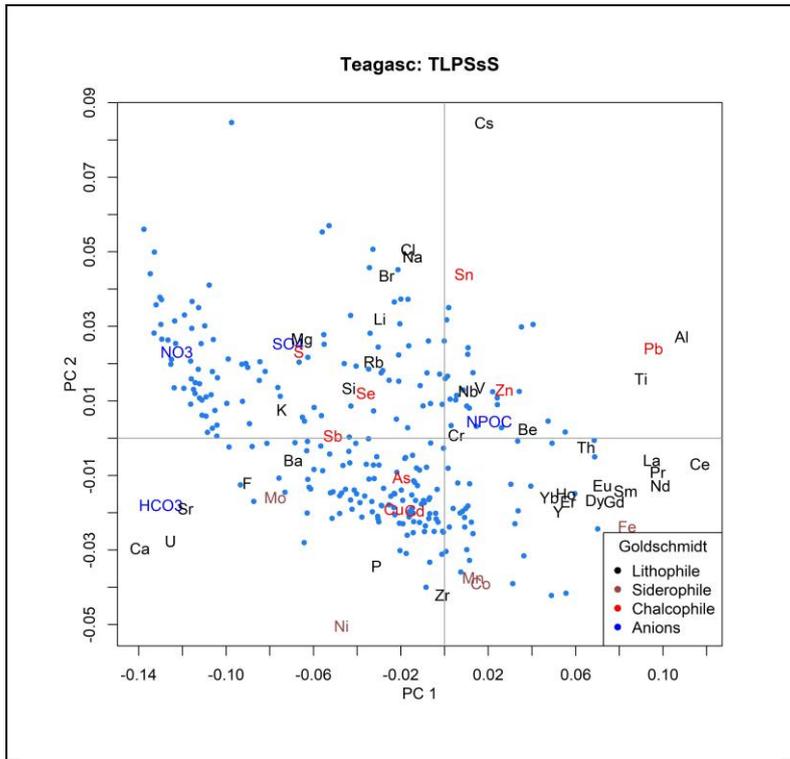


Figure A.45 PCA1 v PCA2 biplot for Teagasc subsoil class Till (Lower Palaeozoic sandstone, shale)

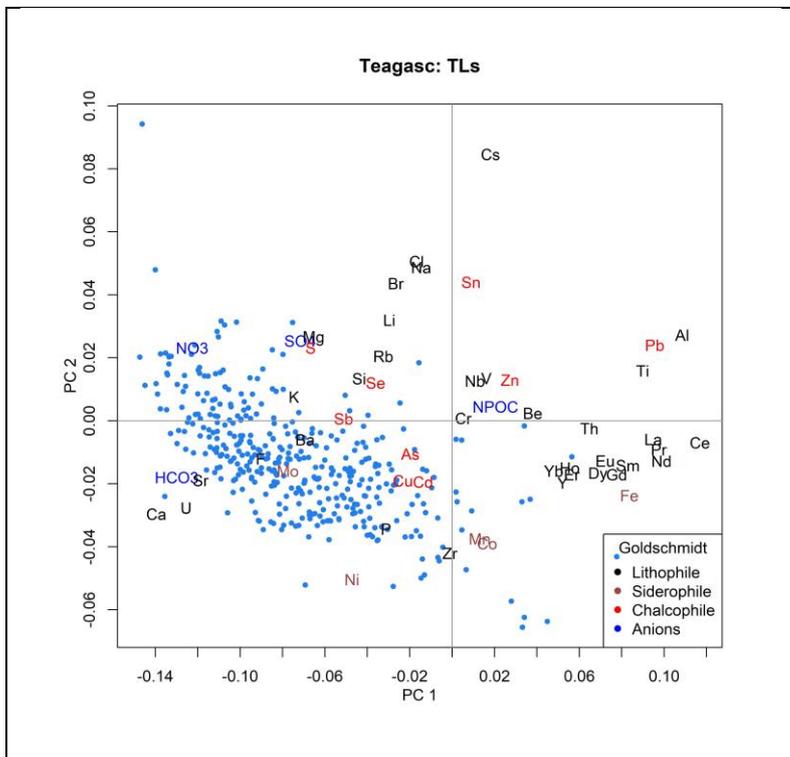


Figure A.46 PCA1 v PCA2 biplot for Teagasc subsoil class Till (limestone)



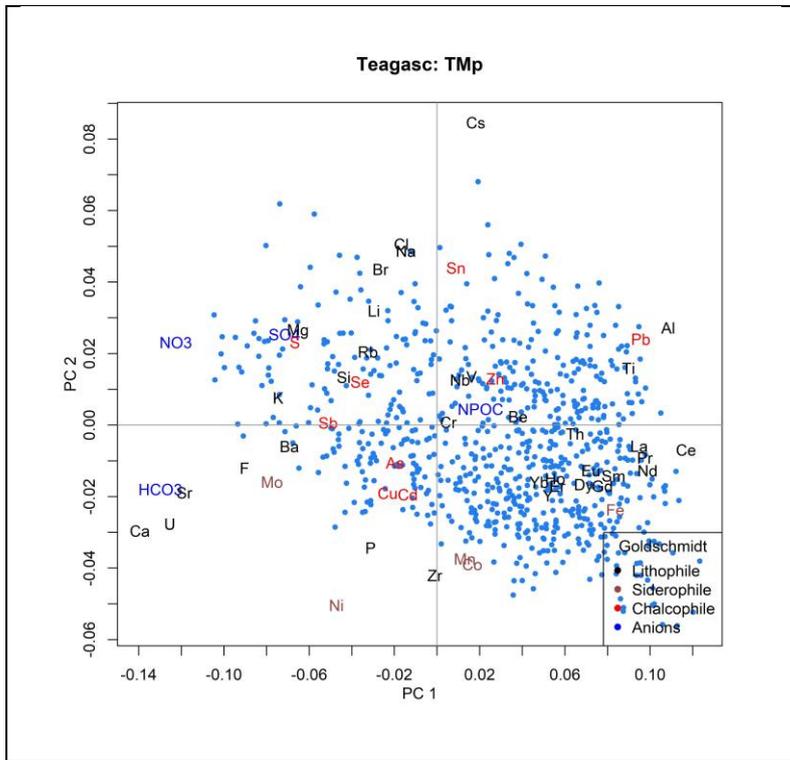


Figure A.47 PCA1 v PCA2 biplot for Teagasc subsoil class Till (metamorphic rocks)

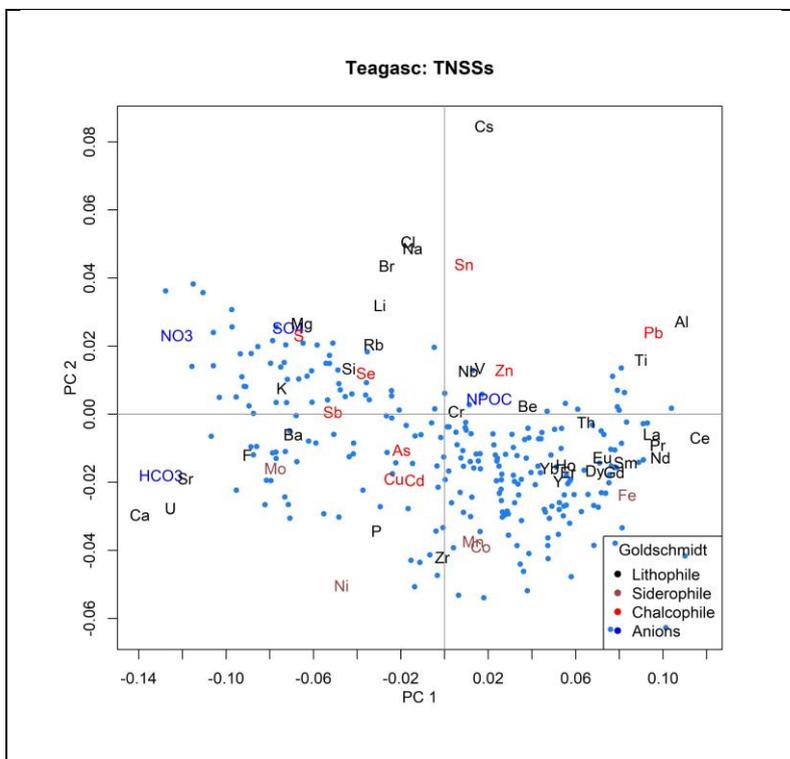


Figure A.48 PCA1 v PCA2 biplot for Teagasc subsoil class Till (Namurian shale, sandstone)



A.3 Principal Components Associated with SRF Domains

Figure A.49 shows the distribution of the SRF Geochemical Domains in the study area.

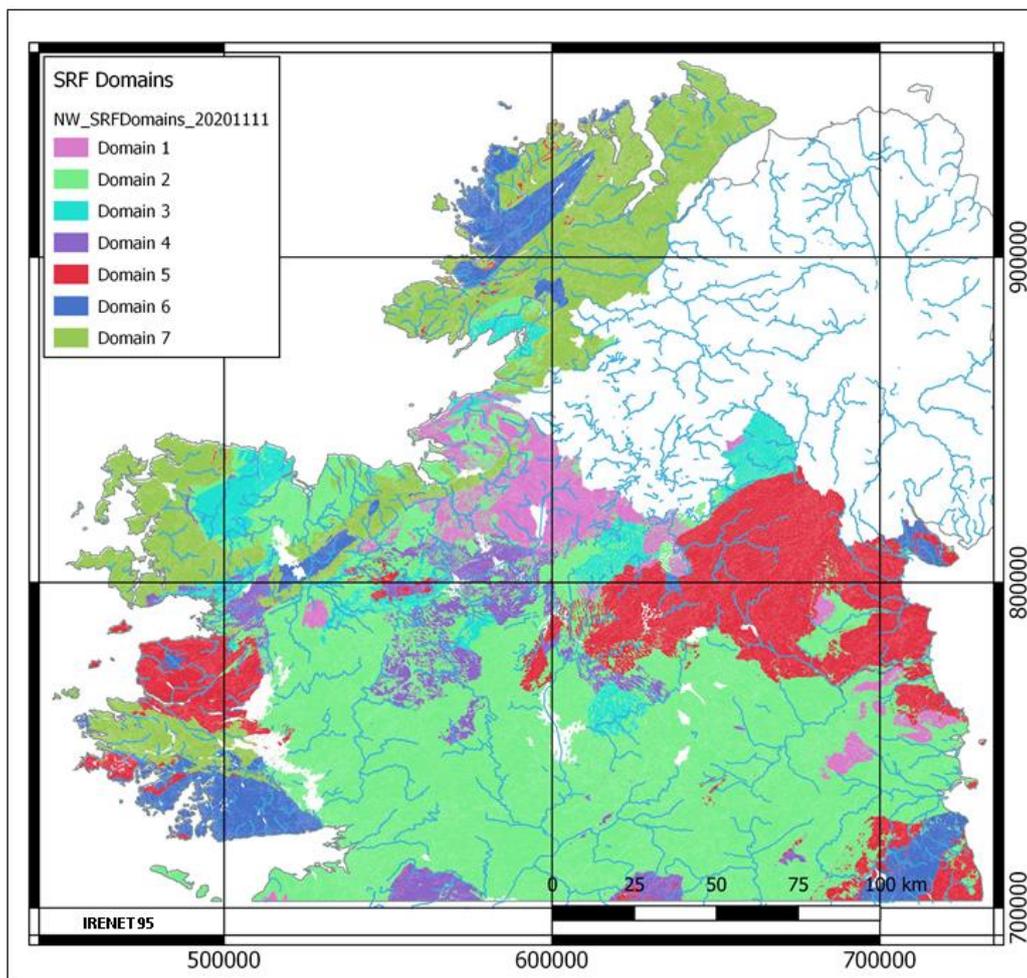


Figure A.49 SRF Domains

Figures A.50 to A.56 show the PC biplots for PC1-PC2, PC2-PC3, PC3-PC4, PC4-PC5, PC5-PC6, PC6-PC7 and PC7-PC8, respectively, for the SRF Geochemical Domains. As described above, the first PC accounts for 50 % of the geochemical variability and the first two components account for > 58 % of the variability. The biplot of PC1 v PC2 (Figure A.52) shows a distinctive pattern, which separates silicate-bearing assemblages (positive PC1 scores) from carbonate-bearing assemblages (negative PC2 scores). The carbonate-bearing assemblages also show an association with the anions NO_3^- , SO_4^{2-} , HCO_3^- and NPOC that may be associated with agricultural activities. The principal component biplots of Figures A.51 to A.56 show minor variations that are associated with distinctive processes as highlighted in the principal component maps of Figures A.23 to A.30, and discussed in preceding sections.



Figures A.57 to A.63 show the PC 1 v PC 2 biplots for stream waters data classified for each SRF Geochemical Domain. Stream waters draining SRF Domain 2 (essentially comprising Carboniferous limestone and related rocks and subsoil derived from them) unsurprisingly have a similar distribution on the biplot (Figure A.58) to those draining limestone-derived till and Carboniferous limestone bedrock (Figure A.15, A.46). However, their distribution also overlaps the area of the biplot associated with silicate-bearing assemblages, reflecting clastic material also present in the domain. Stream waters draining SRF Geochemical Domain 6 (granitic rocks and till derived from them) (Figure A.64) shows a predominant association with silicate-bearing assemblages. The other SRF Domain biplots (Figures A.57, A.59, A.60, A.61 and A.63) reflect a mix of the carbonate-bearing and silicate-bearing assemblages that indicate that these Domains cover both the Phanerozoic sedimentary assemblages and the Proterozoic crystalline and metamorphic assemblages.

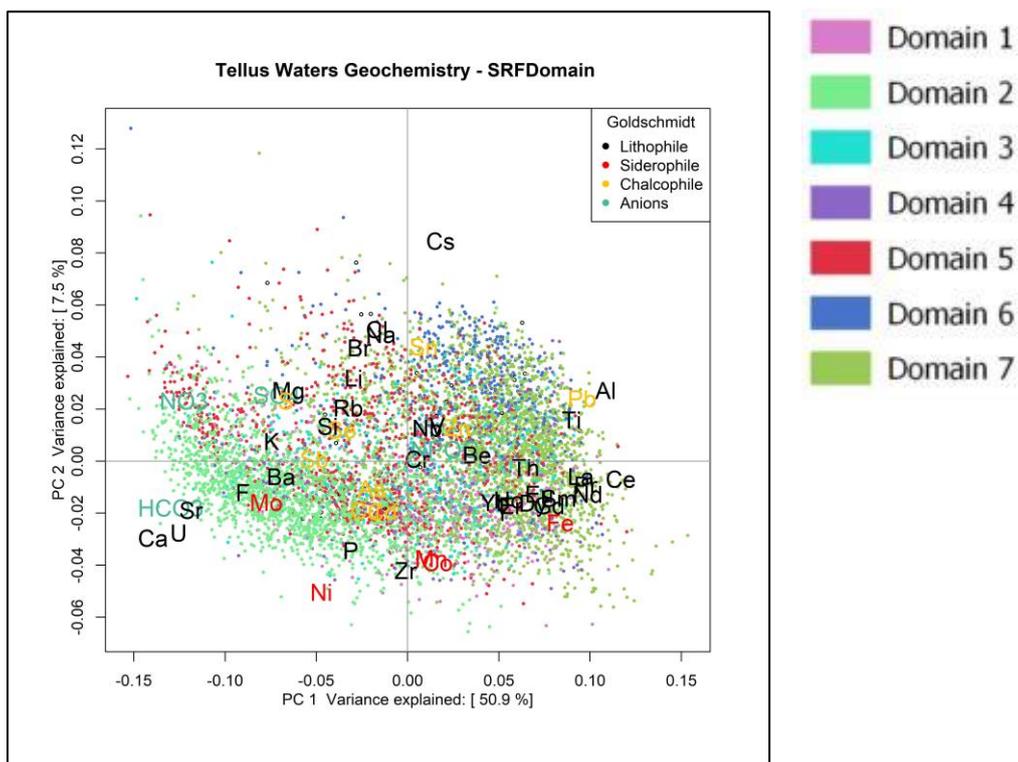


Figure A.50 PC1 v PC2 for SRF Domains (with legend)



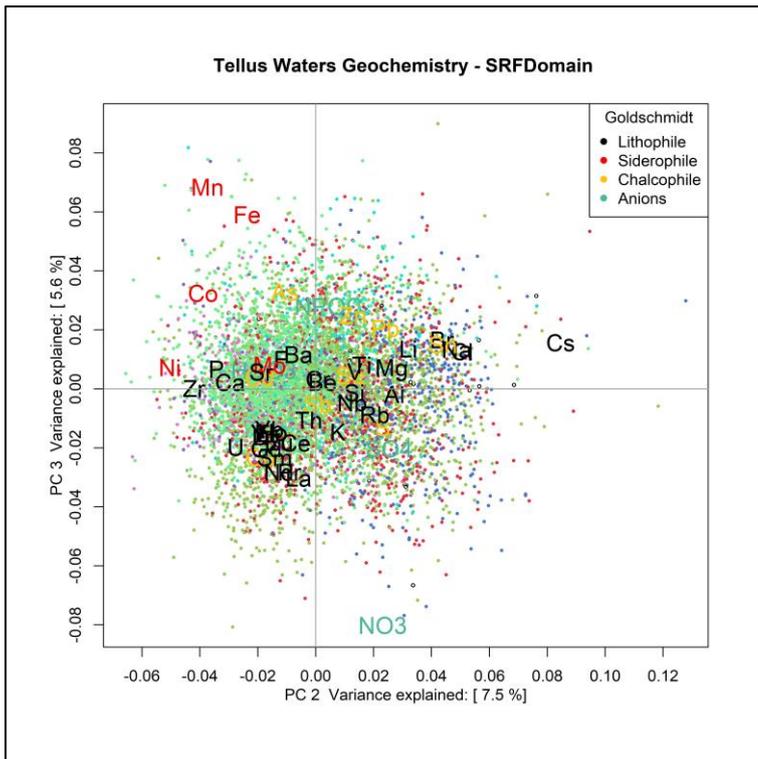


Figure A.51 PC2 v PC3 for SRF Domains (legend as for Figure A.51)

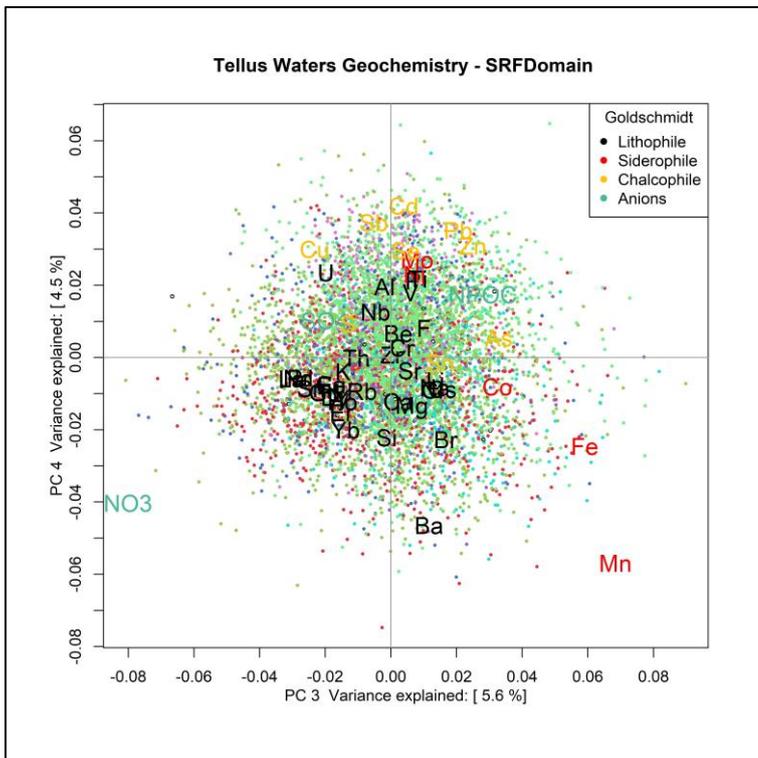


Figure A.52 PC3 v PC4 for SRF Domains (legend as for Figure A.51)



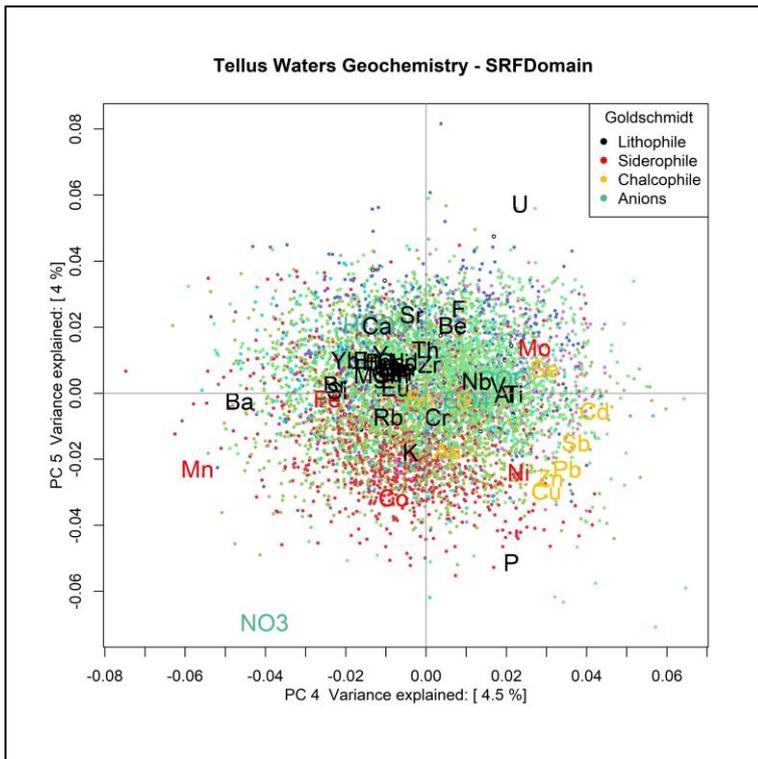


Figure A.53 PC4 v PC5 for SRF Domains (legend as for Figure A.51)

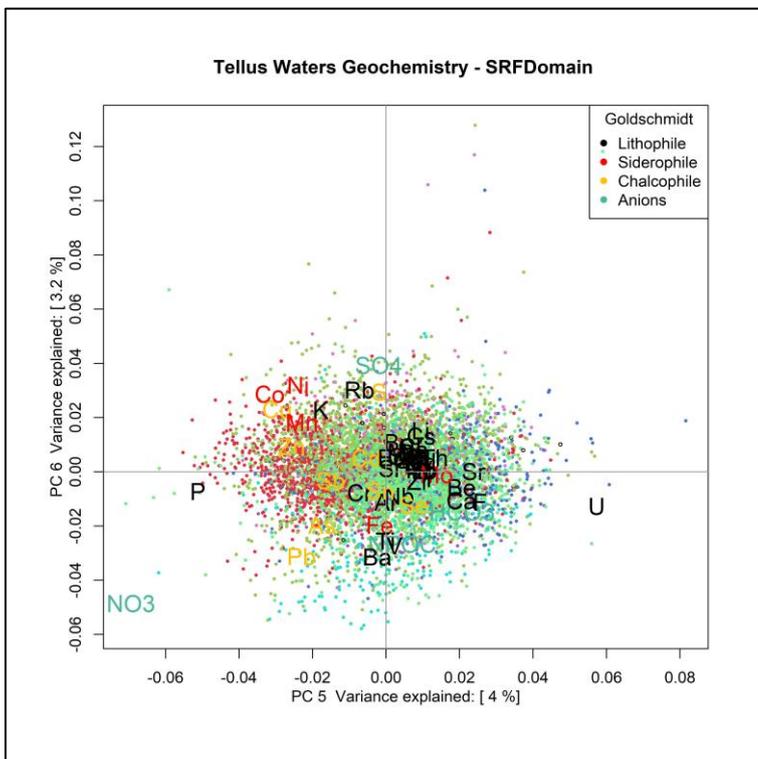


Figure A.54 PC5 v PC6 for SRF Domains (legend as for Figure A.51)



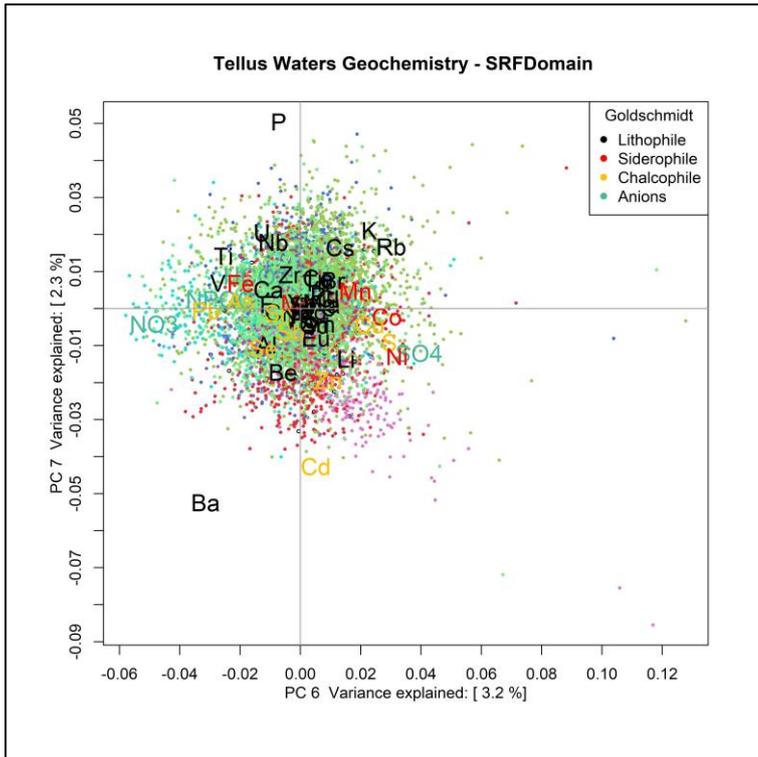


Figure A.55 PC6 v PC7 for SRF Domains (legend as for Figure A.51)

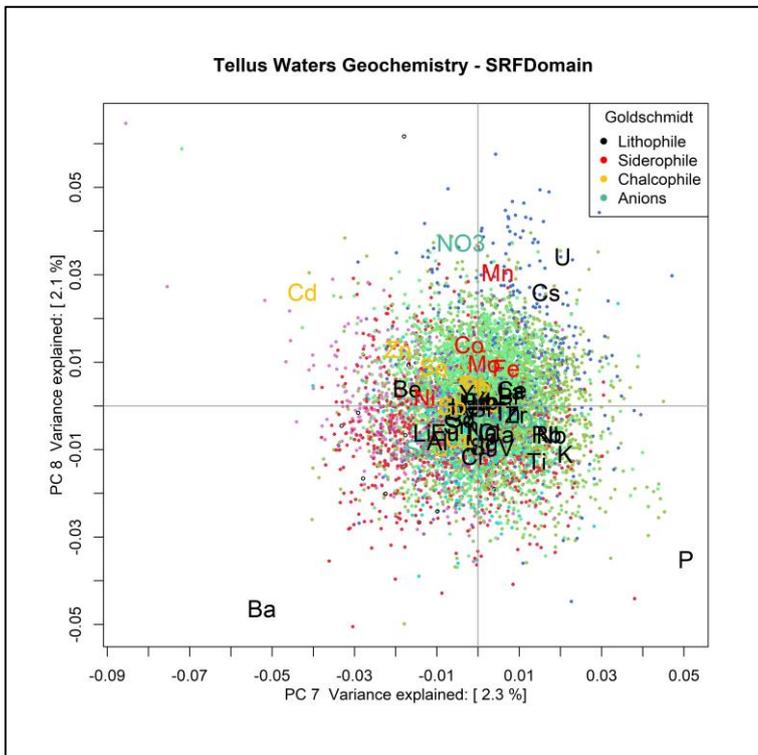


Figure A.56 PC7 v PC8 for SRF Domains (legend as for Figure A.51)



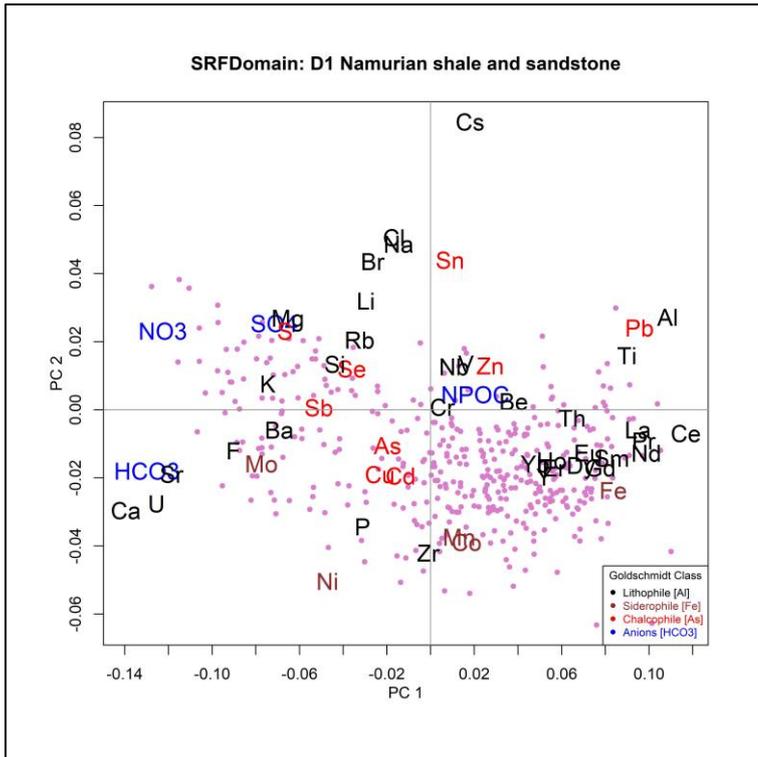


Figure A.57 PC1 v PC2 for SRF Domain 1

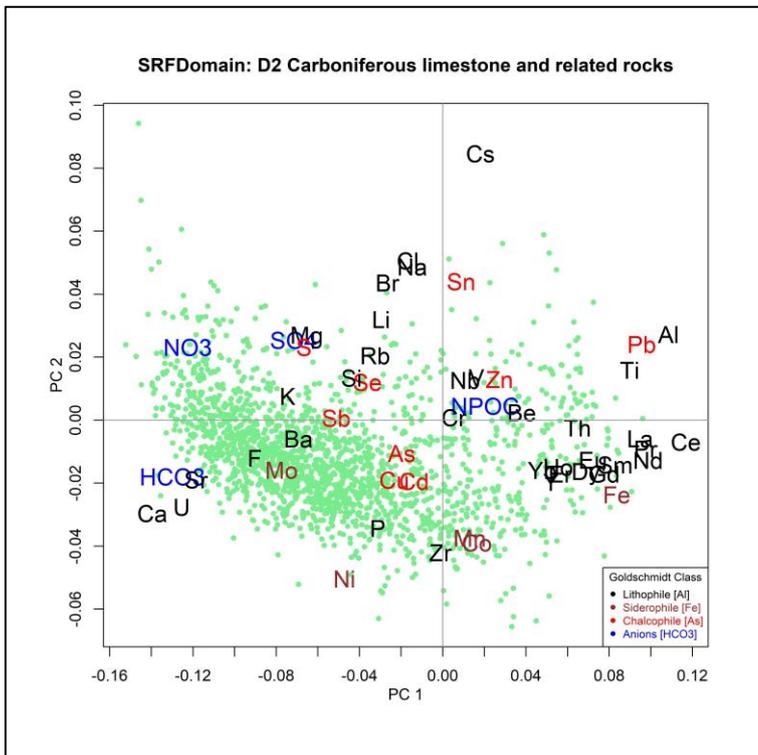


Figure A.58 PC1 v PC2 for SRF Domain 2



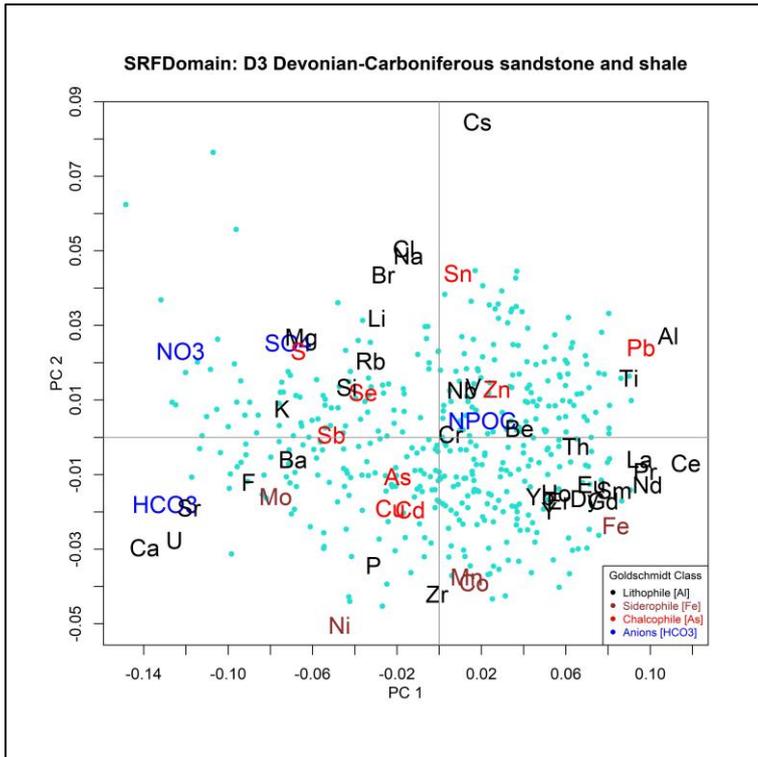


Figure A.59 PC1 v PC2 for SRF Domain 3

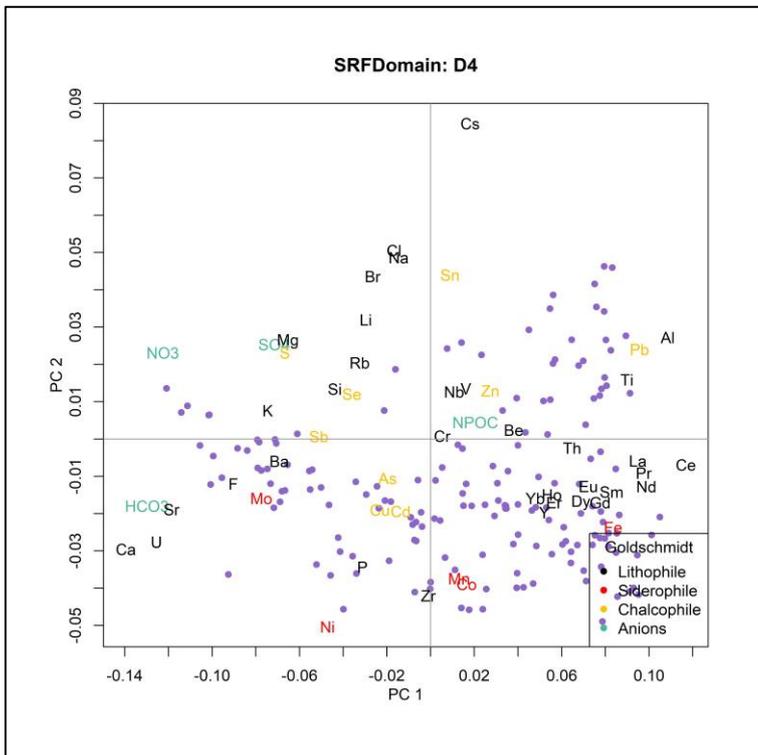


Figure A.60 PC1 v PC2 for SRF Domain 4



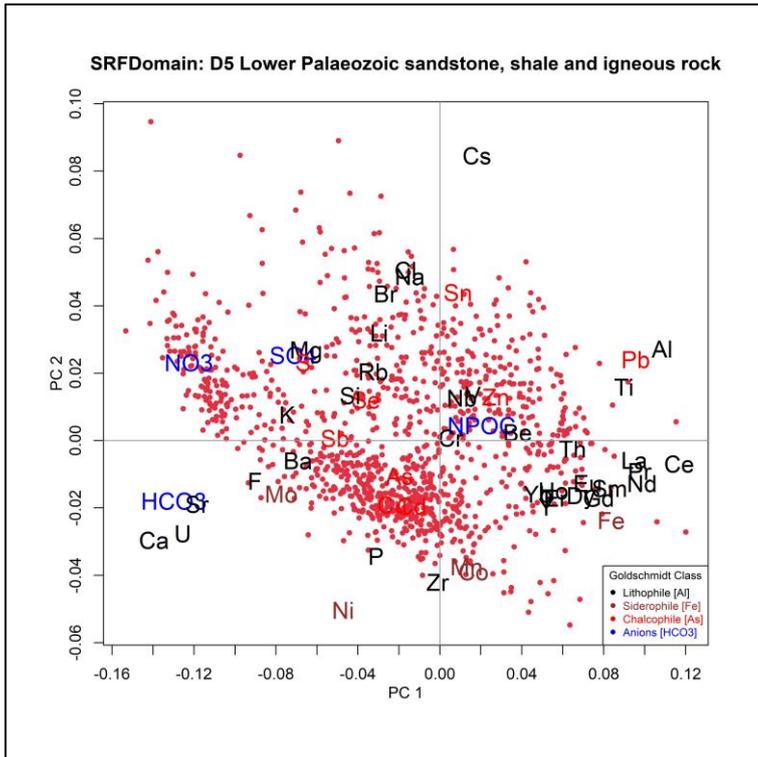


Figure A.61 PC1 v PC2 for SRF Domain 5

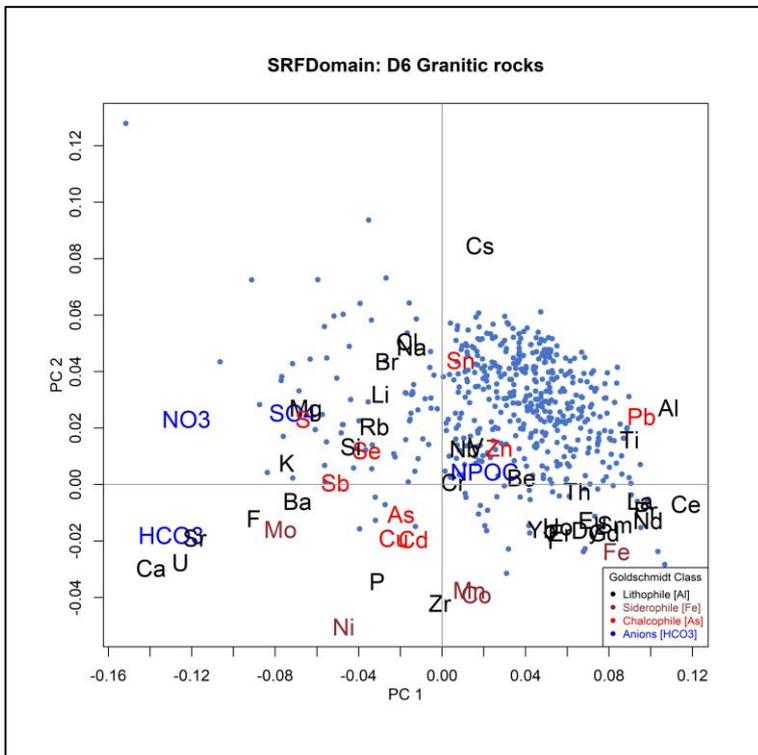


Figure A.62 PC1 v PC2 for SRF Domain 6



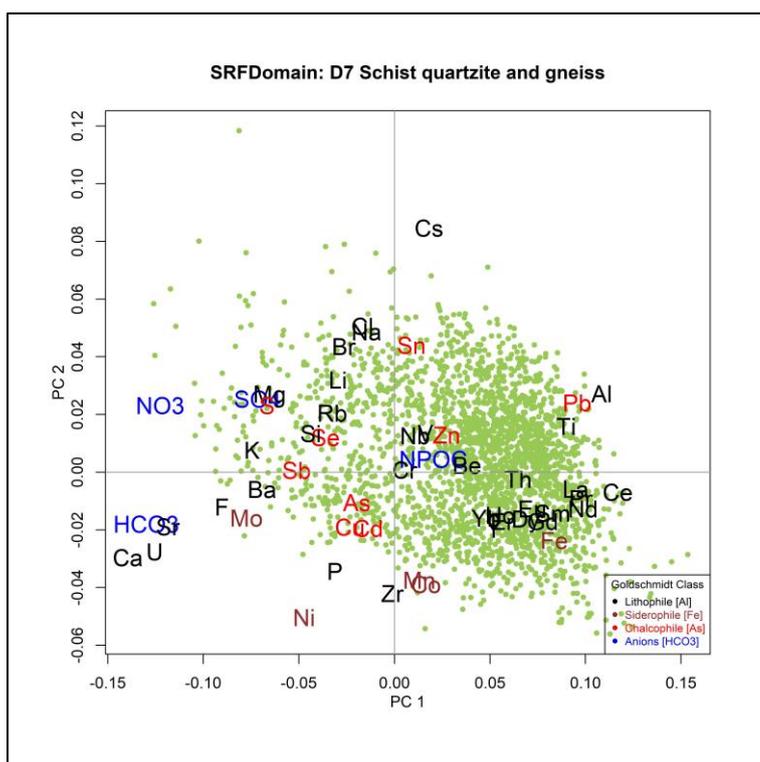


Figure A.63 PC1 v PC2 for SRF Domain 7

A.4 Principal Component Associated with Land Cover

Figures A64 to A70 show the PC biplots for PC1-PC2, PC2-PC3, PC3-PC4, PC4-PC5, PC5-PC6, PC6-PC7 and PC7-PC8, respectively, for stream water data classified according to some Corine Land Cover classes. As described previously, the first PC accounts for 50 % of the geochemical variability and the first two components account for > 58 % of the variability.

The land cover associations observed in the PC1 v PC2 biplot (Figure A.64) are dominated by a distinction between pastures and peat bog. This distribution strongly resembles that observed in biplots for the Rock Types, Teagasc subsoil and the SRF Domains, with the distribution of data for stream water draining pastureland cover largely coincident with that for Carboniferous limestone bedrock and related subsoil. The distribution of data for stream water draining the peat bog land cover overlaps that draining granitic rock and related subsoil. There is a general transition between carbonate-dominated assemblages and assemblages associated with silicate minerals but the stream water draining the peat bog land cover class has a distinct absence of elements related to carbonate rocks as well as anions that are typically associated with agricultural activities. The biplot of PC2-PC3 (Figure A.65) also highlights a relative enrichment trend in chalcophile elements that may represent element adsorption in the organic rich peat. These patterns are reflected in the principal component maps of PC1 and PC2 (Figures A.23 and A.24). The remaining PCs reflect processes that are not recognized, are under sampled or random. Positive PC3 values (Figure A.72) occur widely in the western and northwestern parts of the area (Figure A.25) and have a particular but not exclusive spatial association with peat. The high PC3 areas may reflect the occurrence of Fe-Mn oxyhydroxide coatings on mineral grains. With the exception of pasture (Figure A.75) and peat bog (Figure A.76),



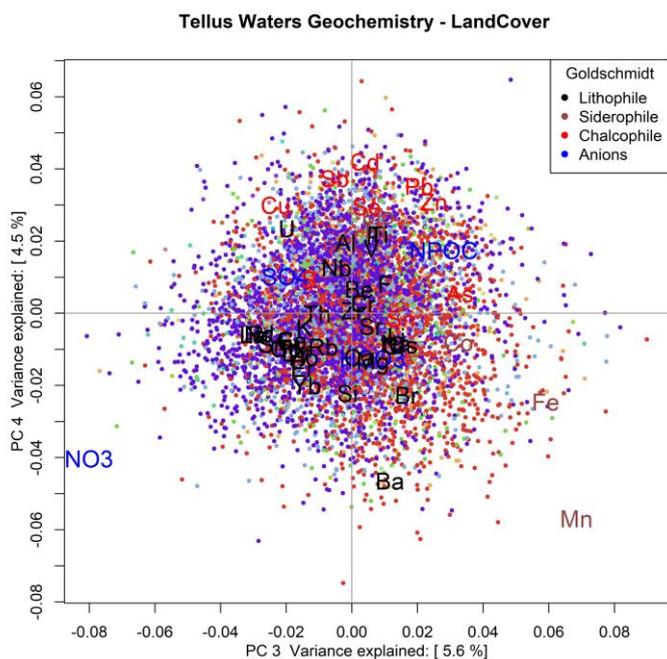


Figure A.66 PC3 v PC4 for Land Cover (Corine data) (Legend as for Figure A.64)

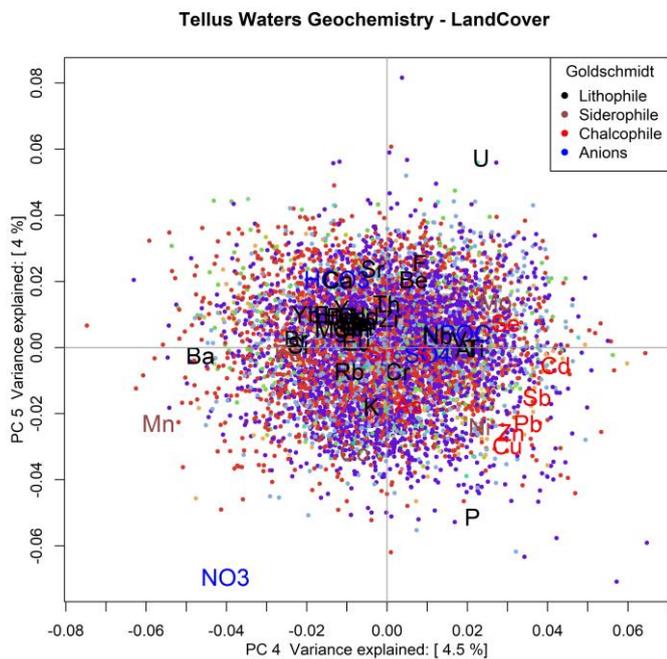


Figure A.67 PC4 v PC5 for Land Cover (Corine data) (Legend as for Figure A.64)



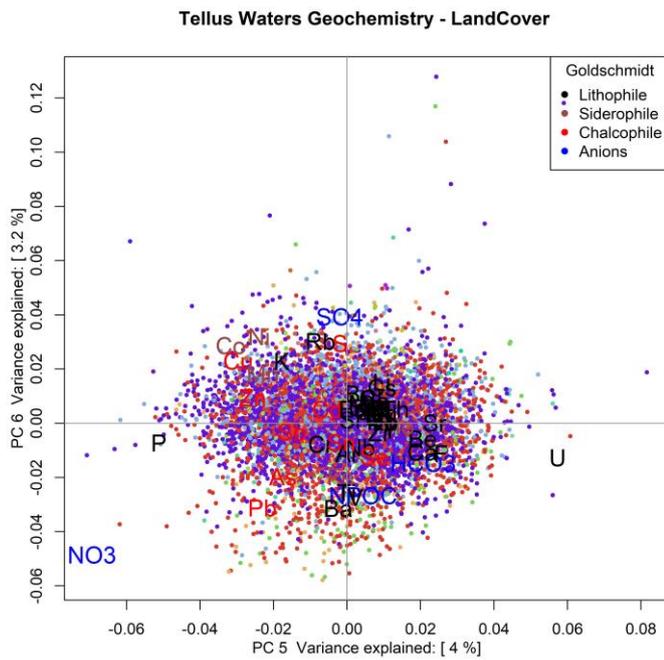


Figure A.68 PC5 v PC6 for Land Cover (Corine data) (Legend as for Figure A.64)

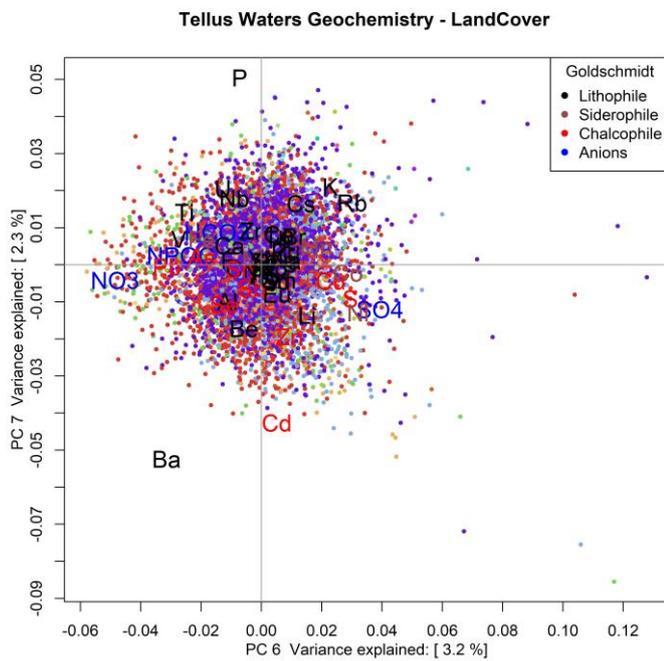


Figure A.69 PC6 v PC7 for Land Cover (Corine data) (Legend as for Figure A.64)



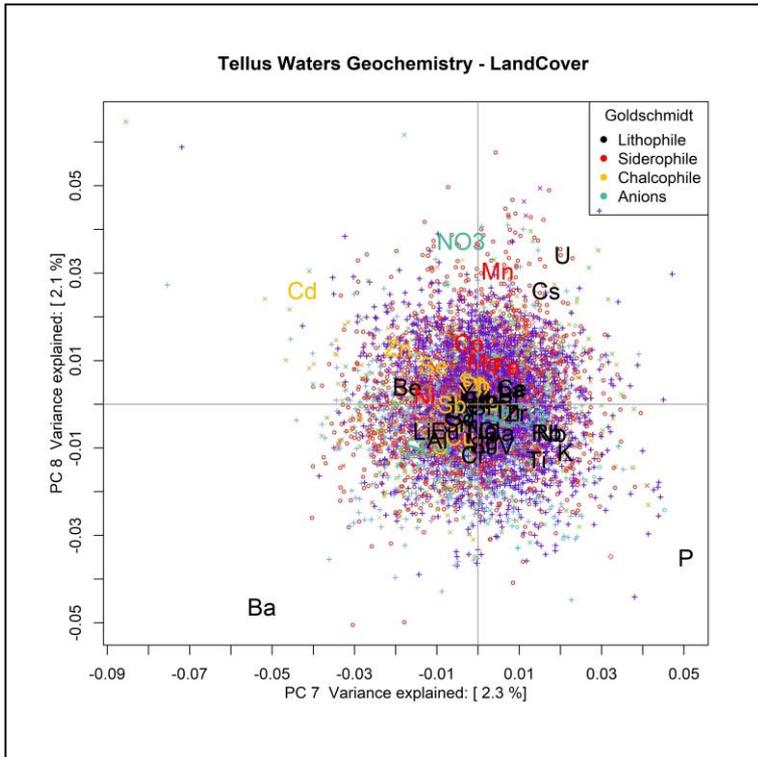


Figure A.70 PC7 v PC8 for Land Cover (Corine data) (legend as for Figure A.66)

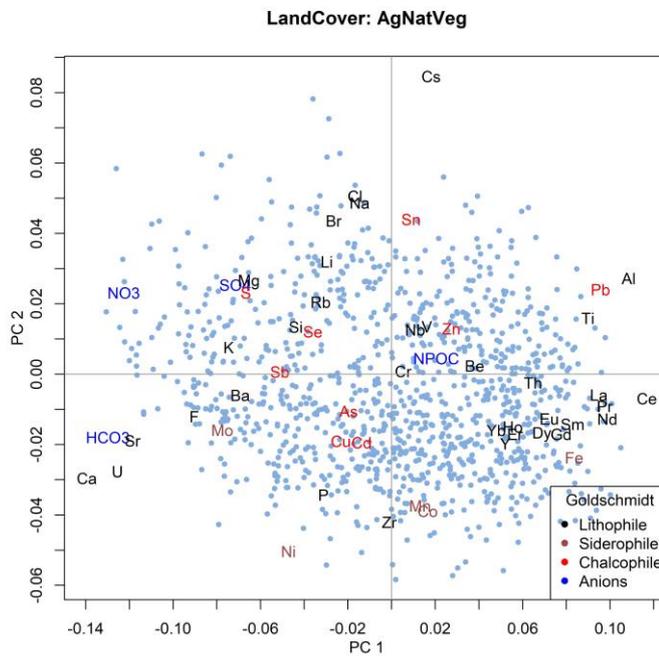


Figure A.71 PC1 v PC2 for Land Cover class Agriculture and Natural vegetation



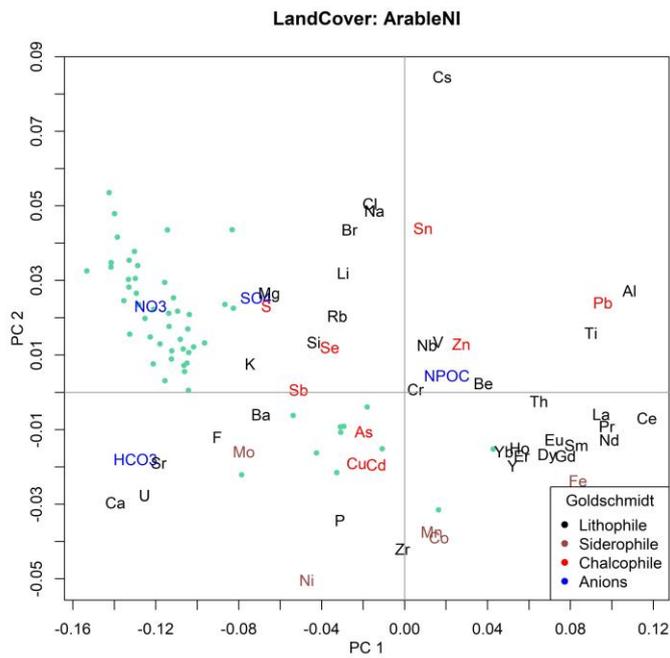


Figure A.72 PC1 v PC2 for Land Cover class Arable land

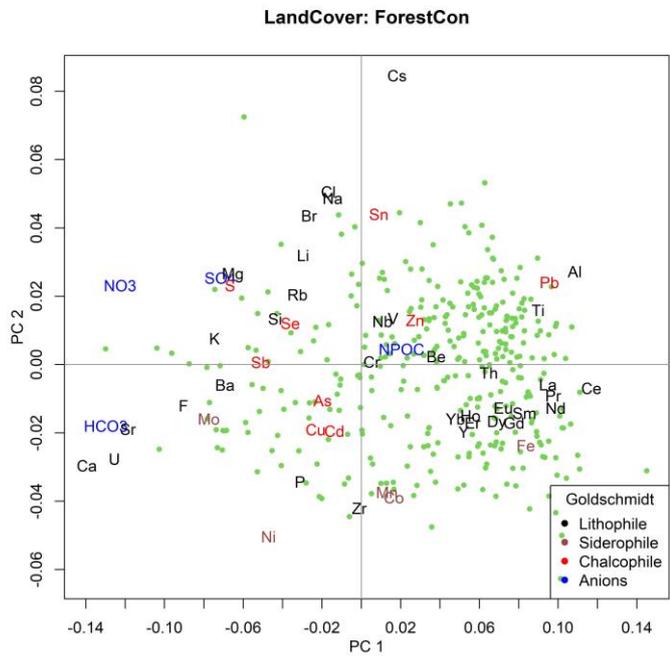


Figure A.73 PC1 v PC2 for Land Cover class Coniferous Forest



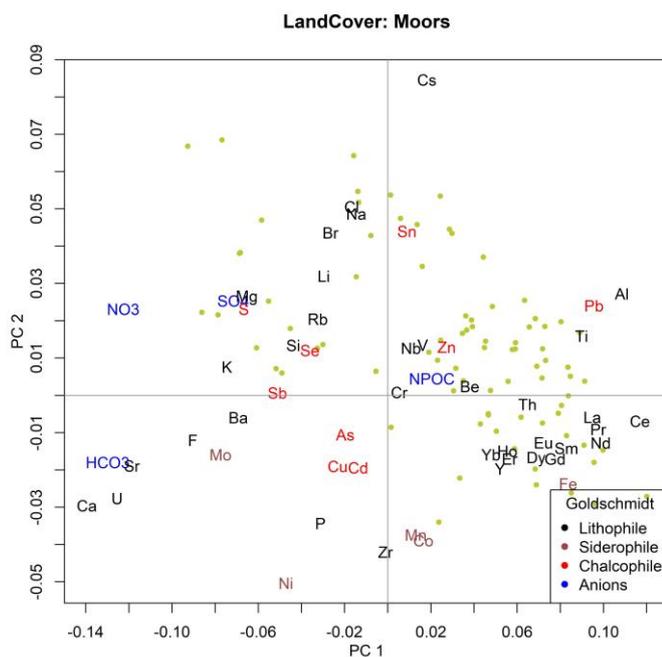


Figure A.74 PC1 v PC2 for Land Cover class Moorland

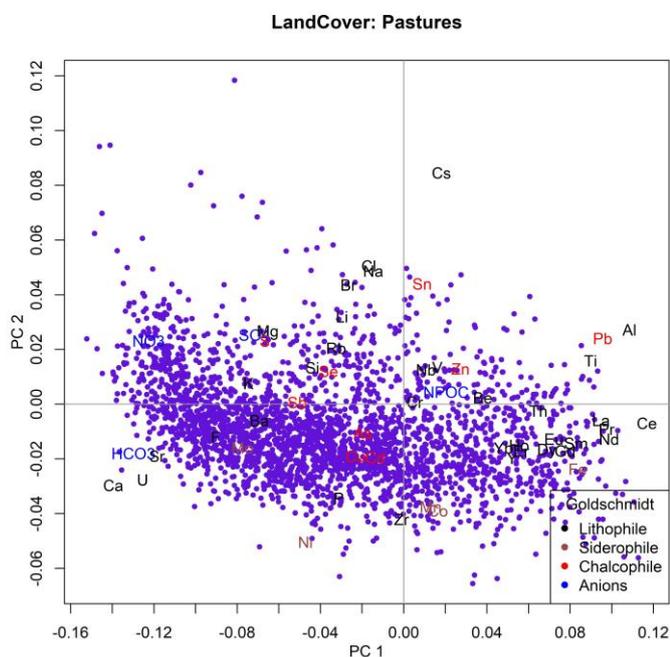


Figure A.75 PC1 v PC2 for Land Cover class Pasture



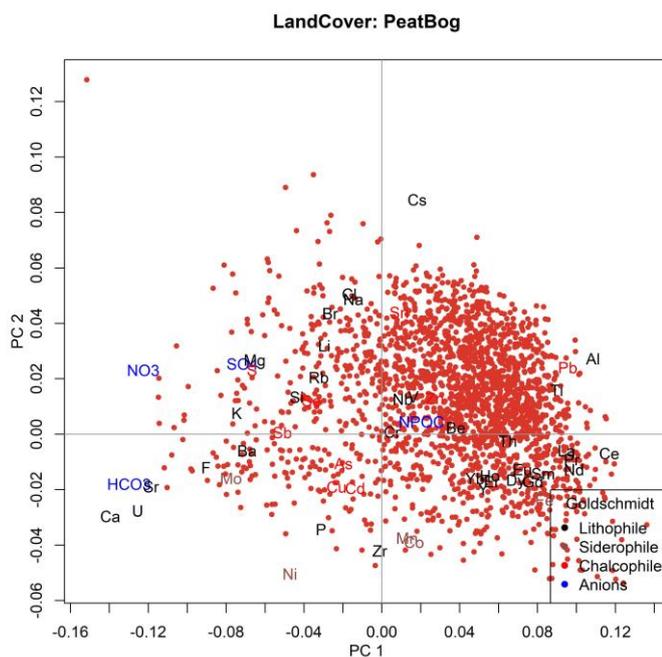


Figure A.76 PC1 v PC2 for Land Cover class Peat bog

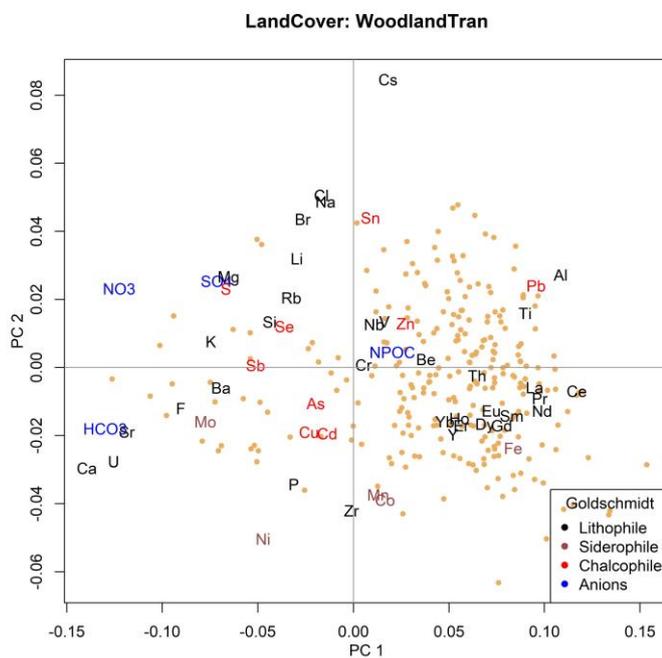


Figure A.77 PC1 v PC2 for Land Cover class Transitional Woodland



Appendix B: Results of Cluster Analysis

Hierarchical clustering is based on the linking of variables (R-mode) or observations (Q-mode) through measures of similarity. The relationships between the variables or observations can be graphically expressed using a dendrogram. Individual clusters can be discriminated by choosing an appropriate value of linkage, which separates internally similar groups of objects into dissimilar groups. Hierarchical clustering assumes that all variables are linked at some level, which may not be a reasonable assumption in some instances.

The correlation coefficient (R-mode) is the most common measure of similarity for clustering. For Q-mode analysis (similarities between the observations), the Euclidean distance can be used as a measure of proximity by which observations can be clustered. However, when the number of observations is large the computation becomes intractable. For the Waters EDA project, the total of 6835 analyses make the use of Q-mode cluster analysis too difficult to view and interpret.

Traditional linkage-based clustering can be carried out using the R package “stats” and the function “hclust”. The function hclust provides linkages and dendrograms based on several linkage criteria. The process of agglomeration can be based on the following methods: Ward’s minimum variance, single, complete, average, McQuitty, median and centroid linkages.

Ward's minimum variance method aims at finding compact, spherical clusters. The complete linkage method finds similar clusters. The single linkage method (which is closely related to the minimal spanning tree) adopts a ‘friends of friends’ clustering strategy. The other methods can be regarded as aiming for clusters with characteristics somewhere between the single and complete link methods. Note that methods "median" and "centroid" do not yield a monotone distance measure, or equivalently the resulting dendrograms can have so-called inversions, which are hard to interpret.

The choice of dendrogram for representing processes depends on the understanding of how the links are made and the relevance of the linking procedure with the processes of interest. The selection of the appropriate model is challenging if a model of the process(es) is not known. Elements of interest (e.g. Zn, Cd, Cr, As, Ni, Sb, Pb) may show linkages associated with the dominant processes; not necessarily the processes of interest such as anthropogenic effects.

B.1 Hierarchical Cluster Analysis with Dendrograms

Hierarchical cluster analysis was applied to the waters geochemical data and seven different hierarchical clustering methods were used to create dendrograms. Only the results from the application of Ward’s method is presented here.

Dendrograms were produced for each of the classes in the Bedrock and Teagasc Subsoil Themes. Only a few of the dendrograms are presented here, notably the dendrograms produced by the application of the different measures of similarity/distance (Figures B.1



to B.6). The interpretation of the dendrograms requires some prior knowledge about how the elements may be linked through processes such as mineralogy, weathering, groundwater effects, mass transport and organic complexation/adsorption. The dendrograms display distinct groupings of the elements for the rock types amphibolite, pelite, limestone, greywacke (Figures B.1 to B.4.) and the Teagasc subsoils blanket peat and till derived from metamorphic rocks (Figures B.5 and B.6). The linkages/associations of the elements at the different levels can assist in understanding the processes that contribute to the element associations. The geochemistry of the waters may reflect the influence of both natural geogenic processes, such as mineralogy, but also the effects of anthropogenic activities such as industry and agriculture.

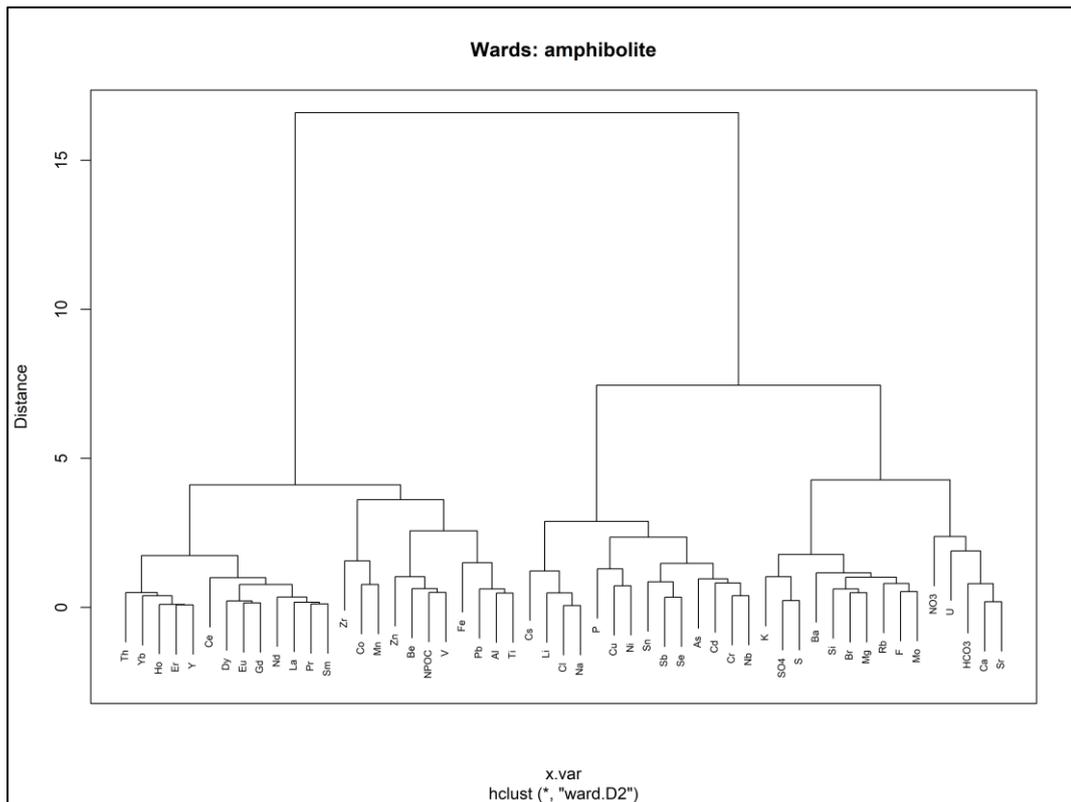


Figure B.1 Ward's Hierarchical clustering for waters classified by Bedrock = amphibolite

Figure B.1 shows a dendrogram for stream water data draining amphibolite according to the bedrock geology map. The first link at the left side of the dendrogram comprises two clusters, the first including rare earth elements that reflect a felsic association and the second including siderophile elements that indicate a mafic association. The links on the right side of the dendrogram have clusters of chalcophile, lithophile, anionic and alkalic elements that reflect a mixture of mineral and organic material.



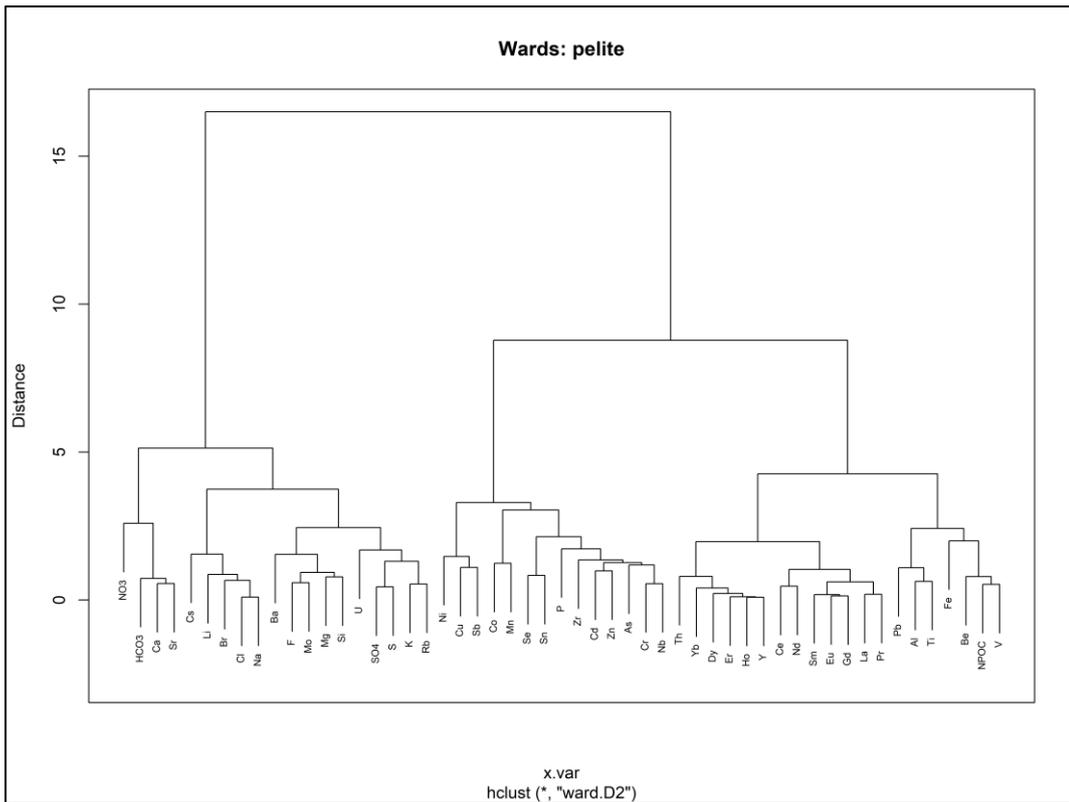


Figure B.2 Ward's Hierarchical clustering for waters classified by Bedrock = pelite.

Figure B.2 shows a dendrogram for stream water data draining pelite according to the bedrock geology map. The left side of the dendrogram shows clusters of alkalic and lithophile elements, possibly representing material of felsic origin. The right side of the dendrogram shows clusters of siderophile, rare earth elements and chalcophile elements that most likely represent the composition of shales and pelite.



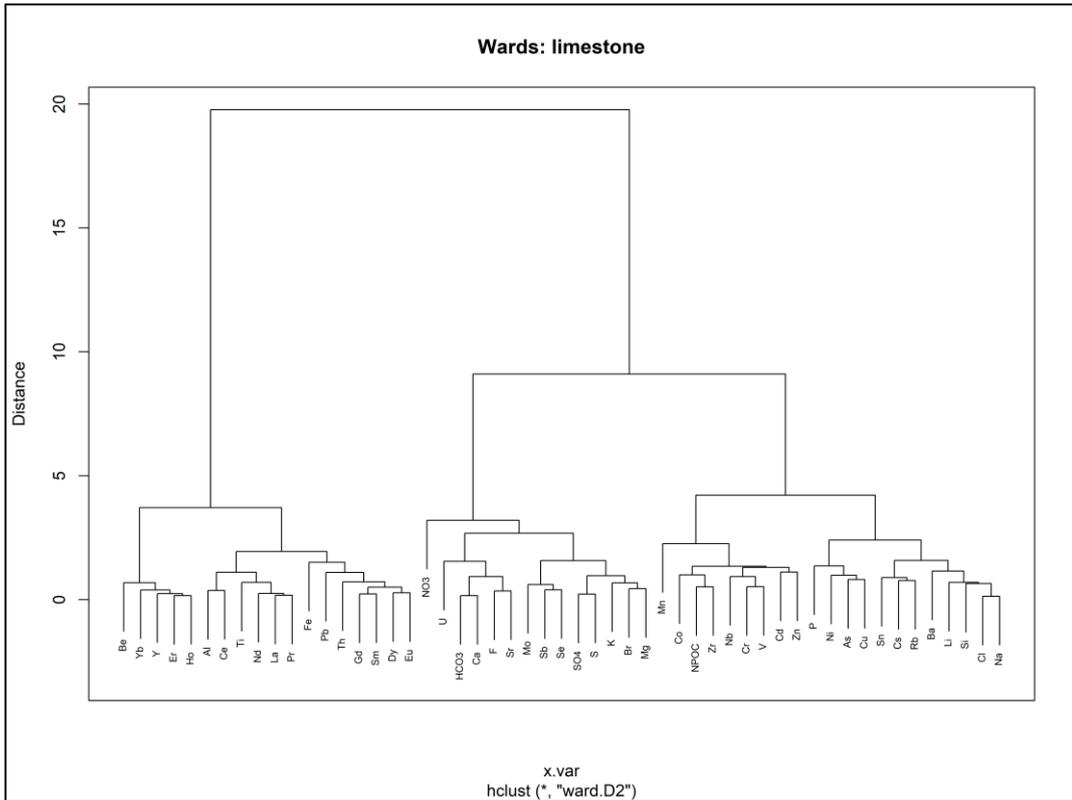


Figure B.3 Ward’s Hierarchical clustering for waters classified by Bedrock = limestone.

Figure B.3 shows a dendrogram for stream water data draining limestone according to the bedrock geology map. The dendrogram shows a distinct cluster of rare earth elements on the left side that is independent of the two other dominant clusters, one related to the composition of limestone and the other enriched in chalcophile elements, most likely associated with organic-rich subsoil or mineralization, notably the Zn-Cd pair.



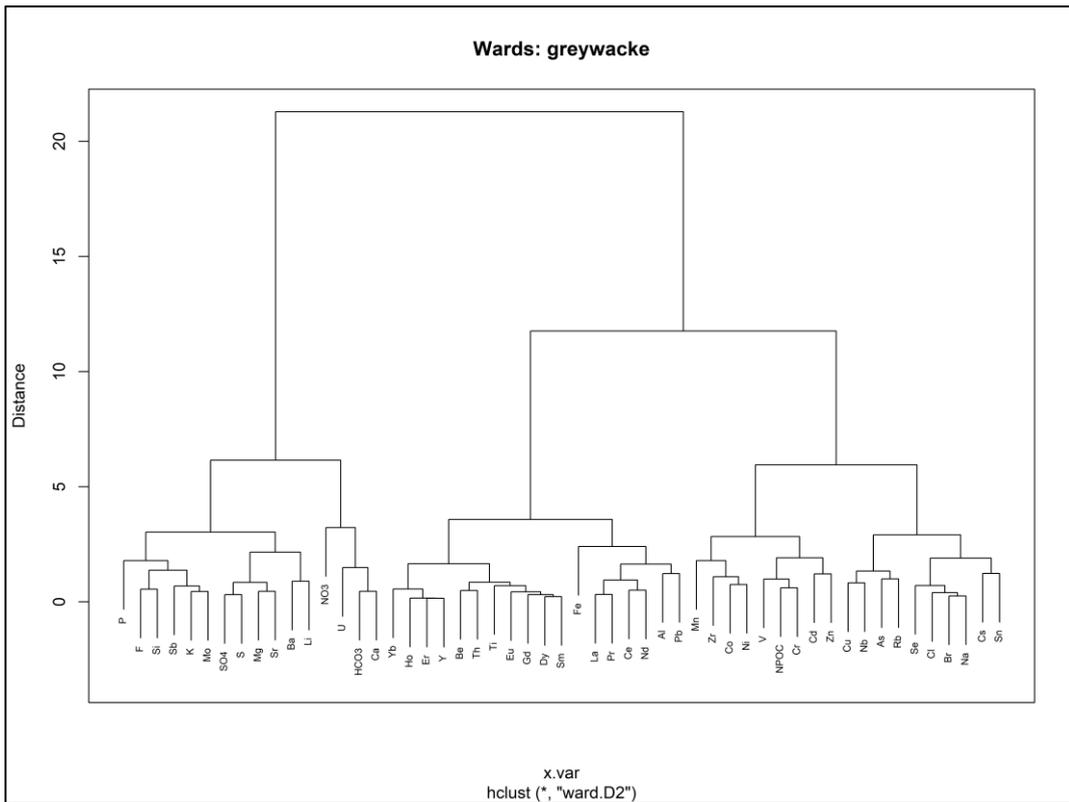


Figure B.4 Ward’s Hierarchical clustering for waters classified by Bedrock = greywacke.

Figure B.4 shows a dendrogram for stream water data draining greywacke bedrock according to the bedrock geology map. The clusters on the left of the dendrogram reflect an alkalic composition. Rare earth elements dominate the cluster in the middle. The cluster on the right includes lithophile, siderophile and chalophile elements. The Zn-Cd pair noted for the limestone bedrock cluster (Fig. B.3) is also present.



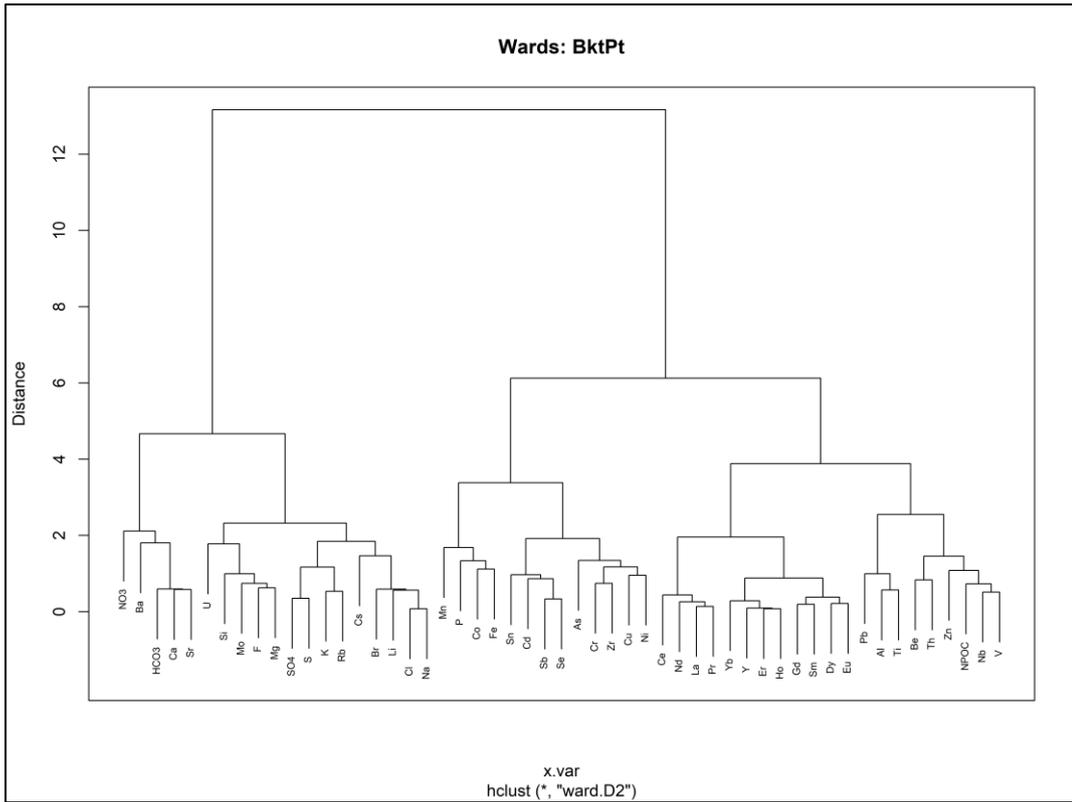


Figure B.5 Ward’s Hierarchical clustering for waters classified by Teagasc subsoil = Blanket Peat.

Figure B.5 shows a dendrogram for stream water data draining blanket peat, according to the Teagasc subsoil map. There are two dominant clusters in this dendrogram. The cluster on the left reflects an anionic and alkalic association, while the one on the right includes distinct groupings of siderophile/chalcophile, siderophile, rare earth elements and lithophile/chalcophile elements.



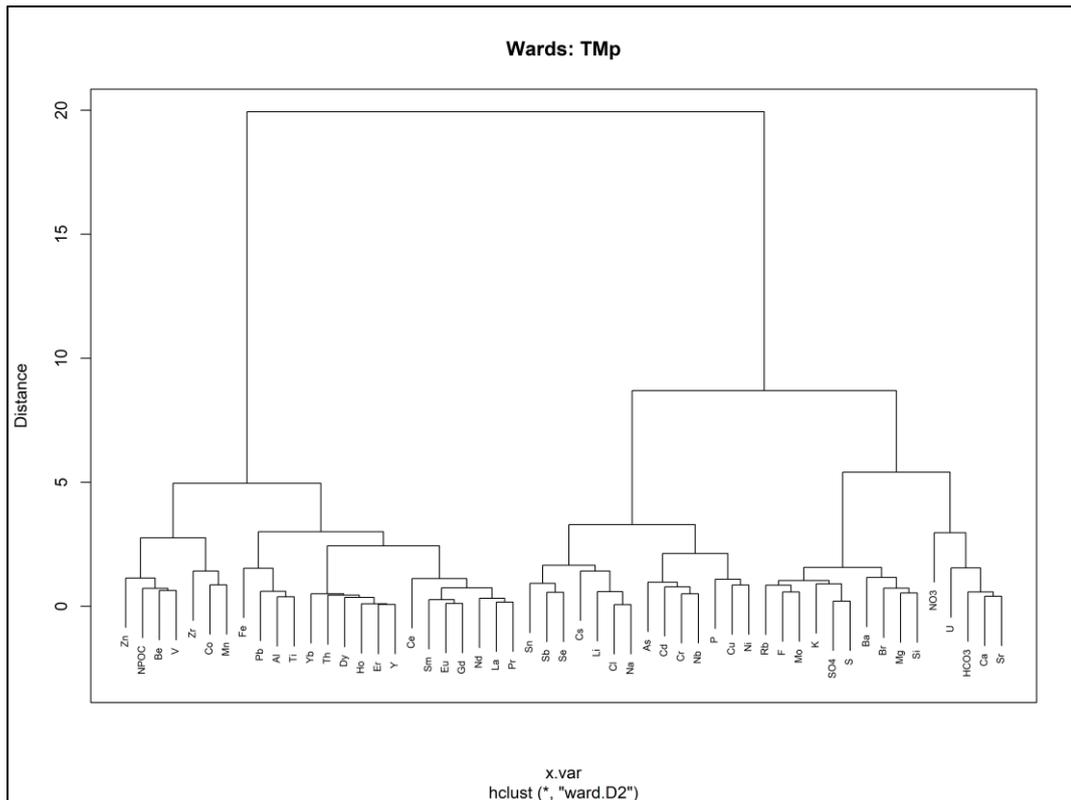


Figure B.6 Ward's Hierarchical clustering for waters classified by Teagasc subsoil = metamorphic till.

Figure B.6 shows a dendrogram for stream water data draining till derived from metamorphic rocks, according to the Teagasc subsoil map. The clusters on the left of the dendrogram show a mix of lithophile, rare earth element and siderophile elements. The clusters on the right highlight associations of chalcophile, siderophile, anionic and lithophile elements. The cluster on the extreme right has an association of elements/anions that are commonly found in agricultural fertilisers.



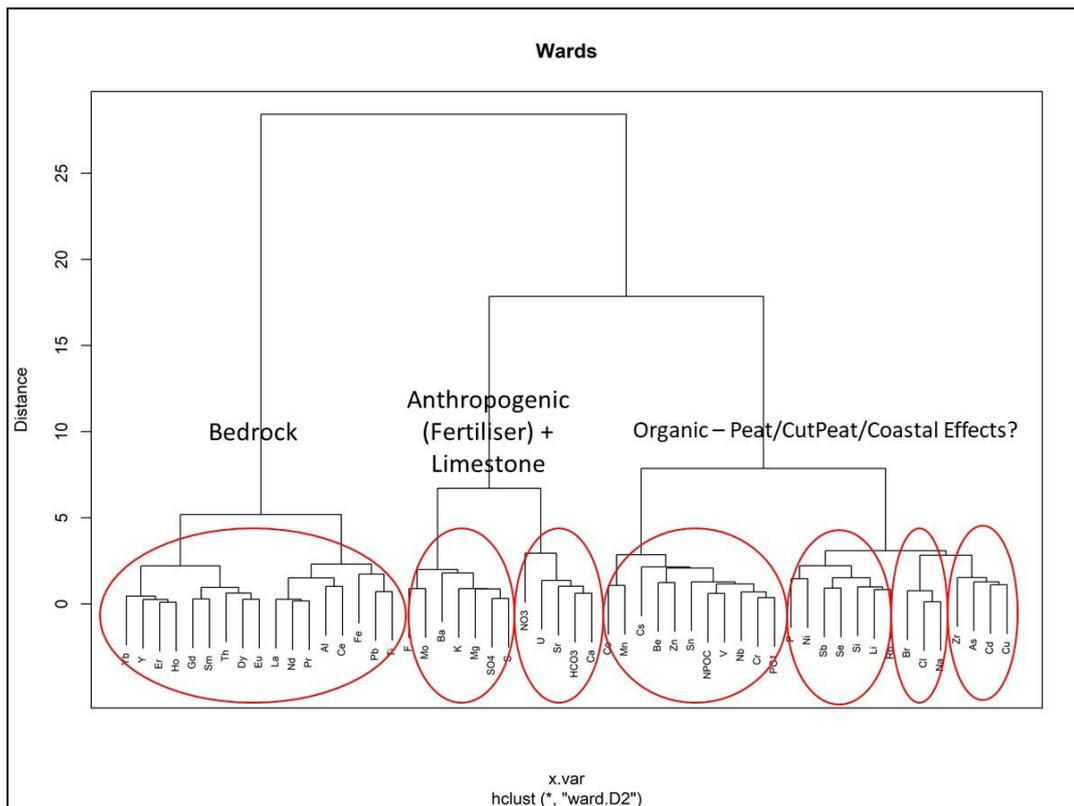


Figure B.7. Ward's hierarchical clustering for waters based on the logcentred (clr) transform.

Figure B.7 shows a dendrogram for all stream water data irrespective of bedrock or subsoil. The groupings of the element highlight some general patterns that may be interpreted as follows:

- 1) A multi-element signature related to bedrock is located on the left side of the dendrogram. This grouping contains several lithophile and rare earth elements that are associated with granitoid and metamorphic rock types along with mixtures of Phanerozoic sediments (greywacke, shale, sandstone).
- 2) A multi-element group that includes anions/nutrients associated with anthropogenic (agricultural) activities overlying limestone bedrock.
- 3) The right side of the dendrogram shows at least three groups that likely represent variations on peat and the underlying bedrock. A unique group that shows a Br-Na-Cl association likely reflects marine influence along the coast.

In summary, the dendrograms derived from hierarchical cluster analysis have the potential to identify specific processes but do not provide any geospatial association. The order of analyses can be identified based on the distance matrix determined from all of the analyses and this can be visualized geospatially but it is difficult to tie the observed correlations within and between clusters with known or speculative processes. However, creating geospatial maps of the q-mode order is technically challenging. It is more informative to use other methods of clustering to render geospatial associations.



B.2 Cluster Analysis using MClust

mclust is a contributed R package for model-based clustering, classification and density estimation based on finite normal mixture modelling. It provides functions for parameter estimation via the EM algorithm for normal mixture models with a variety of covariance structures, and functions for simulation from these models. Also included are functions that combine model-based hierarchical clustering, EM (Expectation-Maximum likelihood) for mixture estimation and the Bayesian Information Criterion (BIC) in comprehensive strategies for clustering, density estimation and discriminant analysis. Additional functionalities are available for displaying and visualizing fitted models along with clustering, classification, and density estimation results. The EM algorithm is used by **mclust** for maximum likelihood estimation. Initialisation of EM is performed using the partitions obtained from agglomerative hierarchical clustering.

Figures B.8 to B.10 display graphical results of the model-based clustering method using the R function “mclust”. Figure B.8 shows the Bayesian Information Criteria that indicate the optimum number of clusters (9) based on various configurations of Gaussian mixture models designated by the legend in the figure. The curves for all of the mixture models, with the exception of the spherical model, show that cluster discrimination is not distinct, whereas the spherical (EII) model indicates that discrimination between the clusters increases with an increasing number of clusters. Figure B.9 shows the distribution density of the nine clusters determined by mclust. Figure B.10 shows the designated clusters plotted geospatially. Tables B.1 to B.4 show the assignment of each class for each theme as calculated from the “mclust” function. For all of the themes, except SRF Geochemical Domains, there are many classes that are not specifically associated with clusters. This is due to the inability to statistically recognize these classes as distinct compositions. “Masking” and “swamping” are common effects for datasets where there the number of classes (priors) are under-represented.



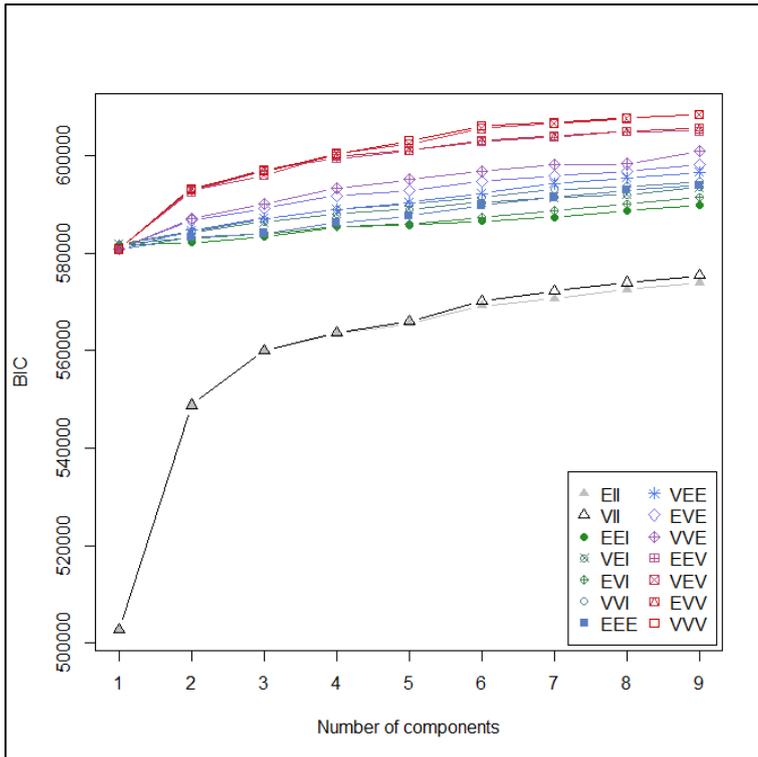


Figure B.8 Bayesian Information Criteria indicating the optimum number of clusters (9) based on various configurations of Gaussian mixture models (legend)

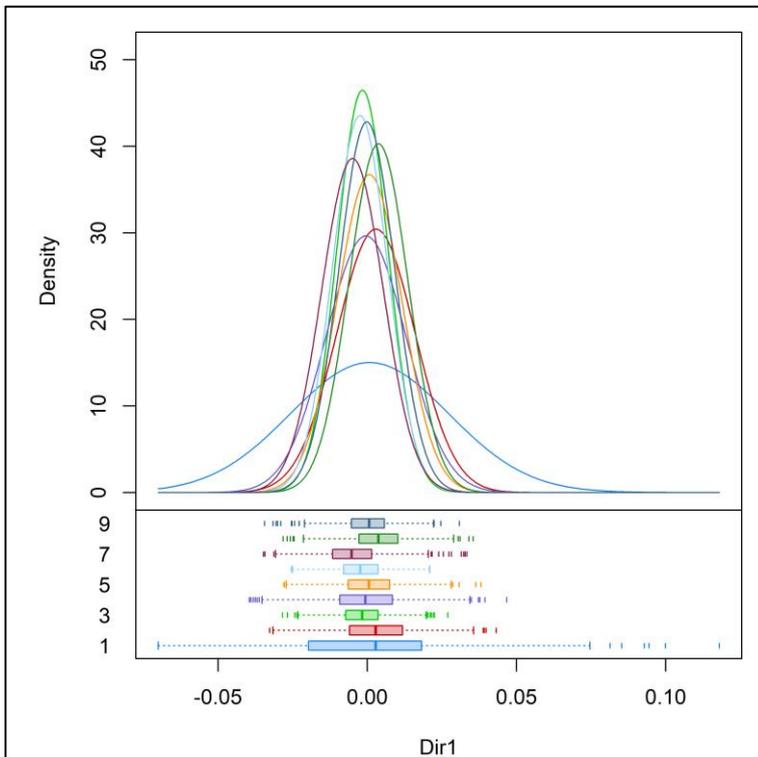


Figure B.9 Distribution density of the nine clusters determined by mclust. The fact that the clusters all overlap indicates that there is little difference between them and that the individual clusters are not unique, although they reflect distinctive geospatial regions.



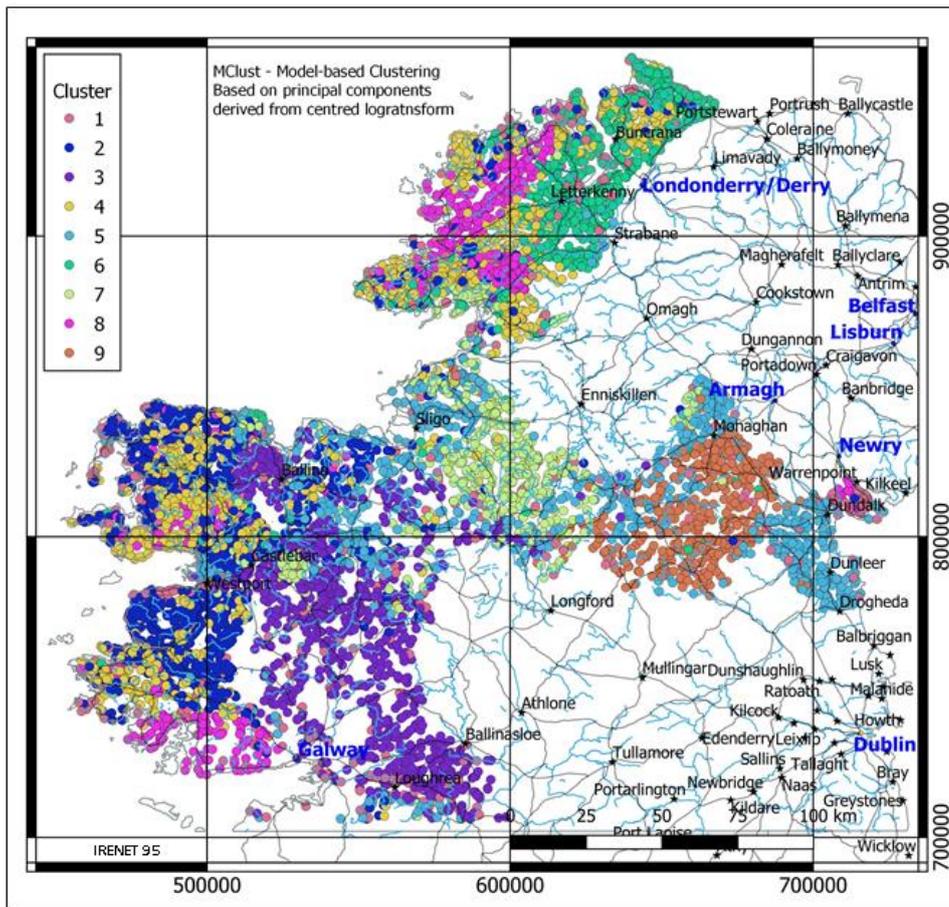


Figure B.10 Spatial distribution of clusters determined by mclust

Comparison of the sample sites classified by assigned cluster number (Figure B.10) and the sample sites classified by rock type (Figure B.11) shows that several clusters are associated with distinct rock types. Cluster 1 is distributed throughout the survey area and has no clear association with any rock type. Survey sites associated with Cluster 2 occur in the western part of the area and are dominated by the Dalradian lithologies in Connemara. Minor amounts of granitic and limestone lithologies are also included in the western part of cluster 2. Cluster 3 is almost exclusively represented by Carboniferous limestone in the western-central part of the survey area. Cluster 4 is dominated by Dalradian rocks in Donegal. Cluster 5 is represented mostly by limestone except in the eastern part of the survey area, in County Louth, where Cluster 5 is almost exclusively greywacke. Cluster 6 occurs in the northern part of the survey area, coinciding with Dalradian amphibolite and pelite bedrock. Cluster 7 coincides with the distribution of Namurian rocks, in the central part of the survey area around Lough Allen and in a small area east of Westport. Cluster 8 corresponds to granitic rocks in Counties Donegal and Galway. Cluster 9 coincides with the greywacke of the Longford-Down Inlier. With the exception of Cluster 1, the model-based clusters represent areas of geospatial continuity. There are some dominant rock types associated with each of the clusters, but there is also overlap of the rock types across the clusters.



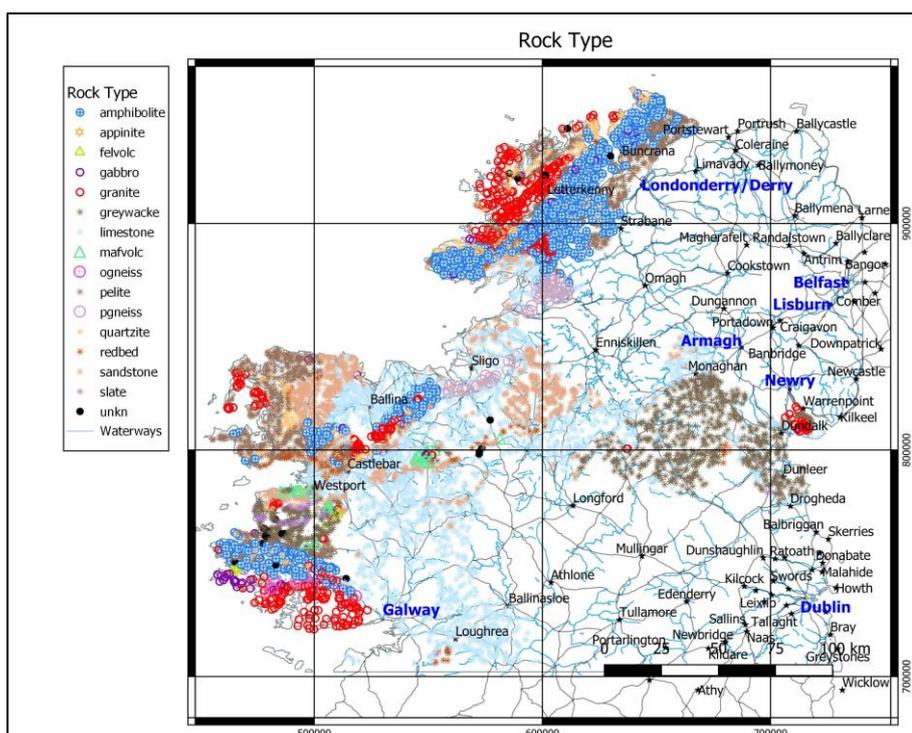


Figure B.11 Stream water samples sites classified by rock type

A comparison of the Teagasc subsoil map (not shown) with the clusters indicate that Clusters 4 and 8 are associated with blanket peat. Limestone and sandstone tills (TLs, TDSs) are associated with Cluster 3. Tills derived from metamorphic rocks (TMP) are exclusively associated with Cluster 6. Cluster 7 has a distinct compositional similarity with TNSSs, tills derived from Namurian shale and sandstone, consistent with the observation for bedrock geology. Cluster 9 has a distinctive geospatial association with till derived from Lower palaeozoic sandstone and shale (TLPSSs).

Comparison of the SRF Geochemical Domains (SRF Domain) map (not shown) with the mode-based clustering shows that individual SRF Geochemical Domains overlap several of the clusters. Nevertheless, more specific associations can be discerned. Cluster 3 tends to be associated with the western part of SRF Domain 2 (Limestone). Cluster 4 is associated with the coastal regions of SRF Domain 7 (Metamorphic rocks). Cluster 6 is associated with the northern part of SRF Domain 7. Cluster 7 is exclusively associated with SRF Domain 1 (Namurian sediments) in the Lough Allen area. Cluster 8 is associated with SRF Domain 6 (Granitic rocks) and Cluster 9 is associated with the eastern portion of SRF Domain 5 (Lower Palaeozoic sediments).

With the exception of the peat bogs, landcover classes do not have any unique or distinctive associations with the model-based clusters based on the principal components derived from the stream water geochemistry.



	1	2	3	4	5	6	7	8	9
amphibolite	48	186	0	428	13	318	6	87	1
appinite	0	0	0	7	0	0	0	3	0
felvolc	0	3	0	5	0	0	0	0	0
gabbro	12	7	0	10	3	7	0	8	1
granite	53	61	2	58	11	9	1	281	0
greywacke	57	203	0	54	179	8	12	6	374
limestone	167	82	897	22	425	16	226	2	26
mafvolc	5	20	0	0	11	0	0	0	0
ogneiss	4	2	0	5	0	0	0	5	0
pelite	82	241	0	248	9	193	5	76	0
pgneiss	5	8	0	38	10	6	2	1	1
quartzite	25	62	2	160	5	37	0	33	1
redbed	7	41	0	48	9	3	7	3	4
sandstone	45	215	45	171	129	15	323	5	4
slate	1	43	0	4	2	0	0	0	1
unkn	2	10	4	3	0	1	0	3	0

Table B.1 Theme Class Assignment from Model-based Clustering: Rock Type.

Table B.1 shows the assignment of themes using model-based clustering with bedrock type. Cluster 3 reflects the composition of stream water associated with limestone. Cluster 8 and 9 reflect compositions associated with granite and greywacke, respectively. Cluster 8 also shows that there is compositional overlap with amphibolite, pelite and quartzite. Cluster 7 shows compositional overlap between limestone and sandstone. With the exception of limestone (Cluster 3), there are no clusters that separate the different rock type classes distinctively.



	1	2	3	4	5	6	7	8	9
A	63	66	117	36	207	65	108	8	141
Ac	0	1	0	0	1	0	0	0	1
Ag	0	4	0	0	0	0	0	0	0
Asi	1	0	1	0	1	0	0	0	0
BasEsk	1	0	3	0	1	0	0	0	0
BktPt	120	580	10	825	37	119	67	370	1
Cut	79	59	459	18	220	5	139	0	141
FenPt	1	0	1	0	1	0	0	0	0
GDCSs	0	0	0	0	3	0	1	0	0
GGr	0	4	0	0	0	0	0	0	0
GLPSs	0	5	0	0	0	0	0	0	0
GLPSsS	1	0	0	0	2	0	0	0	3
GLs	2	3	10	1	2	0	0	0	1
GMp	0	5	0	5	0	7	0	0	0
IrSTLPSsS	1	0	0	0	8	0	0	0	0
KaRck	3	0	7	0	4	0	7	0	0
L	4	0	17	0	6	1	1	0	1
Lsi	0	0	0	0	1	0	0	0	1
Made	3	0	0	0	1	1	0	0	0
Mbs	0	1	0	1	0	0	0	0	0
Mesc	5	2	1	1	1	0	0	0	0
MGs	4	1	1	0	5	2	0	0	0
Mrl	0	0	1	0	0	0	0	0	0
Rck	37	171	3	111	21	41	34	72	19
Scree	1	12	0	3	2	0	2	0	0
TCSsCh	0	0	0	0	2	0	2	0	3
TDCSs	9	36	11	22	40	6	12	4	1
TDCSsS	2	1	1	1	4	2	16	0	0
TDSS	8	19	25	14	16	0	5	0	0
TGr	28	13	1	5	2	3	0	27	0
TLPDSS	0	2	1	0	6	0	0	0	0
TLPSS	10	21	1	6	0	0	0	0	0
TLPSSS	18	64	1	8	67	4	2	0	98
TLs	40	5	270	0	44	2	13	0	1
TMp	47	102	2	182	31	342	18	31	0
TNCSSs	0	0	0	0	0	0	3	0	0
TNSSs	12	3	5	8	66	0	150	0	0
TQz	4	2	1	14	2	8	0	0	1
unkn	0	1	0	0	1	1	2	1	0
Ws	5	1	0	0	1	4	0	0	0
Wsd	4	0	0	0	0	0	0	0	0

Table B.2 Theme Class Assignment from Model-based Clustering: Teagasc subsoil.

Table B.2 shows the assignment of themes using model-based clustering with Teagasc subsoil. The results of the clustering show that blanket peat and cutover peat overlap with every Teagasc subsoil class but each tends to overlap with different subsoil classes. The blanket peat is most strongly linked with TMp (tills derived from metamorphic rocks) and Rck (bedrock close to surface). Cutover peat is predominantly associated with TLs (tills derived from limestone), A (alluvium) and TNSSs (tills derived from Namurian shales and sandstones).



	1	2	3	4	5	6	7	8	9
D1	14	3	7	12	68	0	291	0	2
D2	155	121	888	78	375	7	183	6	23
D3	22	129	18	91	84	12	57	4	5
D4	12	44	26	46	23	3	9	4	0
D5	79	314	2	87	202	12	12	12	379
D6	58	59	2	62	8	8	1	293	0
D7	171	505	4	882	45	570	27	191	3
unkn	2	9	3	3	1	1	2	3	1

Table B.3 Theme Class Assignment from Model-based Clustering: SRF Domains.

Table B.3 shows the assignment of themes using model-based clustering with SRF Geochemical Domains. Cluster 3 distinctively identifies D2 (limestone) and Cluster 9 identifies D5 (Lower Palaeozoic sediments). Cluster 1 shows a mix of D2 and D7 (metamorphic rocks). Cluster 2 shows a mix of D7, D5, D3 (Devonian-Carboniferous sandstones), and D2. Cluster 4 shows a mix of all domains but is dominated by D7. Cluster 5 shows a mix of D2 and D5. Cluster 6 is dominated by D7. Cluster 7 is dominated by D1 (Namurian shales and sandstone) and D2 and Cluster 8 reflects a mixture of D6 and D7. As with the rock type theme, most of the clusters display overlap between the different classes.

	1	2	3	4	5	6	7	8	9
AgNatVeg	112	179	80	132	150	79	248	50	34
ArableNI	5	0	0	0	40	9	0	0	1
BeachDune	5	2	1	0	0	2	0	0	0
Burn	0	0	0	0	0	0	0	3	0
CompCult	1	0	1	6	4	1	0	6	0
ForestBL	5	7	8	5	8	5	3	3	2
ForestCon	14	82	22	121	16	9	47	35	2
ForestMix	1	5	5	4	6	2	5	3	9
GrassNat	5	14	1	13	13	1	8	9	0
MarshInland	3	4	6	0	4	0	2	0	1
MarshSalt	4	0	0	0	1	0	0	0	0
MinMine	1	1	0	0	0	0	0	0	0
Moors	8	16	0	34	6	2	8	8	0
Pastures	227	195	747	91	510	413	141	24	362
PeatBog	103	616	61	739	41	69	78	353	1
Sport	4	0	0	0	0	0	0	0	0
UrbanDis	3	2	2	1	3	11	1	2	1
VegSparse	2	6	0	3	0	0	0	5	0
WoodlandTran	10	55	16	112	4	10	41	12	0

Table B.4 Theme Class Assignment from Model-based Clustering: Land Cover (Corine)

Table B.4 shows the assignment of themes using model-based clustering with Corine Land Cover classes. The results of the model-based clustering for the land cover theme indicate overlap of the peat and pasture classes with all of the clusters, albeit the peat and pasture classes do not overlap strongly with each other. The other classes do not show a distinctive association with any of the clusters.

B.3 Non-hierarchical Cluster Analysis based on K-Means



Non-hierarchical clustering methods are based on an initial selection of arbitrary seeds that define the initial clusters. This approach may offer some advantage over hierarchical methods since the clusters are formed based on multivariate similarities (proximities) rather than individual correlation coefficients. These methods start with an initial number of cluster centres that can be specified or randomly chosen. Each observation is allocated to one of the groups based on proximity to the group centres. The process is iterative and group centres change until there is a stable configuration of clusters. A recent implementation of clustering methods is described in the R package “Flexible Procedures for Clustering” (fpc).

K-means cluster analysis is a method that starts with an initial ‘guess’ of the cluster centres. The distance of each observation from each cluster centre is measured and then provisionally assigned to the closest cluster centre. A new cluster centre is calculated based on the designated observations for each previous centre. The process is iterative until it converges on stable centres. The method requires an initial choice of the number of cluster centres. If the number is too great, there will be many small clusters that have few points. If the number of centres is too few, then the structure of the data may not be realized. A disadvantage of the procedure is that a less than optimal clustering may result if the initial cluster centres do not fall in distinct clusters (Davis 2002, p. 500). Venables and Ripley (2002) provide a method by which a suitable number of starting clusters may be determined by using a combination of hierarchical clustering and PCA.

It is common to apply non-hierarchical clustering methods to PC scores. If one or more principal components can be inferred to represent specific geological/geochemical processes, then the application of cluster analysis can provide further insight in how those processes may be related. Additionally, the component plots provide a reduced set of dimensions for viewing the multi-element associations of the data and thus provide additional visual assistance in examining grouped associations.

K-means cluster analysis was carried out on the Tellus stream waters dataset using the “kmeans” function in the R package “stats”. Using the within-groups sums of squares, as described previously, nine clusters were chosen. The class assignments for each theme are shown in Tables B.5 to B.8. Figure B.12 shows the geospatial assignment of clusters to each stream water sample site. Each of the clusters show geospatially coherent patterns that are similar to the clusters generated by the method of model-based clustering (Figure B.11). Tables B.5 to B.8 show similar patterns to Tables B.1 to B.4, where the dominant classes for each of the themes define distinct geospatial patterns.



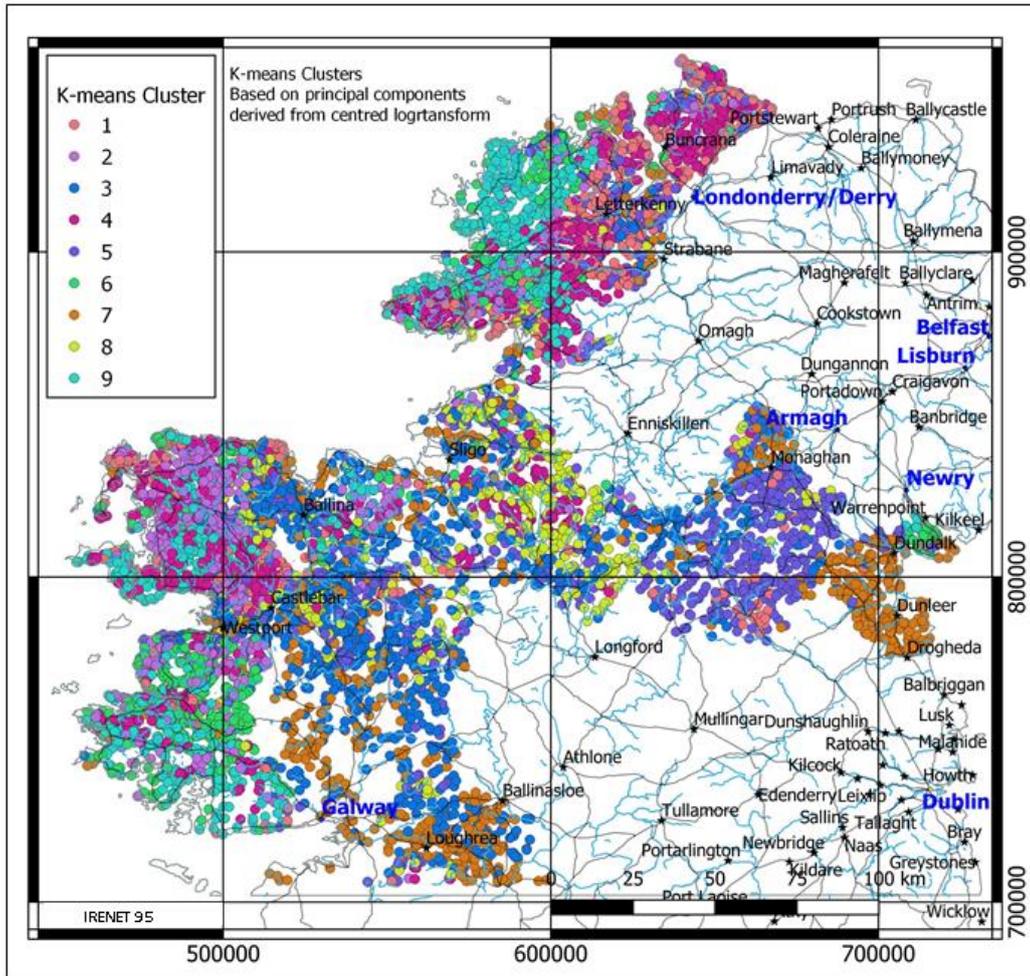


Figure B.12 Spatial distribution of clusters determined by k-means clustering

Cluster	1	2	3	4	5	6	7	8	9
amphibolit	17	84	174	122	110	81	347	14	138
appinite	0	1	2	2	2	0	2	0	1
felvolc	0	3	2	1	2	0	0	0	0
gabbro	3	11	6	3	8	4	8	3	2
granite	8	32	135	21	206	10	42	5	17
greywacke	153	126	30	50	49	328	21	59	77
limestone	511	33	19	25	14	301	16	760	184
mafvolc	1	7	0	3	1	12	2	5	5
ogneiss	0	2	3	7	3	0	1	0	0
pelite	14	87	168	133	100	67	197	8	80
pgneiss	2	4	25	8	4	8	14	3	3
quartzite	1	30	96	30	65	14	53	8	28
redbed	1	3	22	24	3	16	37	5	11
sandstone	42	44	81	180	13	133	115	117	227
slate	3	14	2	21	6	0	3	0	2
unkn	0	4	4	2	7	0	0	5	1

Table B.5 Theme Class Assignment from k-means Clustering: Rock Type



Cluster	1	2	3	4	5	6	7	8	9
A	197	36	11	33	23	206	42	150	113
Ac	0	1	0	0	0	0	0	2	0
Ag	0	2	0	0	1	0	1	0	0
Asi	2	0	0	0	0	0	0	1	0
BasEsk	3	0	0	0	0	1	0	1	0
BktPt	13	162	570	432	330	57	421	42	102
Cut	188	11	4	21	1	312	13	436	134
FenPt	2	0	0	0	0	1	0	0	0
GDCSs	1	1	0	0	0	0	0	1	1
GGr	0	4	0	0	0	0	0	0	0
GLPSs	0	2	0	1	2	0	0	0	0
GLPSsS	0	0	0	1	0	3	0	1	1
GLs	3	0	1	1	2	4	0	7	1
GMp	0	1	6	1	0	1	3	0	5
IrSTLPSSs	9	0	0	0	0	0	0	0	0
KaRck	9	1	1	0	1	2	0	4	3
L	12	0	0	0	0	3	0	15	0
Lsi	0	0	0	0	0	1	0	0	1
Made	2	0	0	0	0	0	0	2	1
Mbs	0	2	0	0	0	0	0	0	0
Mesc	5	0	0	1	1	1	1	0	1
MGs	8	1	0	0	2	0	0	0	2
Mrl	1	0	0	0	0	0	0	0	0
Rck	13	85	73	45	117	40	80	17	39
Scree	1	4	3	3	7	1	0	0	1
TCSsCh	0	0	0	0	0	3	0	2	2
TDCSs	10	8	12	18	4	22	16	27	24
TDCSsS	1	0	1	1	0	4	3	2	15
TDSSs	9	2	0	4	1	14	20	26	11
TGr	2	24	4	1	33	5	3	2	5
TLPDSs	0	0	0	0	0	3	0	1	5
TLPSs	1	19	5	4	0	3	2	3	1
TLPSsS	63	38	1	16	7	95	4	20	18
TLs	148	1	0	0	1	54	1	159	11
TMp	23	73	61	44	53	106	210	20	165
TNCSSs	0	0	0	0	0	0	0	0	3
TNSSs	22	5	11	2	1	27	31	46	99
TQz	1	0	5	2	4	3	7	2	8
unkn	0	1	0	1	1	2	0	1	0
Ws	5	1	0	0	0	0	0	1	4
Wsd	2	0	0	0	1	0	0	1	0

Table B.6 Theme Class Assignment from k-means Clustering: Teagasc subsoil

Cluster	1	2	3	4	5	6	7	8	9
D1	26	5	18	4	4	38	61	58	183
D2	474	26	40	74	16	306	34	724	142
D3	27	24	41	107	7	53	28	68	67
D4	9	2	18	21	5	22	39	28	23
D5	173	169	43	106	62	361	37	68	80
D6	7	43	135	23	217	12	33	6	15
D7	40	211	470	294	277	179	626	35	266
unkn	0	5	4	3	5	3	0	5	0

Table B.7 Theme Class Assignment from k-means Clustering: SRF Domain



Cluster	1	2	3	4	5	6	7	8	9
AgNatVeg	69	115	49	71	86	163	127	150	234
ArableNI	44	0	0	0	0	8	0	1	2
BeachDun	3	2	0	0	1	0	1	1	2
Burn	0	0	1	0	0	0	2	0	0
CompCult	6	1	4	2	4	0	1	0	1
ForestBL	9	8	2	2	4	6	2	6	7
ForestCon	5	14	88	55	24	21	77	29	35
ForestMix	2	1	3	1	3	12	6	8	4
GrassNat	1	14	13	1	16	10	5	2	2
MarshInlai	5	2	1	0	0	4	0	6	2
MarshSalt	3	1	0	0	0	1	0	0	0
MinMine	0	0	0	1	0	0	0	0	1
Moors	2	14	15	6	12	2	20	5	6
Pastures	572	124	28	69	48	676	165	681	347
PeatBog	21	181	498	380	375	56	369	83	98
Sport	3	0	0	0	0	0	0	1	0
UrbanDis	4	1	2	0	2	5	1	4	7
VegSparse	1	3	1	0	8	0	3	0	0
Woodland	6	4	64	44	10	10	79	15	28

Table B.8 Theme Class Assignment from k-means Clustering: Land Cover (Corine)

B.4 Summary of Clustering Methods

The application of cluster analysis provides some insight into the similarity of stream water samples that are partly influenced by underlying geology, Teagasc subsoil and SRF Geochemical Domains. The influence of Land Cover is not readily recognized in the initial data investigation undertaken here. In this study, principal component analysis was used as a method to increase the signal to noise ratio and create linear combinations of elements that reflect mineral stoichiometry and other processes. The issue of under-sampled processes as described above can be difficult to resolve. Additional study may include the grouping/consolidation of classes that are believed to similar in process or geochemical composition.



Appendix C: Random Forest

Classification

The following description of Random Forests is adapted from Harris and Grunsky (2015). Random Forests (Breiman 2001) is an ensemble, multiple decision-tree classifier that offers a number of advantages for modelling: geochemical data. Training data (sampling sites where the predictive class has been verified) are required for this approach. Input parameters into the Random Forest (RF) classifier are minimal and include only the number of variables (log-centred geochemistry or principal components) for each tree and the number of trees to create. The RF process performs internal cross-validation through bootstrapping and provides a robust estimate of classification accuracy through out-of-bag (OOB) estimates. A bagging process (i.e., bootstrap sampling) is where approximately two thirds of the training areas (pixels) are randomly selected with replacement and these are used for generating the classification (in-bag data) and the remaining one third (OOB) is used for validation. This random sampling with replacement of the training dataset is undertaken for every tree. The bag data are used to create multiple decision trees, which are applied to produce independent classifications and the OOB data is used to validate the classification by calculating an OOB error.

An ensemble of trees (predictions) is created and a voting procedure is employed to assign the majority class to each observation in the final prediction. According to Breiman (2001), RF is less sensitive to noise or over fitting and there is no need for cross validation as it is performed internally (e.g. OOB). However, as with any supervised classification method, an independent check training dataset of occurrences is still required to calculate an unbiased and more robust estimate of classification accuracy. In addition to the classification map generated by RF, a probability map is also generated that shows the strength of membership of the estimated classes (rock type, subsoil, etc.). Another very useful aspect of RF is that it calculates the importance (predictive power) of each variable in the classification process. The main point of ensemble classifiers, such as RF, is that the process produces not just from a single prediction (decision tree) but from many predictions which are then combined. This is beneficial as this process helps reduce the variance by minimizing any bias that may be the result of a unique single training dataset.

For the Tellus stream waters dataset, Random forest classification was applied to each of the four themes (rock type, subsoil, SRF Geochemical Domains, land cover) with the function "randomForest" from the R package "RandomForest", using the parameters listed in Table C.1.



Parameter	Comment
ntree = 5	Default number of t grown for each sample site
type = "classification"	Type of prediction for each class for each theme: "regression", "classification" or "unsupervised".
proximity = TRUE	Proximity (distance) between observations are measured as part of clustering process
classwt = prior probabilities	Prior probabilities that are measured for each class for each theme
importance = TRUE	Relative importance of predictors
norm votes = TRUE	The final results of votes are expressed as fractions

Table C.1 Random Forest Parameters

Because of the imbalance of classes in the themes for rock type, subsoils and land cover, prior probabilities were included in the random forest classification. Imbalance refers to unequal numbers of observations for each class. Classes with large numbers of observations can "mask" and "swamp" classes with a small number of observations. This is particularly the case for the Teagasc subsoil classes, where numerous classes with very limited geographical extent intersect relatively few stream water sites.

For predictions, the values of "type" are set to "response" for the class prediction and "prob" for the posterior probabilities (normalized votes).

A measure of variable significance is provided by the Gini index, which calculates the amount of probability of a specific class that is classified incorrectly when selected randomly. Increasing Gini index values indicates that the variable is better at discriminating between the classes. Variables with low Gini index values are less significant for discriminating between the classes.

C.1 Random Forests – Rock Type

Figure C.1 shows the variable significance obtained from the application of rock type based on the principal components. The figure shows the mean decrease measure of the Gini index. Increasing Gini indexes indicates that the variable is better at discriminating between the classes. Variables with low Gini index values are less significant for discriminating between the classes. For the theme rock type, PCs 1, 5, 2, 7, 6 are the main variable for classifying the rock types from the waters chemistry. PC1 is by far the better variable for classification alone.



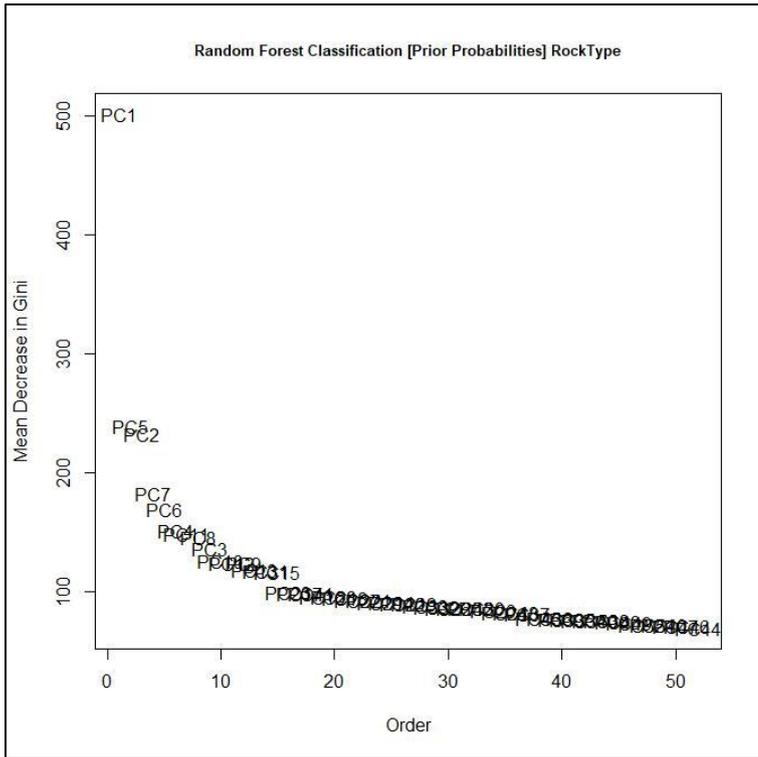


Figure C.1 Mean decrease of Gini index (Rock Type)

Tables C.2 and C.3 show the individual classification accuracy for each rock type. The matrix of counts (Table C.2) shows the individual point counts associated with the prediction of each class. The accuracy % table (Table C.3) shows the point count assignments in a percentage form. The diagonal along the accuracy matrix shows the accuracy of prediction for each rock type. The off-diagonal elements indicate the misclassification/confusion of the each rock type with respect to the others in the theme. The overall accuracy is 69.7 %, which is the average predictive accuracy over all of the classes, defined as the number of correct predictions divided the number of total predictions. Measures of recall (true positives)/(true positives + false negatives) and precision (true positives)/(true positives + false positives) indicate the performance of each class. Recall captures as many positives as possible and precision captures the true accuracy of the prediction.

The precision of the predicted rock types indicates that paragneiss, redbeds, quartzite, granite and greywacke are well predicted. Classes including appinite, felsic volcanics, gabbro, mafic volcanics and orthogneiss are not well predicted because (i) they have very low counts in the initial dataset and (ii) they overlap with other classes, despite the correction applied by the prior probabilities. This is also reflected in the measure of recall. Low recall values indicates the success of capturing the positive results. High recall values are noted for limestone, greywacke and amphibolite. Comparison of the precision and recall (Table C.4) indicates that the two measures do not correlate well for most of the classes. This also indicates significant confusion in the prediction. Sources for error or uncertainty in the predictions may be due to compositional overlap of the chemistry of the waters between the rock types, misclassification of the original site and the variability created by the range of influences on stream water chemical composition.



	Counts														
	amphibolite	appinite	felvolc	gabbro	granite	greywacke	limestone	mafvolc	ogneiss	pelite	pgneiss	quartzite	redbed	sandstone	slate
amphibolite	872	0	0	0	14	13	99	0	0	78	0	5	0	6	0
appinite	8	0	0	0	2	0	0	0	0	0	0	0	0	0	0
felvolc	2	0	0	0	1	4	0	0	0	1	0	0	0	0	0
gabbro	20	0	0	0	7	3	10	0	0	7	0	0	0	1	0
granite	60	0	0	0	318	13	41	0	0	35	0	2	0	7	0
greywacke	19	0	0	0	7	748	93	0	0	4	0	0	0	21	1
limestone	23	0	0	0	2	37	1735	0	0	2	0	0	1	63	0
mafvolc	3	0	0	0	0	10	18	0	0	0	0	0	0	5	0
ogneiss	4	0	0	0	5	1	1	0	0	5	0	0	0	0	0
pelite	242	0	0	0	16	29	61	0	0	489	0	3	0	14	0
pgneiss	35	0	0	0	0	1	20	0	0	7	6	0	0	2	0
quartzite	148	0	0	0	10	8	31	0	0	66	0	55	0	7	0
redbed	18	0	0	0	1	5	36	0	0	17	0	0	23	22	0
sandstone	69	0	0	0	0	22	338	0	0	22	0	0	0	501	0
slate	5	0	0	0	2	34	4	0	0	2	0	0	0	1	3

Table C.2 RF matrix of counts for Rock Type

	Accuracy %														
	amphibolite	appinite	felvolc	gabbro	granite	greywacke	limestone	mafvolc	ogneiss	pelite	pgneiss	quartzite	redbed	sandstone	slate
amphibolite	80.2	0	0	0	1.3	1.2	9.1	0	0	7.2	0	0.5	0	0.6	0
appinite	80	0	0	0	20	0	0	0	0	0	0	0	0	0	0
felvolc	25	0	0	0	12.5	50	0	0	0	12.5	0	0	0	0	0
gabbro	41.7	0	0	0	14.6	6.2	20.8	0	0	14.6	0	0	0	2.1	0
granite	12.6	0	0	0	66.8	2.7	8.6	0	0	7.4	0	0.4	0	1.5	0
greywacke	2.1	0	0	0	0.8	83.8	10.4	0	0	0.4	0	0	0	2.4	0.1
limestone	1.2	0	0	0	0.1	2	93.1	0	0	0.1	0	0	0.1	3.4	0
mafvolc	8.3	0	0	0	0	27.8	50	0	0	0	0	0	0	13.9	0
ogneiss	25	0	0	0	31.2	6.2	6.2	0	0	31.2	0	0	0	0	0
pelite	28.3	0	0	0	1.9	3.4	7.1	0	0	57.3	0	0.4	0	1.6	0
pgneiss	49.3	0	0	0	0	1.4	28.2	0	0	9.9	8.5	0	0	2.8	0
quartzite	45.5	0	0	0	3.1	2.5	9.5	0	0	20.3	0	16.9	0	2.2	0
redbed	14.8	0	0	0	0.8	4.1	29.5	0	0	13.9	0	0	18.9	18	0
sandstone	7.2	0	0	0	0	2.3	35.5	0	0	2.3	0	0	0	52.6	0
slate	9.8	0	0	0	3.9	66.7	7.8	0	0	3.9	0	0	0	2	5.9
Overall Accuracy	69.73														

Table C.3 RF accuracy % matrix for Rock Type

Precision		Recall	
amphibolite	0.57	amphibolite	0.80
appinite	NA	appinite	0.00
felvolc	NA	felvolc	0.00
gabbro	NA	gabbro	0.00
granite	0.83	granite	0.67
greywacke	0.81	greywacke	0.84
limestone	0.70	limestone	0.93
mafvolc	NA	mafvolc	0.00
ogneiss	NA	ogneiss	0.00
pelite	0.67	pelite	0.57
pgneiss	1.00	pgneiss	0.08
quartzite	0.85	quartzite	0.17
redbed	0.96	redbed	0.19
sandstone	0.77	sandstone	0.53
slate	0.75	slate	0.06

Table C.4 RF Measures of Precision and Recall.

Figures C.2 and C.3 show two maps, one classifying the stream water data according to bedrock type and the other showing the predicted classes (bedrock types) based on random forest analysis. From the scale of the maps as presented in Figures C.2 and C.3, the predictions of the rock types are reasonably close to the mapped rock types.



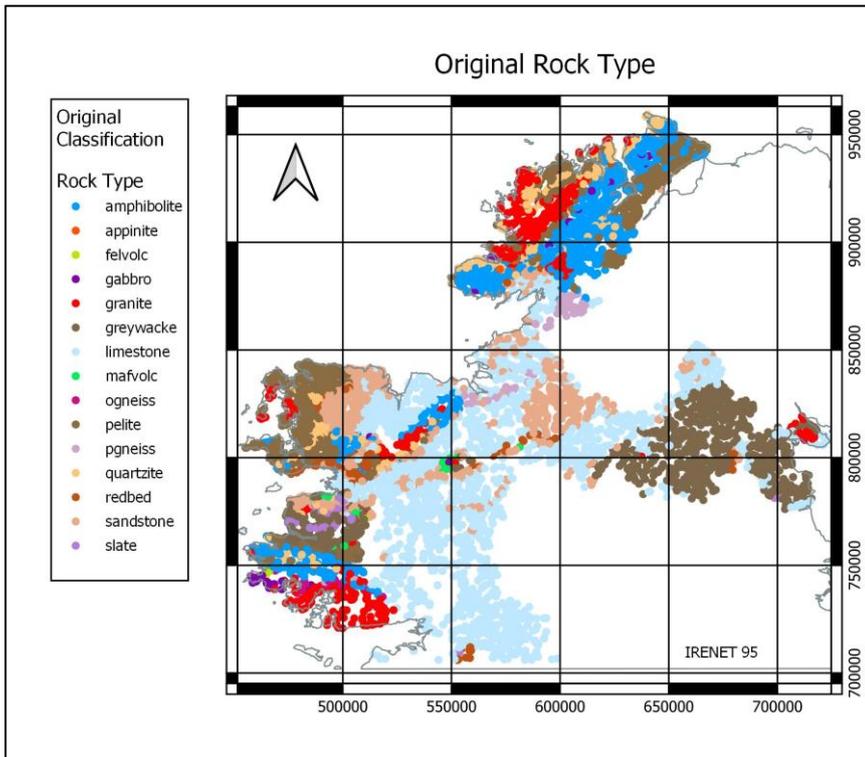


Figure C.2 Stream water sites classified by mapped rock type.

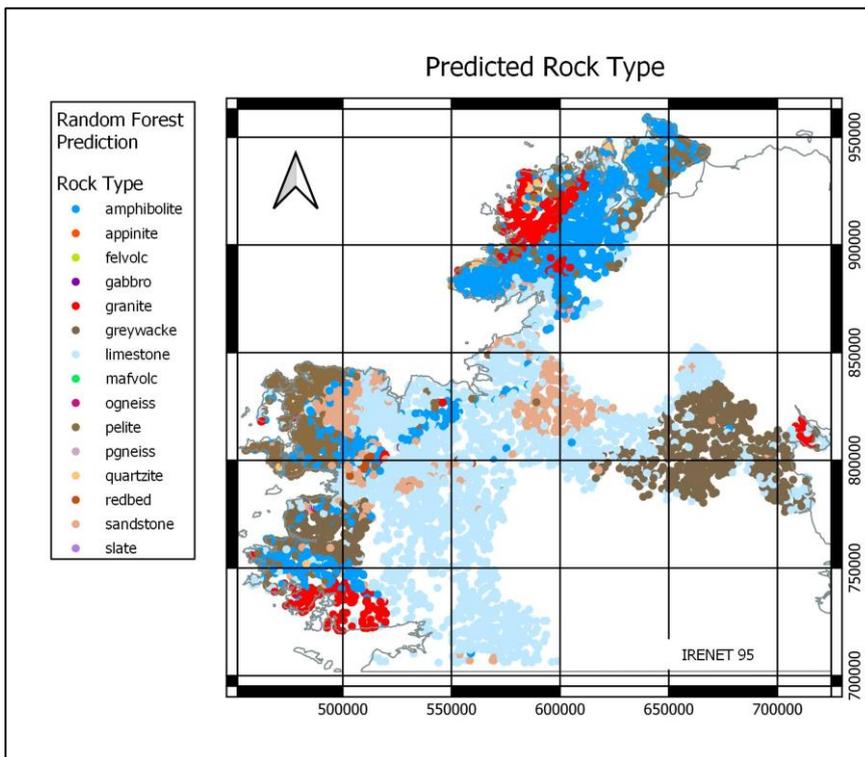


Figure C.3 Stream water sites classified by RF prediction.



C.2 Random Forests – Teagasc Subsoil

Figures C.4, C.5 and C.6 show the variable significance (Gini index) and the original and predicted maps for Teagasc subsoil. The table showing the prediction accuracies for the Teagasc subsoil classes (Table C.5) reflects the very large number of subsoil classes (40). Many of these classes have only a few sites within the waters survey area and have low or zero accuracy scores. The under-representation of these classes within any classification scheme can yield uncertain results. Figure C.4 shows that PC2 is the dominant principal component for discriminating between the subsoils. Additionally, in descending order, PC13, PC1, PC7, PC4, PC8 and PC5 can help discriminate between the subsoil classes. For the 40 classes of subsoil, the only classes that show any form of classification accuracy are alluvium (20 %), blanket peat (97 %), cut away peat (73 %), rock (3 %) and several classes of till with dominant clasts derived from the Phanerozoic sedimentary assemblages including Lower Palaeozoic sandstones (8 %), Lower Palaeozoic sandstone and shale (5 %), Carboniferous limestone (11 %), metamorphic rocks (25 %) and Namurian shales and sandstones (29 %). Measures of precision and recall are low and/or inconsistent. Although there is significant uncertainty in the class prediction, the method of random forests, predicts the classes in a consistent manner as shown in Figures C.5 and C.6 where the broad subsoil classes of the original sites are very similar to the predicted subsoil sites. Despite the generally low individual prediction accuracies, the overall accuracy is 49.4 %, which suggests that the dominant classes are reasonably well predicted.

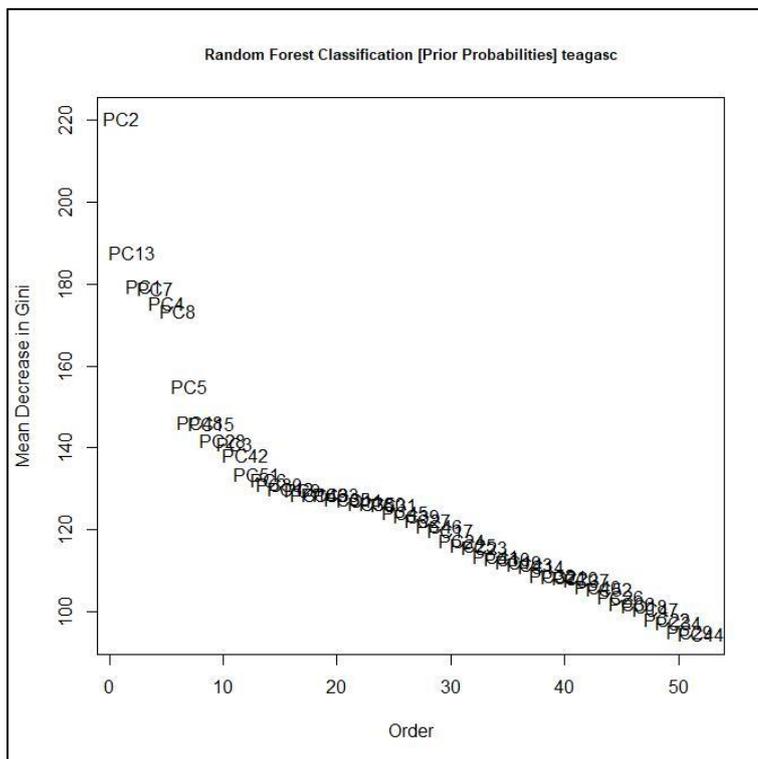


Figure C.4 Mean decrease of Gini index (Teagasc subsoil classes)



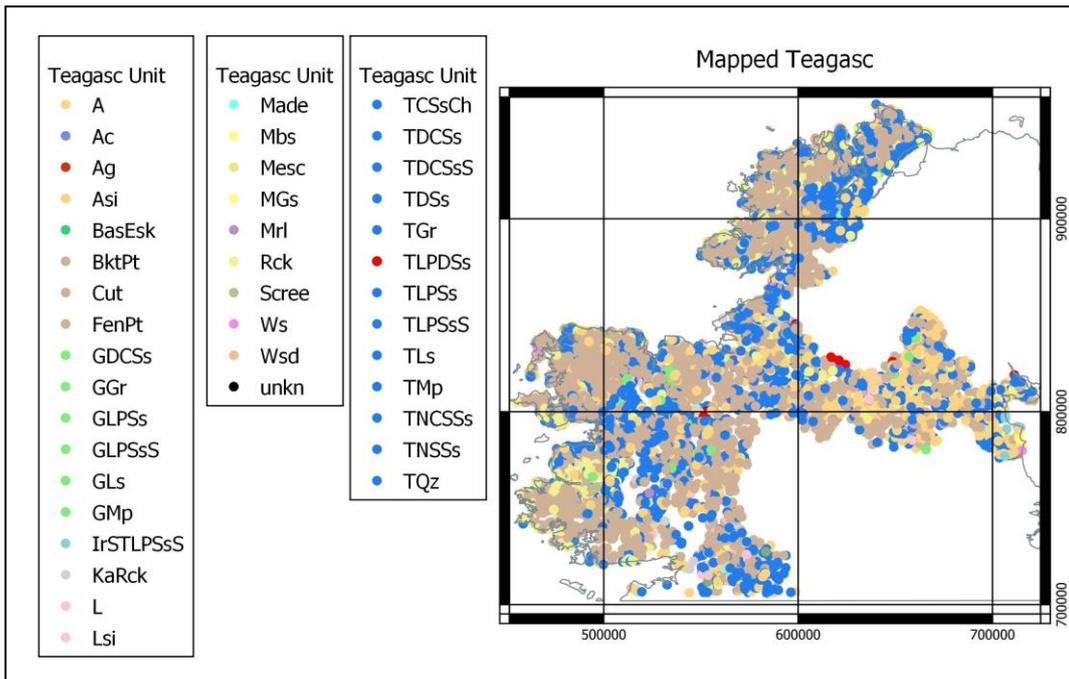


Figure C.5 Stream water sites classified by mapped subsoil.

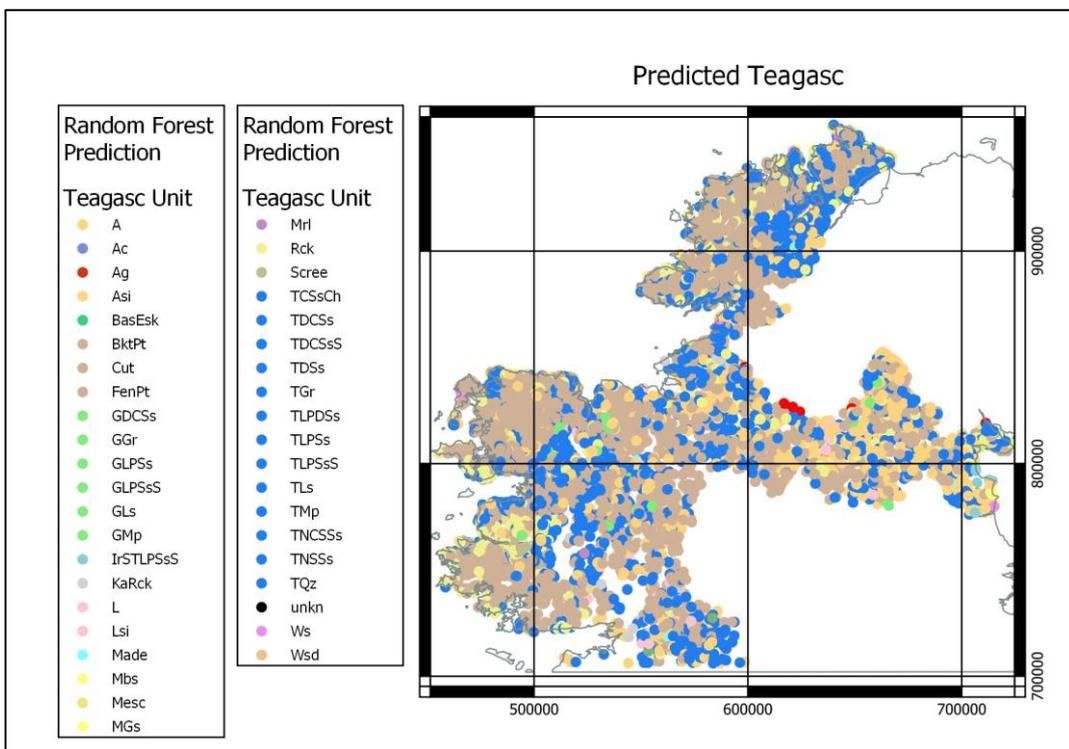


Figure C.6 Stream water sites classified by subsoil types as predicted by Random Forests

Further refinement in the use of Teagasc subsoils through the reduction of the number of classes by increased generalization of the class features may be warranted. This process largely underpins the SRF Domain classification.



C.3. Random Forests – SRF Geochemical Domains

Random forests classification and prediction of the SRF Geochemical Domains are shown in Figures C.7, C.8 and C.9 and in Table C.6. Figure C.7 shows the variable significance of the principal components. For the SRF Geochemical Domain, several principal components contribute to the discrimination of the seven domains. The main components in descending order of importance are PC2, PC1, PC7, PC5, PC6, PC13, PC4 and PC8. The overall prediction accuracy is 76.8 %. The accuracies of the predicted domains are 59 % (Domain 1), 84 % (Domain 2), 23 % (Domain 3), 11 % (Domain 4), 71 % (Domain 5), 46 % (Domain 6) and 97 % (Domain 7) (Table C.6). The measures of precision and recall are better than those reported for the prediction of rock type and subsoil. Comparison of Figures C.8 and C.9 shows that the geospatial extent for each of the domains in the predicted map is similar to that in the original map. Some misclassification is observed between Domain 5 and Domain 7 in the west and between Domain 6 and Domain 7 in the northwest. Overlap or confusion between Domain 1 and Domain 2 is also evident in the northwest. The predicted geospatial extent of Domain 1 is reduced from the extent observed in the original map. The generally strong agreement between the original domain map and the map predicted from random forest analysis both validates the geochemical domain map itself and provides further support for the observation that the stream waters data carries a strong geogenic signal.

Random Forest SRF Domain Prediction Accuracy									
	Accuracy Count							Precision Recall	
	D1	D2	D3	D4	D5	D6	D7		
D1	235	110	0	0	2	0	50	0.9073	0.5919
D2	23	1551	13	1	55	0	193	0.7686	0.8448
D3	1	145	97	2	7	0	170	0.8509	0.2299
D4	0	70	2	18	4	0	73	0.8571	0.1078
D5	0	78	0	0	779	7	235	0.9058	0.7088
D6	0	13	0	0	6	227	245	0.9080	0.4623
D7	0	51	2	0	7	16	2322	0.7062	0.9683
	Accuracy %								
	D1	D2	D3	D4	D5	D6	D7		
D1	59.2	27.7	0	0	0.5	0	12.6		
D2	1.3	84.5	0.7	0.1	3	0	10.5		
D3	0.2	34.4	23	0.5	1.7	0	40.3		
D4	0	41.9	1.2	10.8	2.4	0	43.7		
D5	0	7.1	0	0	70.9	0.6	21.4		
D6	0	2.6	0	0	1.2	46.2	49.9		
D7	0	2.1	0.1	0	0.3	0.7	96.8		
Overall									
Accuracy 76.78414									

Table C.6 RF matrix of counts and accuracy % for SRF Domains



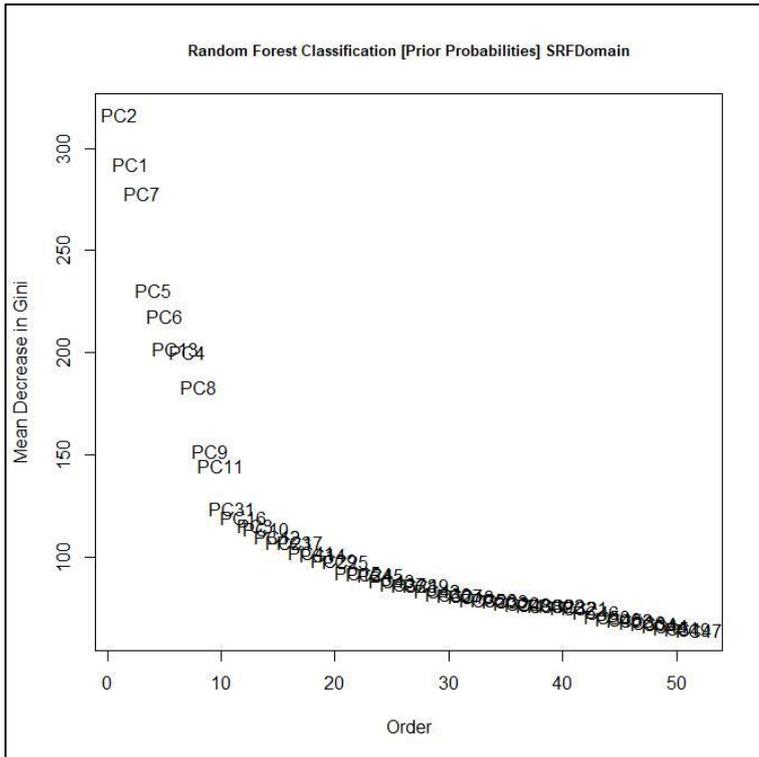


Figure C.7 Mean decrease of Gini index (SRF Geochemical Domains)

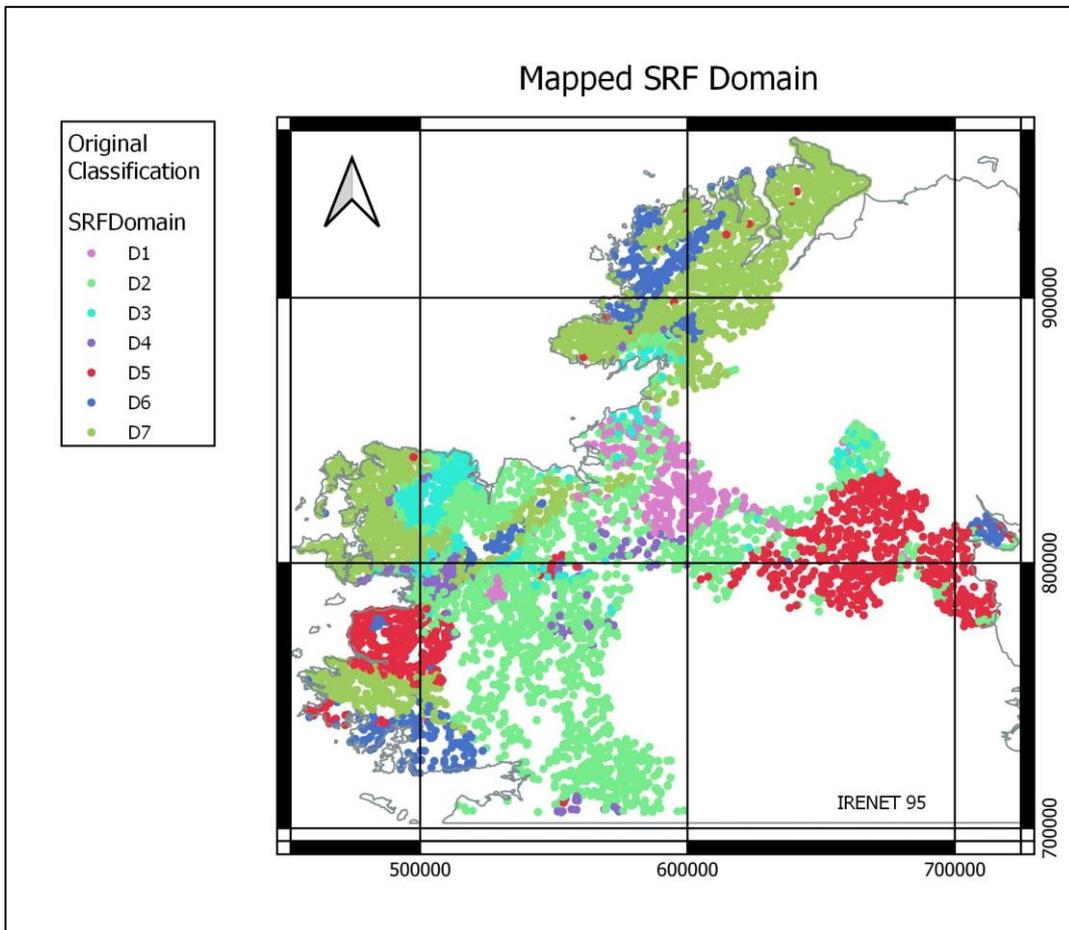


Figure C.8 Stream water sites classified by mapped SRF Geochemical Domain.



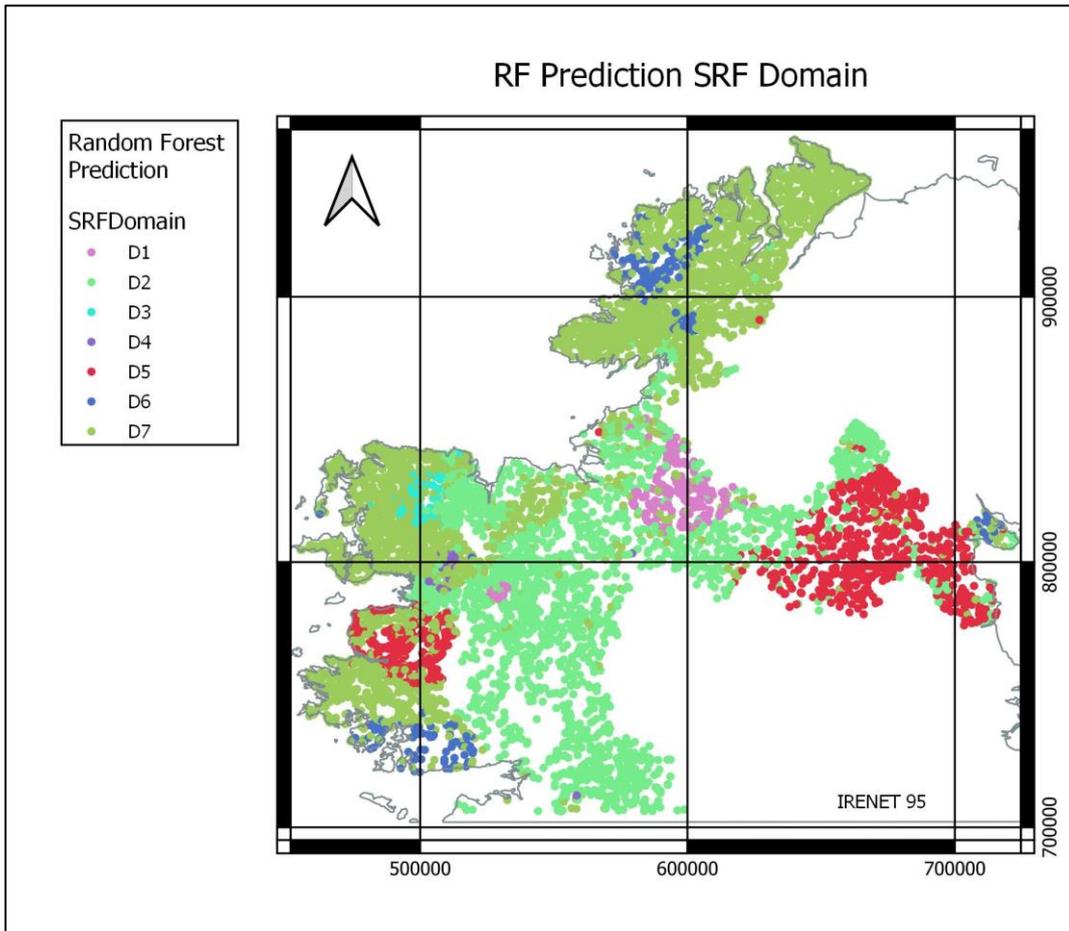


Figure C.9 Stream water sites classified by SRF Geochemical Domain as predicted by random forests.

C.4. Random Forests – Land Cover

The geospatial distributions of many of the Corine Land Cover classes have very limited geospatial extent. The geographically most extensive land cover classes in the area covered by Tellus stream water data are pasture and peat bog. This is reflected in the prediction accuracies where these two classes dominate (Table C.7). The variance of the waters compositions for the pastures and peat bog overlap with all of the other classes, which results in a sub-optimal classification. Figure C.10 shows that PC1, 2, and 8 are the most significant principal components for discriminating the classes. Figures C.11 and C.12 indicate that the broad geospatial pattern of the original and predicted data are similar. The degree of confusion/misclassification is difficult to see at the scale of the figures. Nonetheless, the overall prediction accuracy of 64.5 % indicates that there is useful prediction capacity for peat bog and pasture land cover classes.



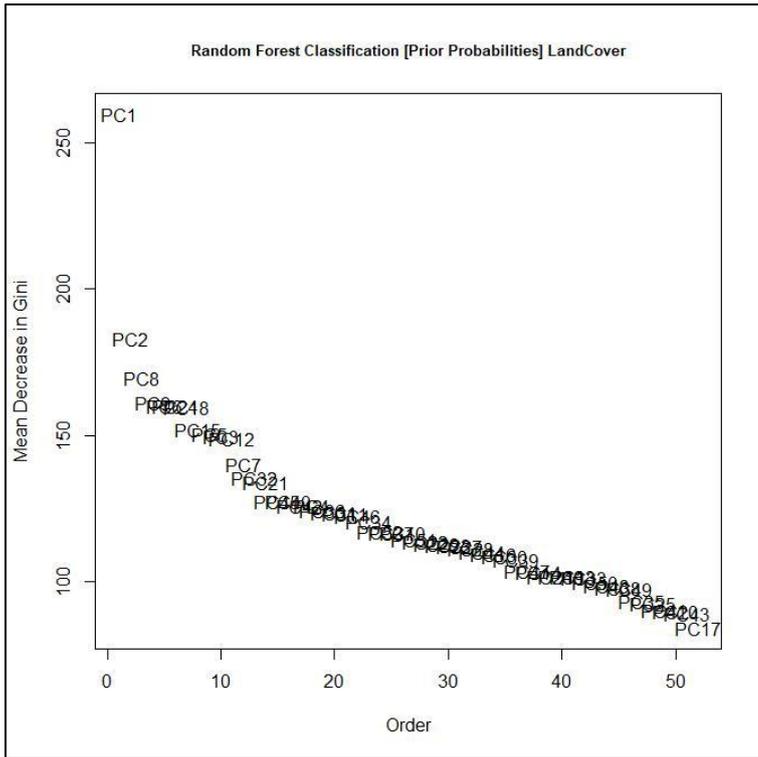


Figure C.10 Mean decrease of Gini index (Land Cover)

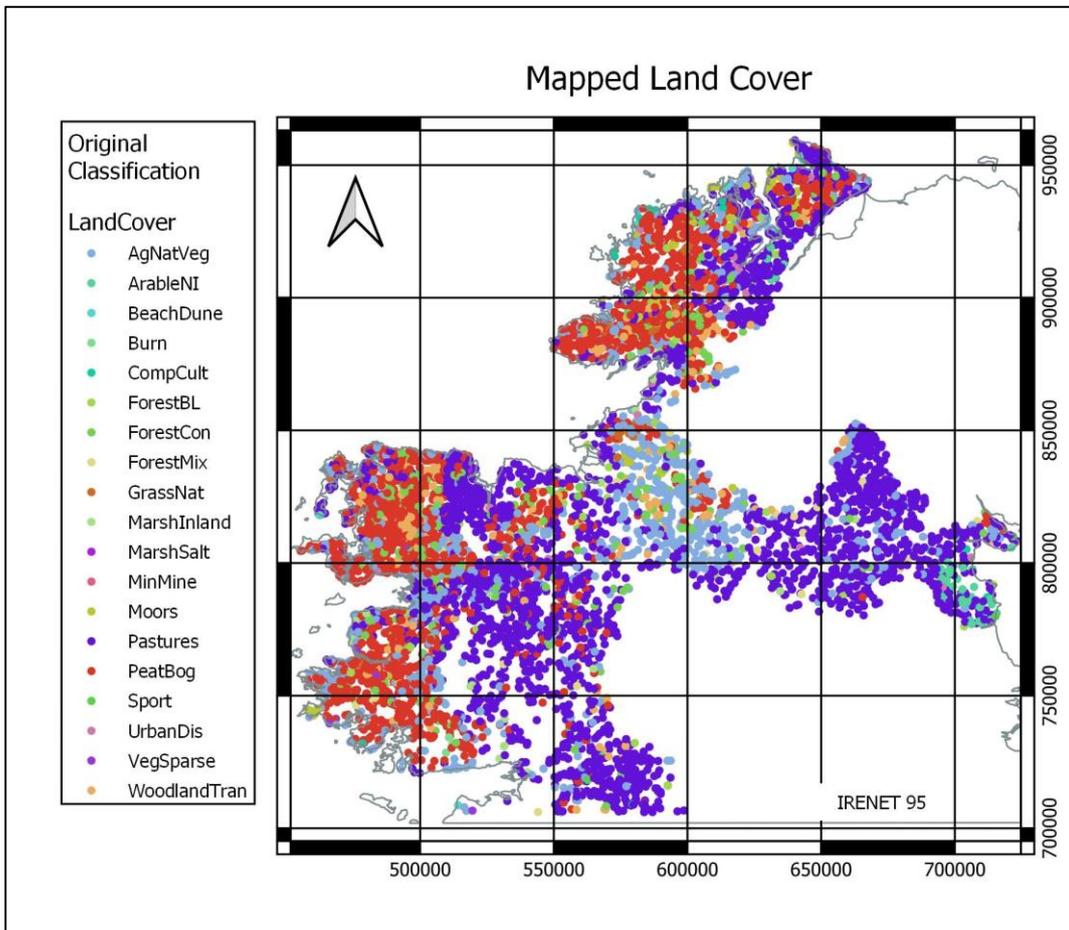


Figure C.11 Stream water sites classified by mapped Land Cover (Corine).



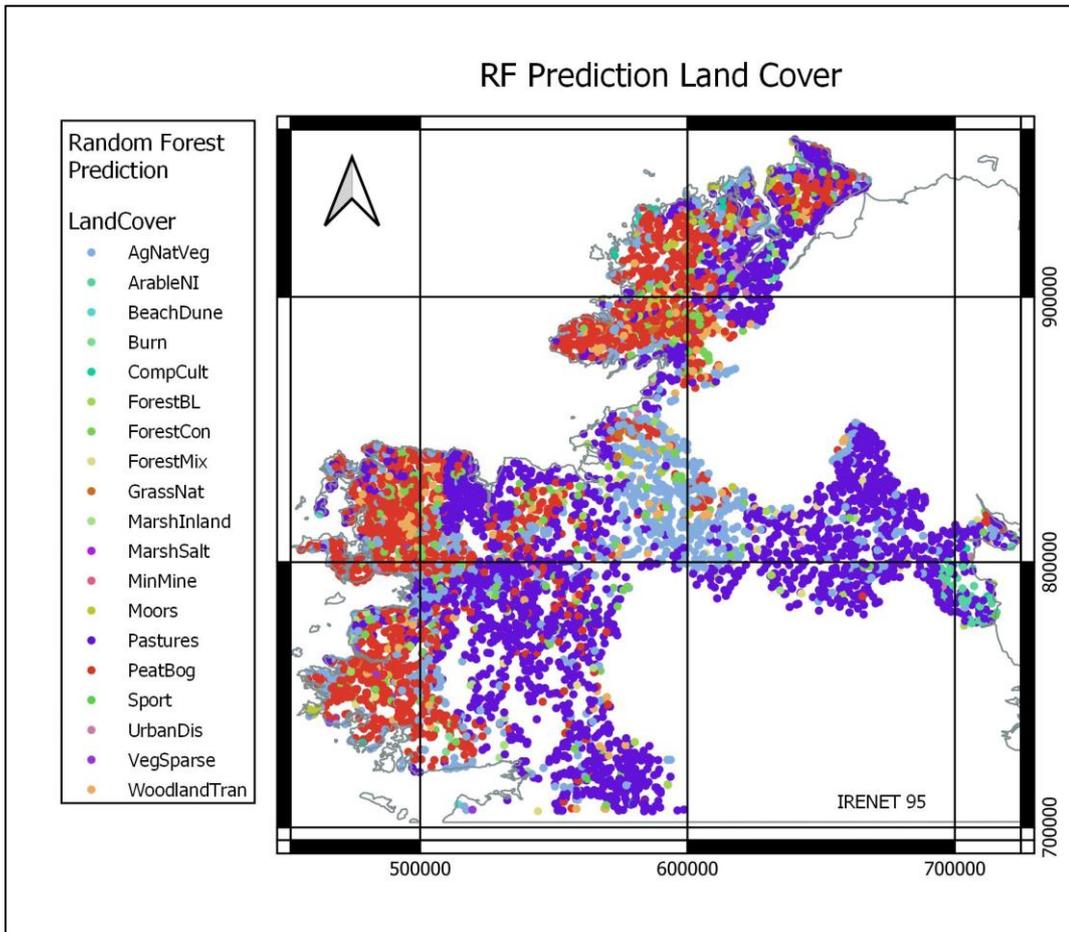


Figure C.12 Stream water sites classified by random forest prediction.



	% Accuracy																			
	AgNatVeg	ArableNI	BeachDune	Burn	CompCult	ForestBL	ForestCon	ForestMix	GrassNat	MarshInland	MarshSalt	MinMine	Moors	Pastures	PeatBog	Sport	UrbanDis	VegSparse	WoodlandTran	
AgNatVeg	10.2	0	0	0	0	0	0	0	0	0	0	0	0	0	55.8	33.8	0	0	0	0.1
ArableNI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
BeachDune	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	10	0	0	0	0
Burn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
CompCult	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47.4	52.6	0	0	0	0
ForestBL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	71.7	28.3	0	0	0	0
ForestCon	6.9	0	0	0	0	0	2	0	0	0	0	0	0	0	25.3	65.8	0	0	0	0
ForestMix	2.5	0	0	0	0	0	0	0	0	0	0	0	0	0	67.5	30	0	0	0	0
GrassNat	1.6	0	0	0	0	0	0	0	0	0	0	0	0	0	35.9	62.5	0	0	0	0
MarshInland	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	10	0	0	0	0
MarshSalt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
MinMine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	50	0	0	0	0
Moors	2.4	0	0	0	0	0	0	0	0	0	0	0	0	0	23.2	74.4	0	0	0	0
Pastures	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	92.5	7.3	0	0	0	0
PeatBog	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	16	83.1	0	0	0	0
Sport	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
UrbanDis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76.9	23.1	0	0	0	0
VegSparse	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12.5	87.5	0	0	0	0
WoodlandTran	5	0	0	0	0	0	0	0	0	0	0	0	0	0	23.1	71.2	0	0	0	0.8
Overall																				
Accuracy	63.46744696																			

Table C.7 RF matrix of % accuracy for Land Cover classes

C.5 Summary of Random Forest Classification

The random forest classifications for the four themes (rock type, subsoil, SRF Geochemical Domains and land cover) indicate varying success in classification accuracies. A significant issue is the under-representation of many classes relative to the dominant classes. In particular, the classes peat bog (Teagasc subsoil, Corine Land Cover) and pasture (Corine Land Cover) display significant compositional variation, overlapping all of the other less dominant classes. Further studies of these themes may generalize some of the classes so that the compositional variance of the newer generalized classes will assist in reducing the influence of the dominant classes.

As an independent check on the validity of the application of Random Forests for the classification of the four themes, linear discriminant analysis was also carried out. The results are similar to those produced by Random Forest analysis but the Random Forests are generally better predictors for individual classes and have better overall accuracy prediction for the themes.



Appendix D: Principal Component analysis of chalcophile mineralization in Tellus stream waters

The accompanying figures and maps display the results of a principal component analysis applied to a suite of chalcophile elements from the Tellus stream water data. Examination of the figures provides additional information and insight into the relationships of the chalcophile elements with subsoil and bedrock with an emphasis on the detection of base metal mineralization.

Additional measures were added to assist in decoupling the surficial and bedrock / mineralized environment:

- A measure of Zn vs 100*Cd indicates association with base metal mineralization when Zn and Cd exceed their respective 98th percentile values (Zn \geq 20 ppb and 100*Cd \geq 10 ppb) (see Section 3).
- An additional measure of “distance to closest mineral occurrence” shows which waters sampling sites are close to known mineral occurrences.

D.1 Screeplot and PC loadings

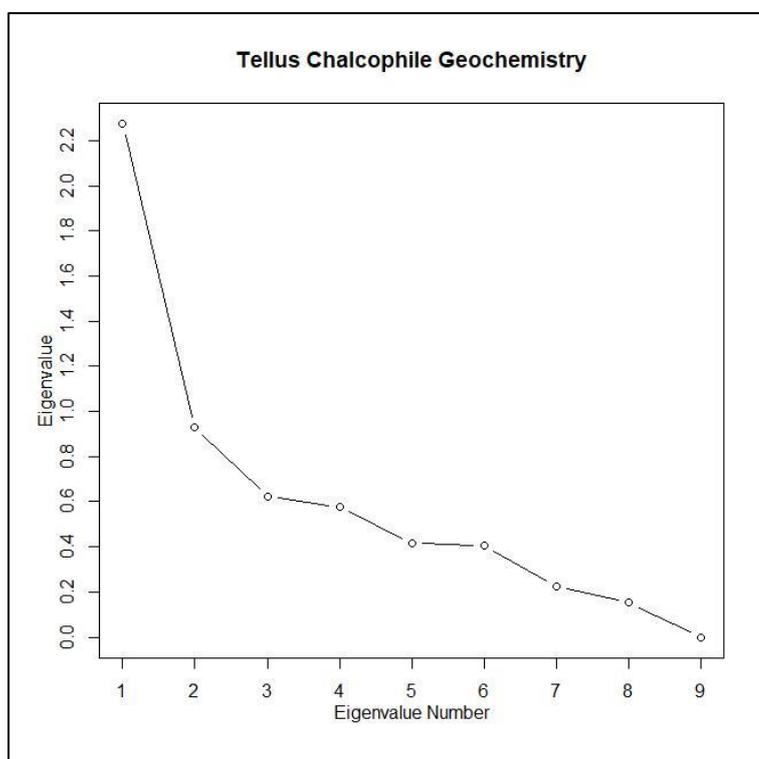


Figure D.1 Ordered screeplot of PCA applied to chalcophile elements



Figure D.1 shows an ordered screeplot of the PCA, where it can be observed that eight principal components account for all of the variability of the data. This is demonstrated quantitatively in Table D.1 where the first eight components account for 100 % of the variability of the data. Note that the number of principal components is one less than the number of elements. This is due to the nature of the logcentred transform, the last of which is zero, and reflects the closure property of compositional data.

Tellus Chalcophile Geochemistry								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
λ	2.27	0.93	0.62	0.58	0.42	0.4	0.22	0.15
$\lambda\%$	40.6082	16.6369	11.0912	10.3757	7.5134	7.1556	3.9356	2.6834
$\Sigma\lambda\%$	40.6082	57.2451	68.3363	78.712	86.2254	93.381	97.3166	100
R-Scores								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cu	-0.3103	-0.2808	-0.2614	-0.406	-0.2721	0.2447	-0.0154	0.0773
Zn	0.4721	-0.2401	0.2711	-0.3182	0.4029	0.0853	0.0723	-0.0258
As	0.0027	0.2284	-0.3697	0.3206	0.2647	0.3001	-0.0941	0.0274
Se	-0.2125	0.1027	0.0457	0.1305	0.0128	-0.1767	0.2533	0.2642
S	-0.7106	0.1314	0.0983	-0.1296	0.1155	-0.2826	-0.2888	0.0166
Cd	-0.085	-0.5796	0.2921	0.4159	-0.145	0.0625	-0.0622	-0.0392
Sn	0.1668	0.613	0.3806	-0.0314	-0.2296	0.1892	0.0078	-0.059
Sb	-0.4168	0.0251	-0.22	0.0134	9.00E-04	-0.1537	0.2223	-0.2626
Pb	1.0937	-2.00E-04	-0.2367	0.0048	-0.1501	-0.2687	-0.0951	0.001
Relative Contributions								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cu	17.5561	14.3792	12.4601	30.0566	13.4982	10.9179	0.0433	1.0886
Zn	35.3331	9.1401	11.6549	16.0529	25.7322	1.1529	0.8286	0.1053
As	0.0015	11.3079	29.6293	22.2745	15.1848	19.52	1.9189	0.163
Se	18.803	4.3913	0.8694	7.0953	0.0681	13.0028	26.7122	29.0579
S	69.5926	2.3809	1.3328	2.315	1.8385	11.0054	11.4967	0.0382
Cd	1.1431	53.1753	13.5025	27.3762	3.3292	0.6178	0.6128	0.2431
Sn	4.3372	58.5805	22.5815	0.1537	8.2163	5.5779	0.0095	0.5434
Sb	47.6073	0.1732	13.2677	0.0492	2.00E-04	6.4738	13.5382	18.8904
Pb	88.2148	0	4.1321	0.0017	1.6611	5.3229	0.6673	1.00E-04
Absolute Contributions								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cu	4.233	8.4883	10.9335	28.5724	17.7866	14.827	0.106	3.9422
Zn	9.7988	6.2059	11.7629	17.5521	38.9997	1.8008	2.33	0.4387
As	3.00E-04	5.6163	21.8747	17.8155	16.8349	22.3036	3.9471	0.4965
Se	1.986	1.1355	0.3342	2.9547	0.0393	7.7353	28.6073	46.0967
S	22.2057	1.8599	1.5476	2.9123	3.206	19.7786	37.1957	0.1831
Cd	0.3176	36.17	13.6523	29.9871	5.0548	0.9667	1.7262	1.0142
Sn	1.2234	40.456	23.1812	0.1709	12.666	8.862	0.0273	2.302
Sb	7.639	0.068	7.7477	0.0311	2.00E-04	5.8507	22.0262	45.5259
Pb	52.5962	0	8.966	0.0039	5.4125	17.8754	4.0342	7.00E-04

Table D.1 Eigenvalues, R-scores, relative and absolute contributions for first nine Principal Components



D.2 PC analysis of chalcophile elements classified by rock type: Zn-Cd Enrichment

The PC biplots below show the results of PC analysis of stream waters classified by bedrock type with additional classification based on abundance of Zn and Cd. Negative PC2 scores, in particular, is useful for discriminating high concentrations of Zn and Cd in stream waters draining all rock types while positive PC5 scores reflect high Zn concentrations in stream waters draining limestone.

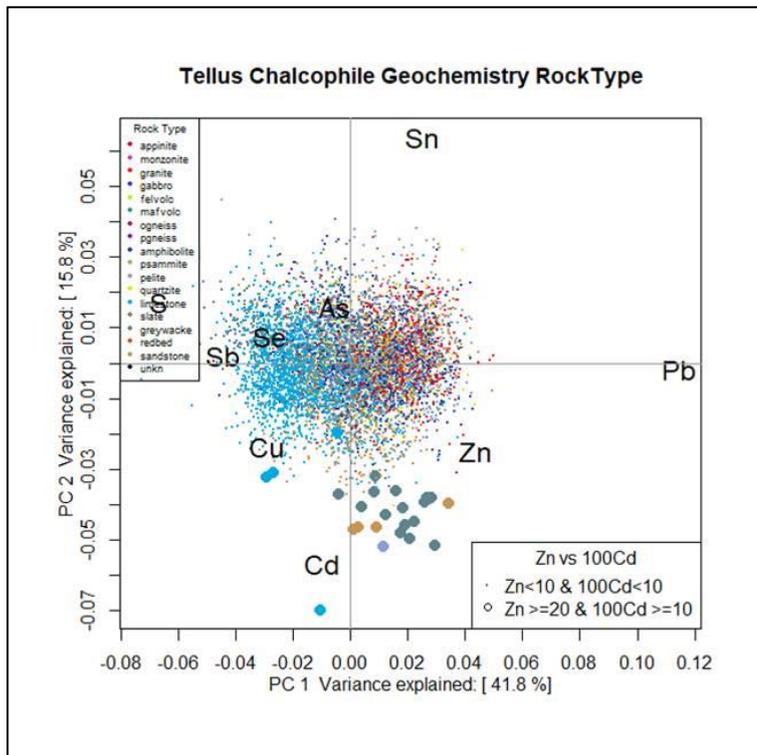


Figure D.2 PC1 v PC2 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance. Lithologies with higher abundances of Cd-Zn-Cu include Carboniferous limestone, Lower Palaeozoic sediments and Namurian sediments.



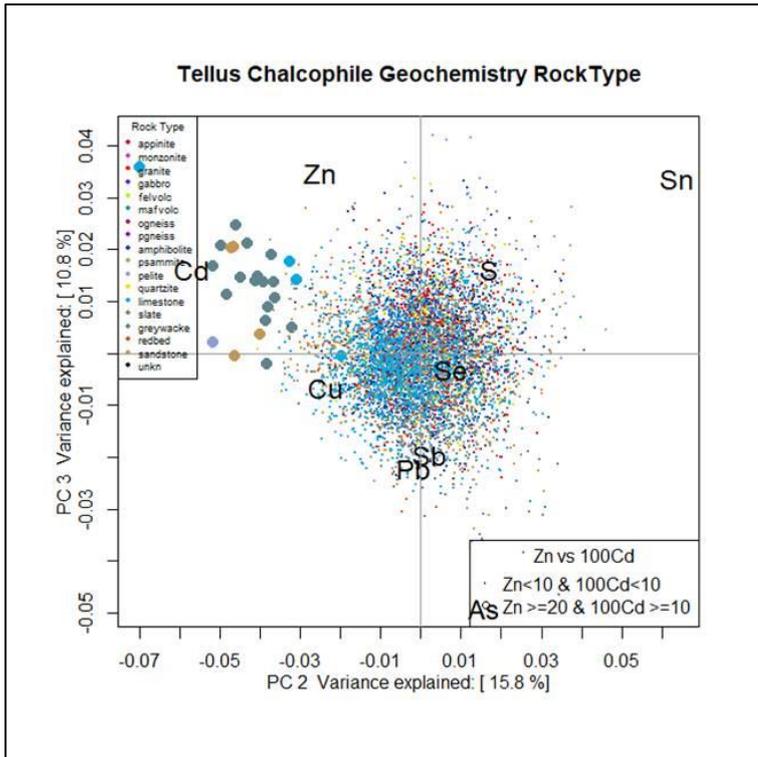


Figure D.3 PC2 v PC3 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance.

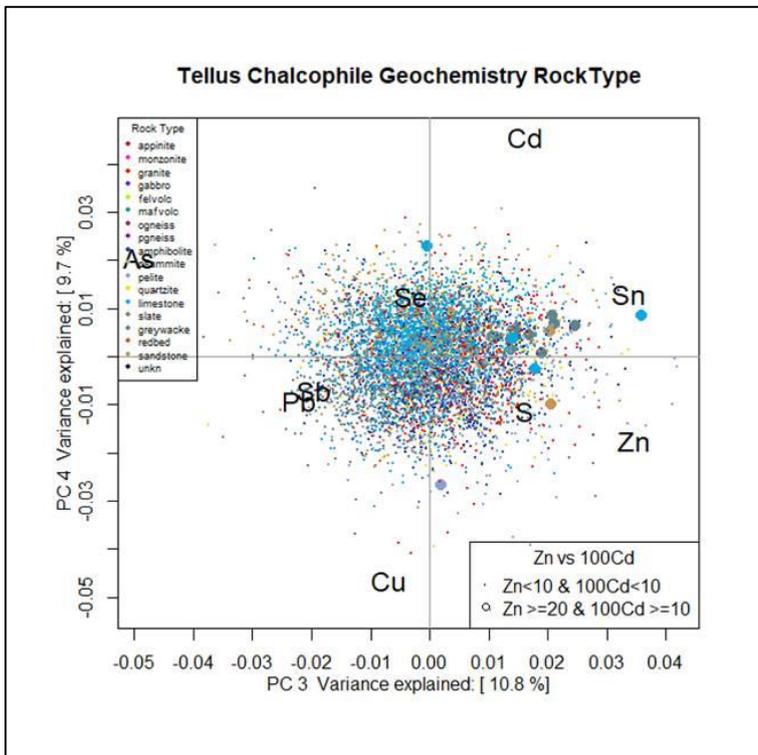


Figure D.4 PC3 v PC4 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance



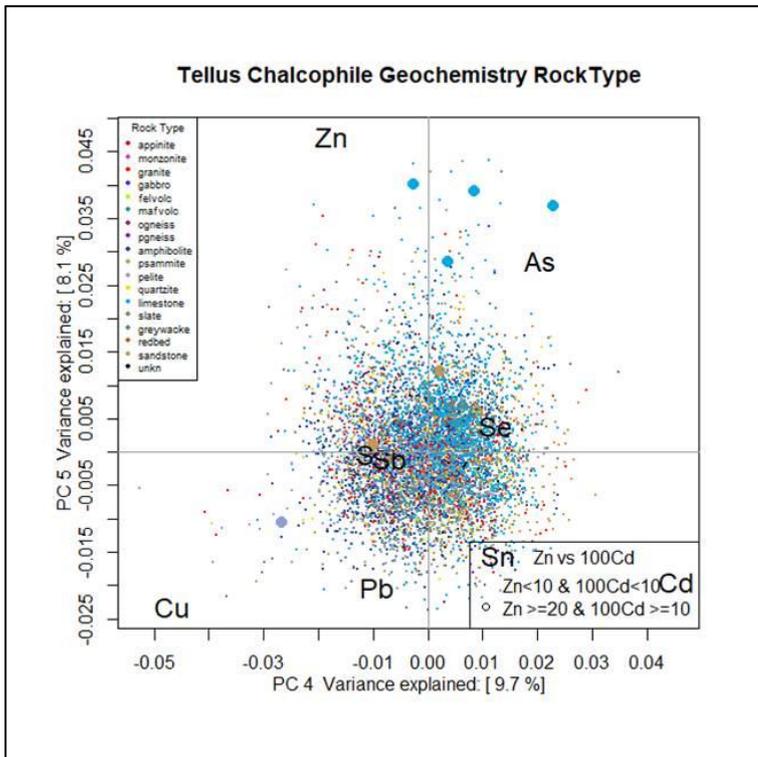


Figure D.5 PC4 v PC5 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance. Positive PC5 scores reflect high Zn in stream water draining Carboniferous limestone.

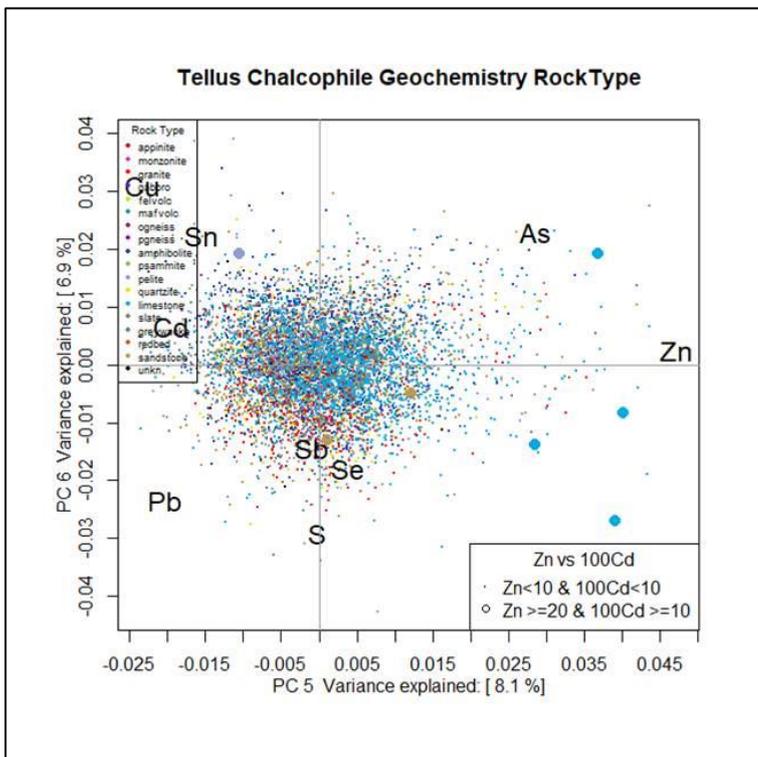


Figure D.6 PC5 v PC6 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance. Positive PC5 scores reflect high Zn in stream water draining Carboniferous limestone.

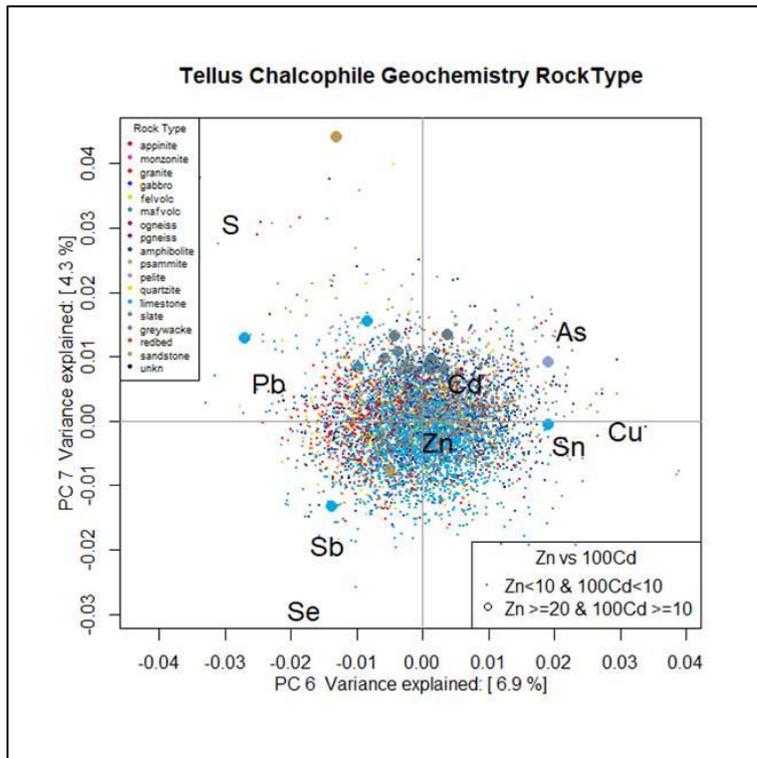


Figure D.7 PC6 v PC7 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance.

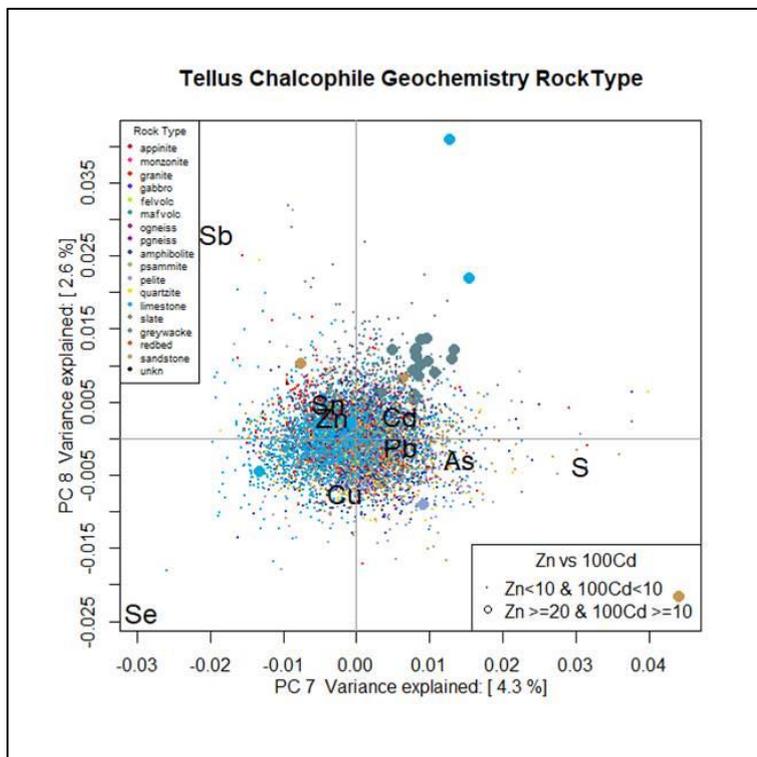


Figure D.8 PC7 v PC8 biplot of chalcophile elements for stream water classified by rock type and by Zn-Cd abundance.

D.3 PC analysis of chalcophile elements classified by rock type: distance to mineral localities (MinLocs)

The GSI Mineral Localities database (MinLocs) includes known occurrences of mineralization across Ireland, from minor shows to active and historic mines. The PC biplots below show the results of a PCA of stream water data classified by bedrock type with additional classification based on proximity to the nearest MinLocs occurrence that includes chalcophile element(s) as a component.

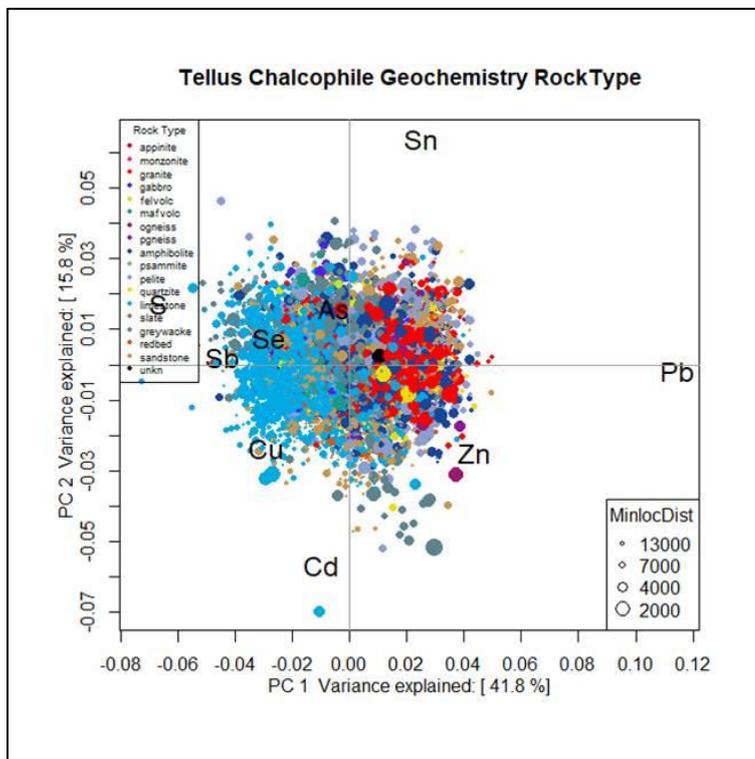


Figure D.9 PC1 v PC2 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality. The shortest distances to mineral localities with high Zn-Cd are for stream water draining Lower Palaeozoic greywackes and other sedimentary rocks in the area of Pb-Zn mineralization in County Monaghan but limestone-related mineralization is also highlighted.



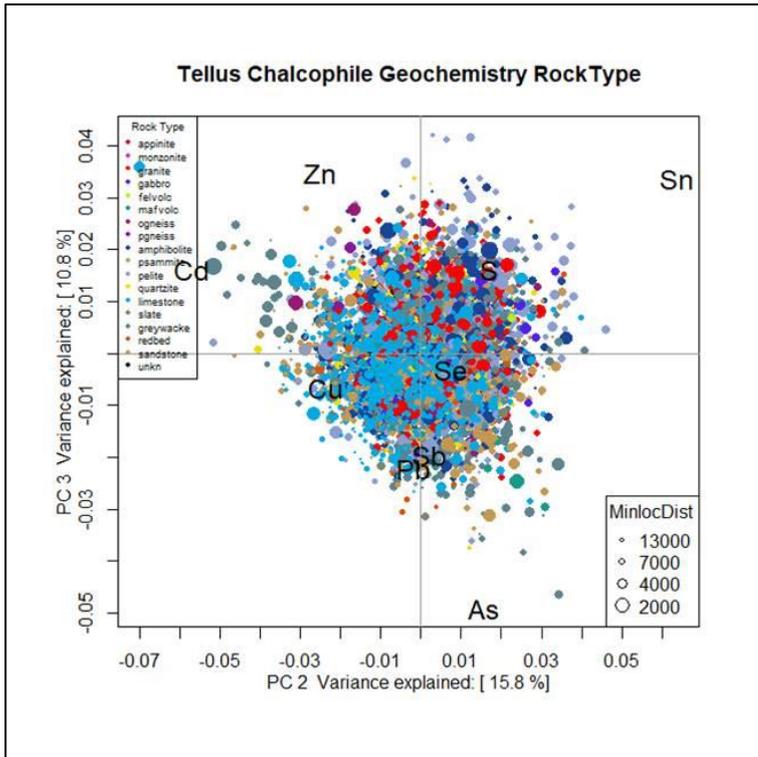


Figure D.10 PC2 v PC3 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality. Areas underlain by Lower Palaeozoic greywacke and limestone show proximity with mineral occurrences associated with Cd and Zn.

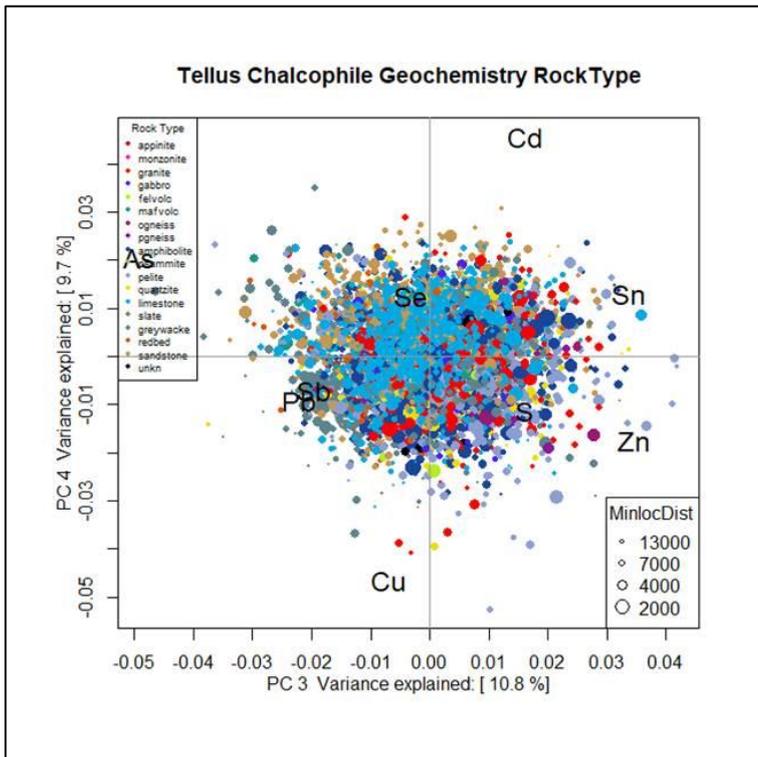


Figure D.11 PC3 v PC4 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality.



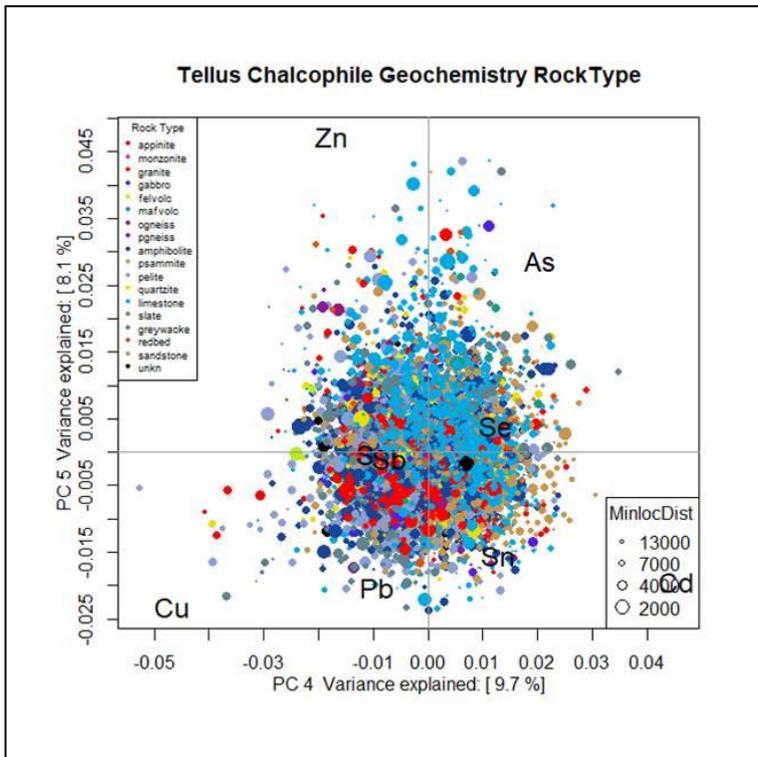


Figure D.12 PC4 v PC5 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality. Sites underlain by limestone show relative enrichment in Zn-As with a closer proximity to mineral occurrences (positive PC5).

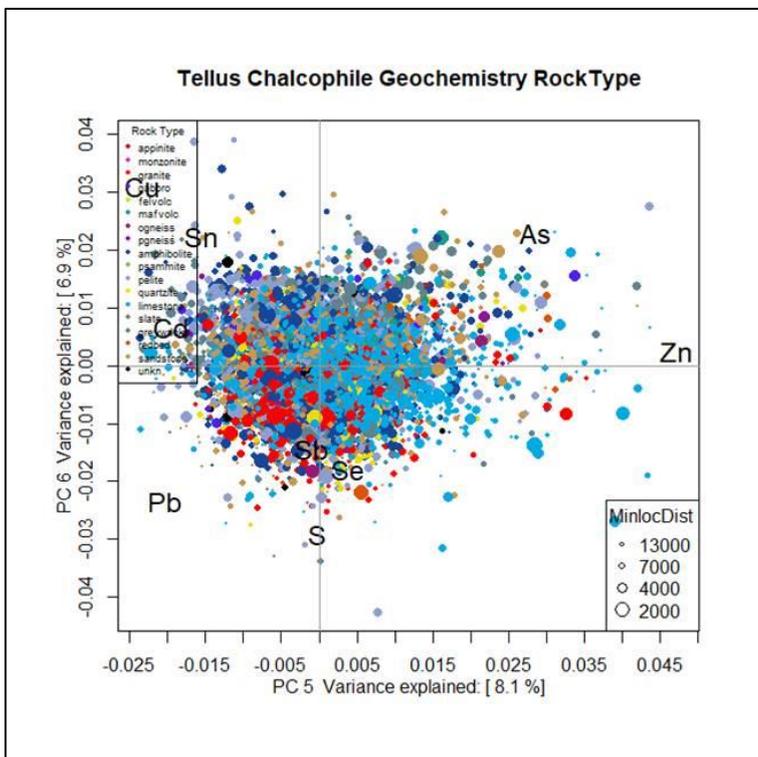


Figure D.13 PC5 v PC6 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality.

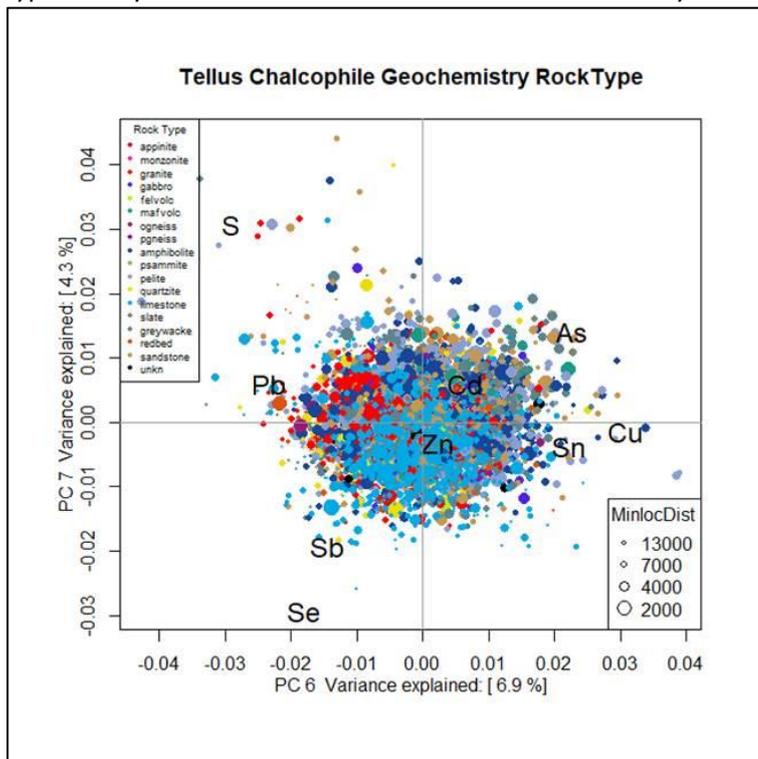


Figure D.14 PC6 v PC7 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality.

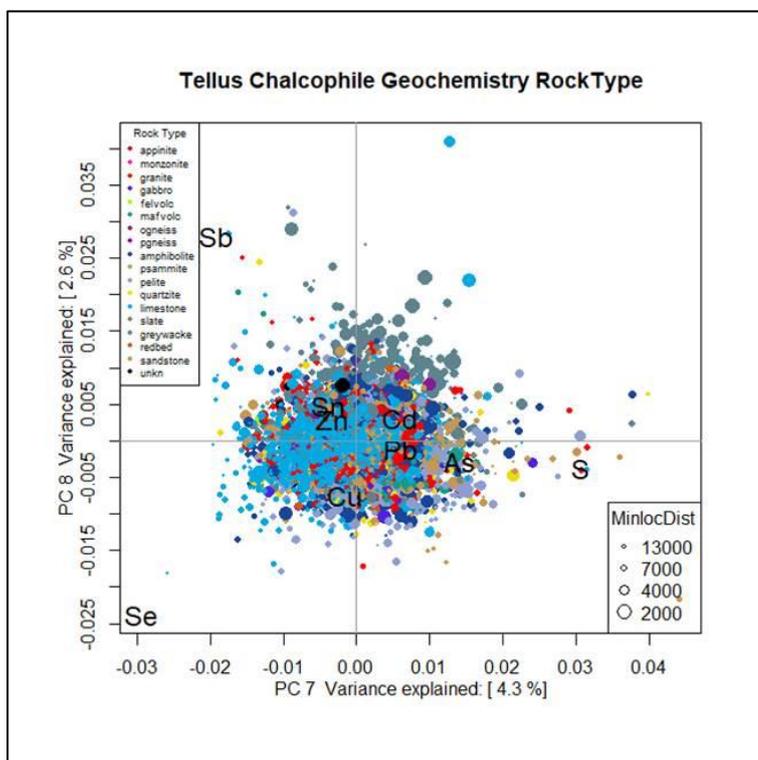


Figure D.15 PC6 v PC7 biplot of chalcophile elements for stream water classified by rock type and by distance to nearest recorded mineral locality. Stream water sampled from



sites underlain by Lower Palaeozoic sediments and with a relative enrichment in Sb (positive PC8) have a closer proximity to mineral occurrences. This is largely a reflection of Sb mineralization in the County Monaghan area.

D.4 PC analysis of chalcophile elements classified by SRF Geochemical Domain: Zn-Cd Enrichment

The PC biplots coded by the Teagasc subsoil classification can be difficult to interpret, primarily owing to the very large number of subsoil classes and the overwhelming number of sites that drain blanket peat and cut peat. An alternative approach is to utilize the SRF Geochemical Domains for classification since these domains are derived mainly from the Teagasc subsoil classes through amalgamation of these classes, with some input from bedrock geology where the subsoil is not classified, e.g. where bedrock is within 1 m of the surface. Peat or peaty subsoil is not accounted for in this classification.

Figure D16 shows the PC1 v PC2 biplot for the SRF Geochemical Domains, indicating strong association between Zn-Cd enrichment and Domain 5 (Lower Palaeozoic sediments), Domain 2 (Carboniferous limestone) and Domain 1 (Namurian rocks). The association with Domain 5 reflects the large area of Zn and Cd geochemical anomalies in County Monaghan, in the historic lead mining district, while that related to Domain 2 is primarily a reflection of anomalies around Tynagh mine in County Galway.

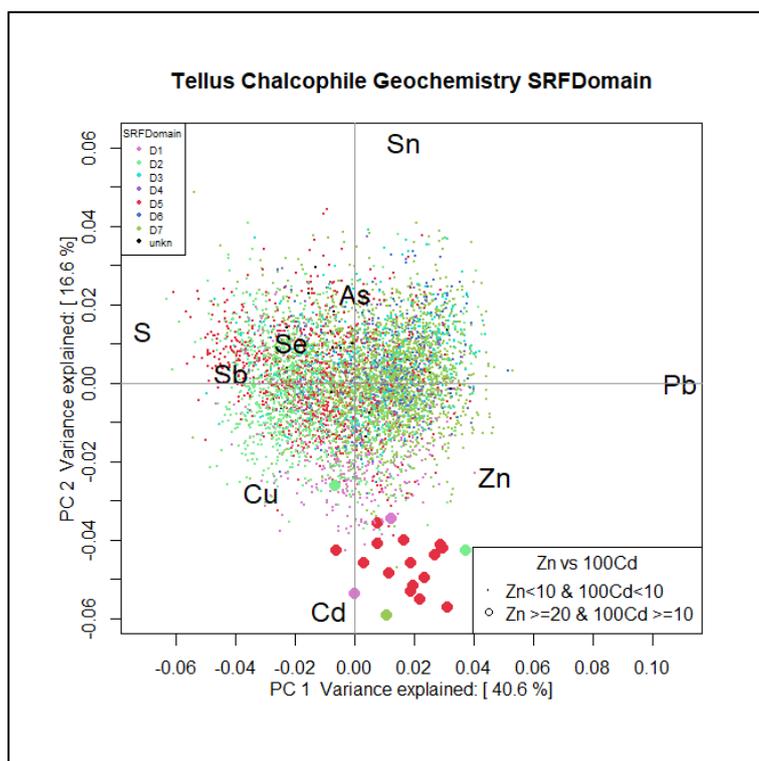


Figure D.16 PC1 v PC2 biplot of chalcophile elements for stream water classified by SRF Geochemical Domain and by relative Zn-Cd enrichment.



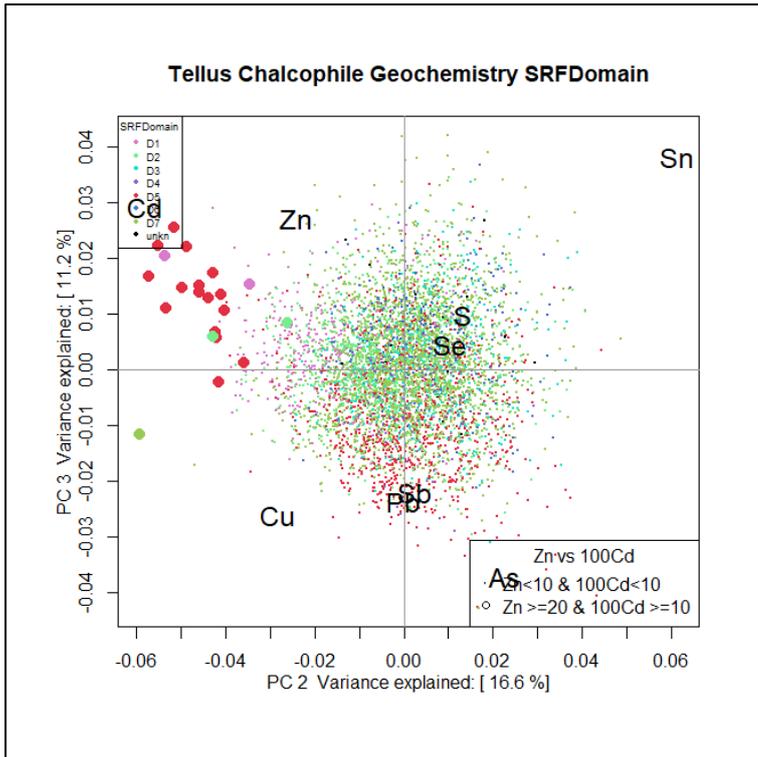


Figure D.17 PC2 v PC3 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and relative Zn-Cd enrichment.

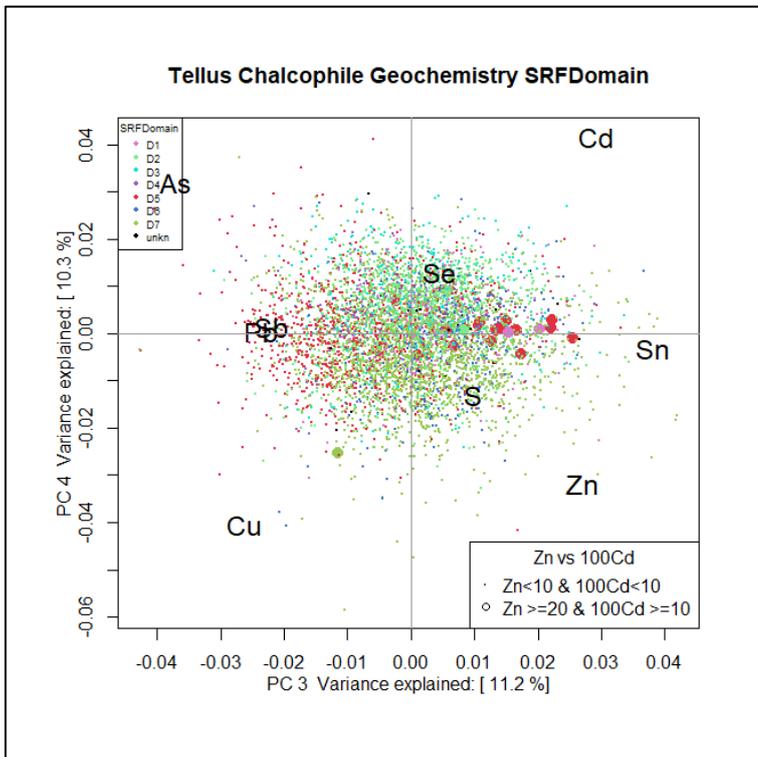


Figure D.18 PC3 v PC4 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by relative Zn-Cd enrichment.



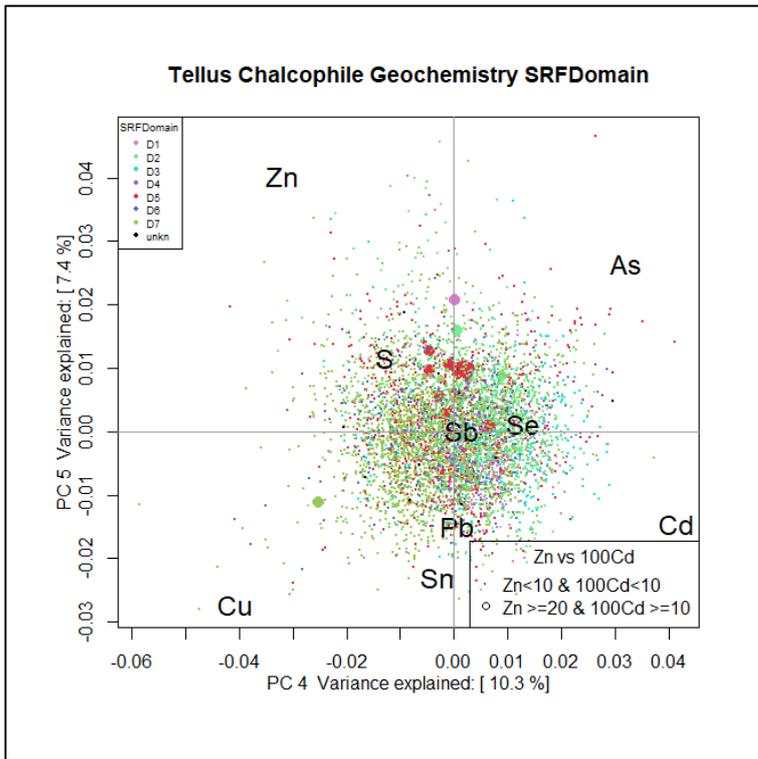


Figure D.19 PC4 v PC5 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by relative Zn-Cd enrichment.

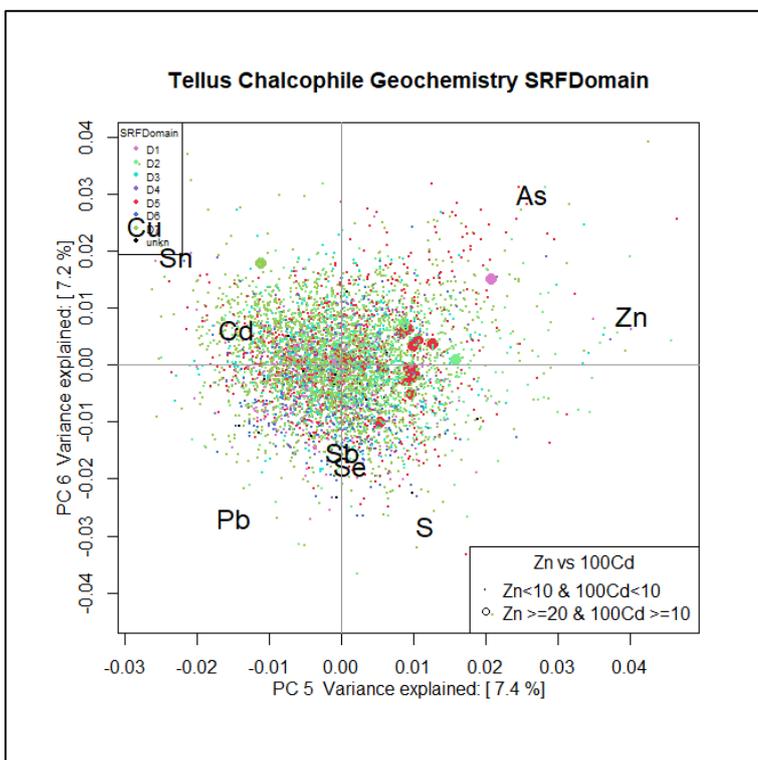


Figure D.20 PC5 v PC6 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by relative Zn-Cd enrichment.

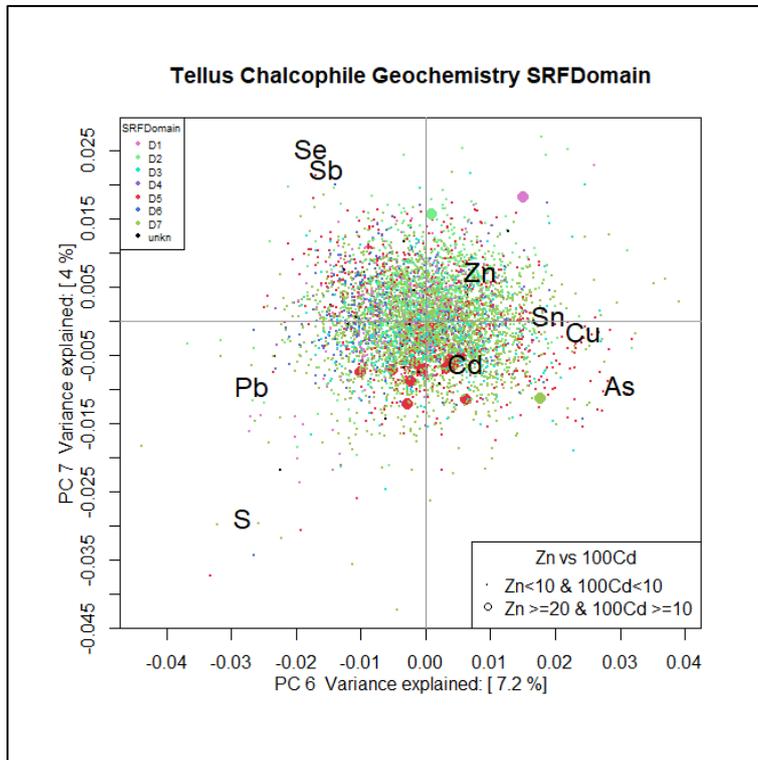


Figure D.21 PC6 v PC7 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by relative Zn-Cd enrichment.



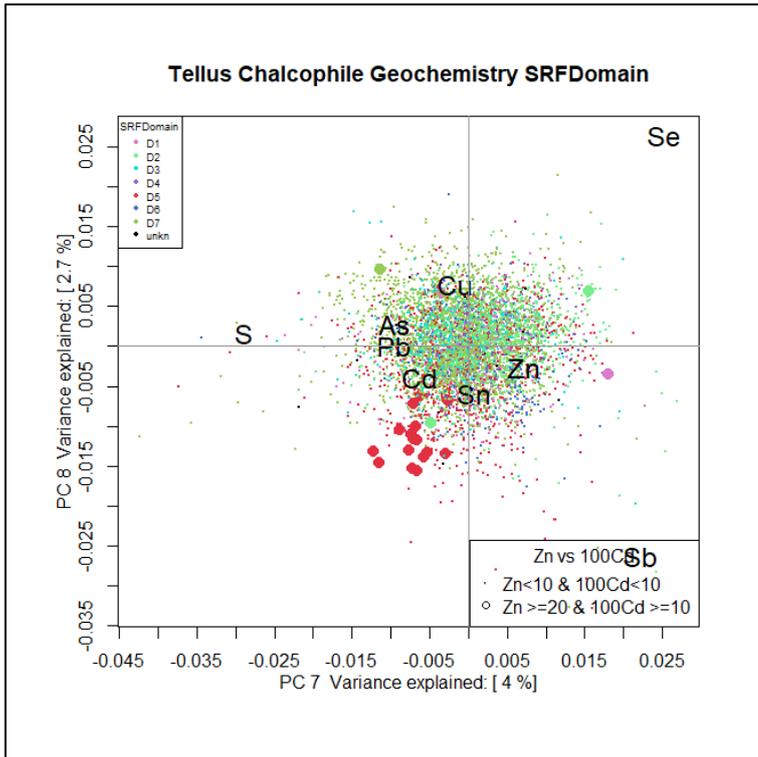


Figure D.22 PC7 v PC8 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by relative Zn-Cd enrichment.

D.5 PC analysis of chalcophile elements classified by SRF Geochemical Domain: distance to mineral localities (MinLocs)

The GSI Mineral Localities database (MinLocs) includes known occurrences of mineralization across Ireland, from minor shows to active and historic mines. The PC biplots below show the results of PC analysis of stream waters classified by SRF Geochemical Domain with additional classification based on proximity to the nearest MinLocs occurrence that includes chalcophile element(s) as a component.



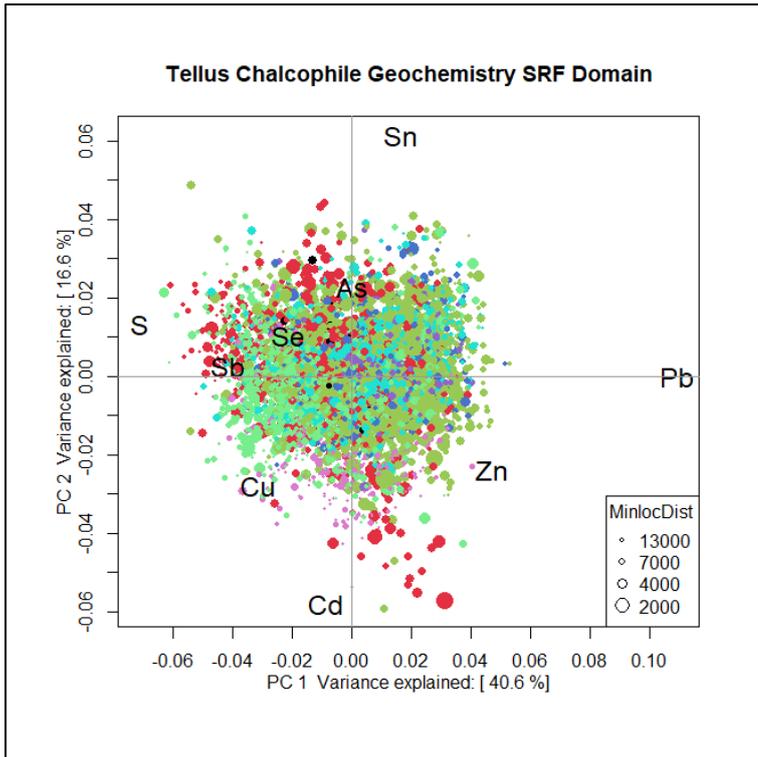


Figure D.23 PC1 v PC2 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.

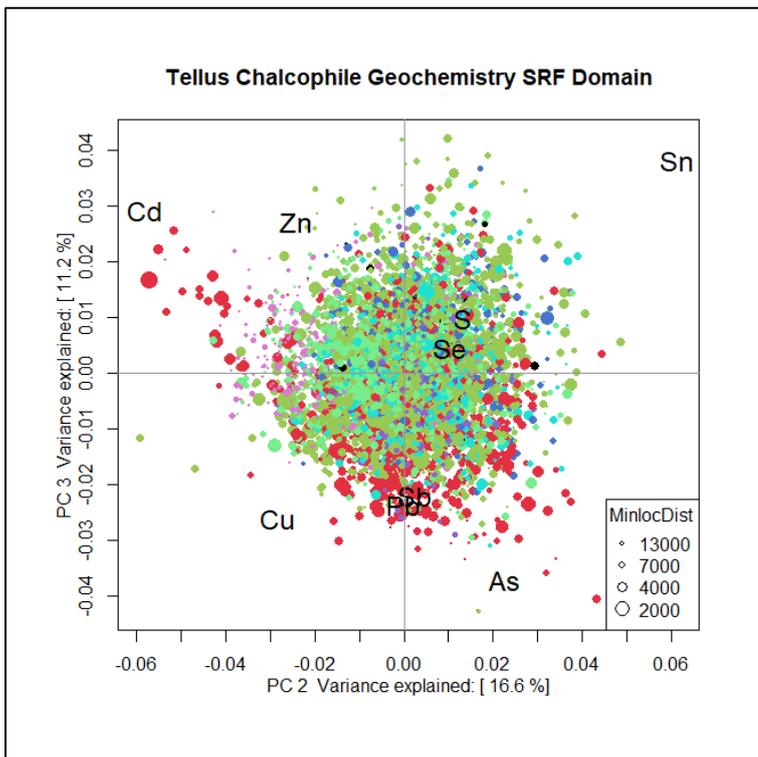


Figure D.24 PC2 v PC3 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.

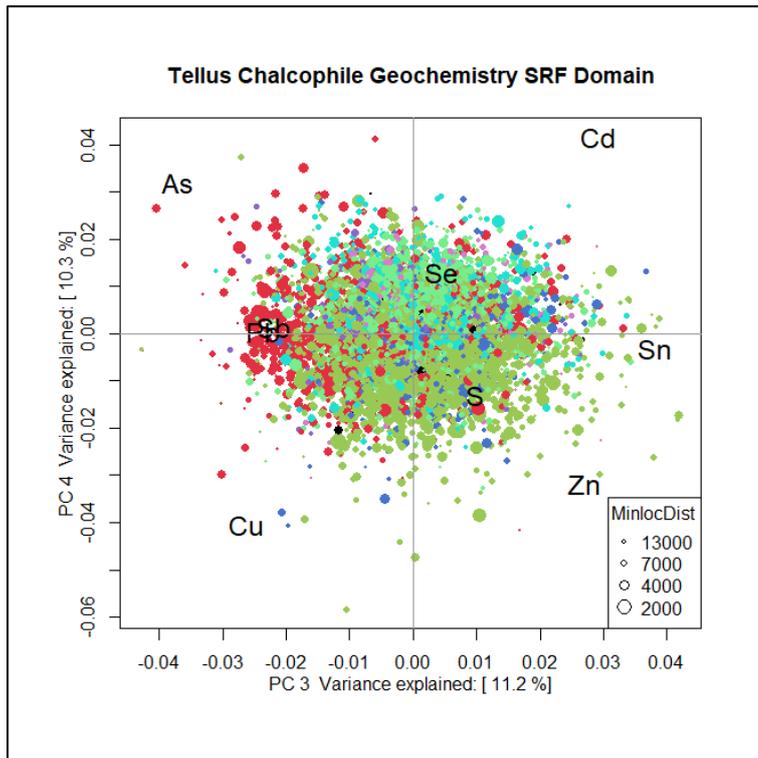


Figure D.25 PC3 v PC4 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.



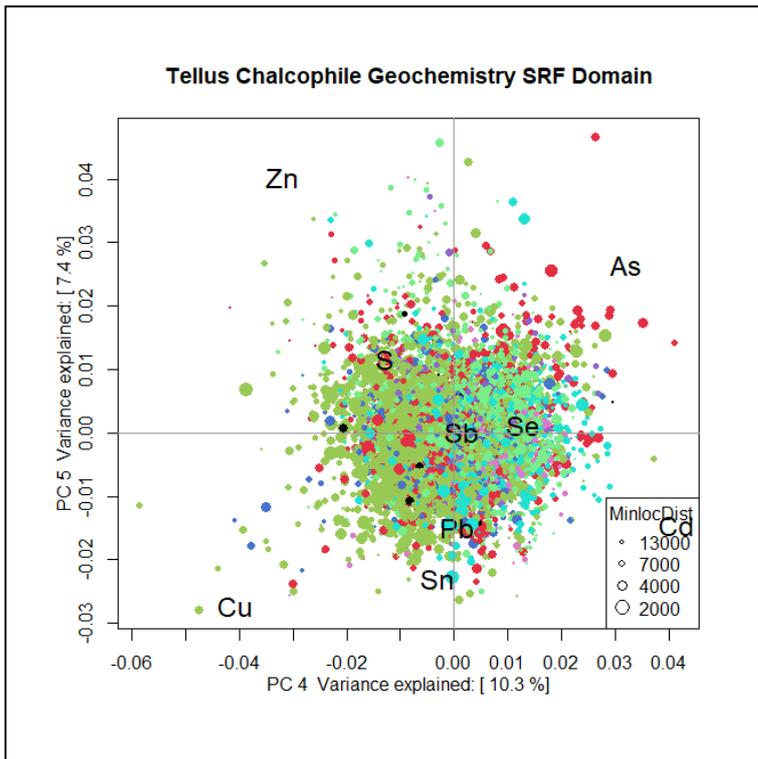


Figure D.26 PC4 v PC5 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.

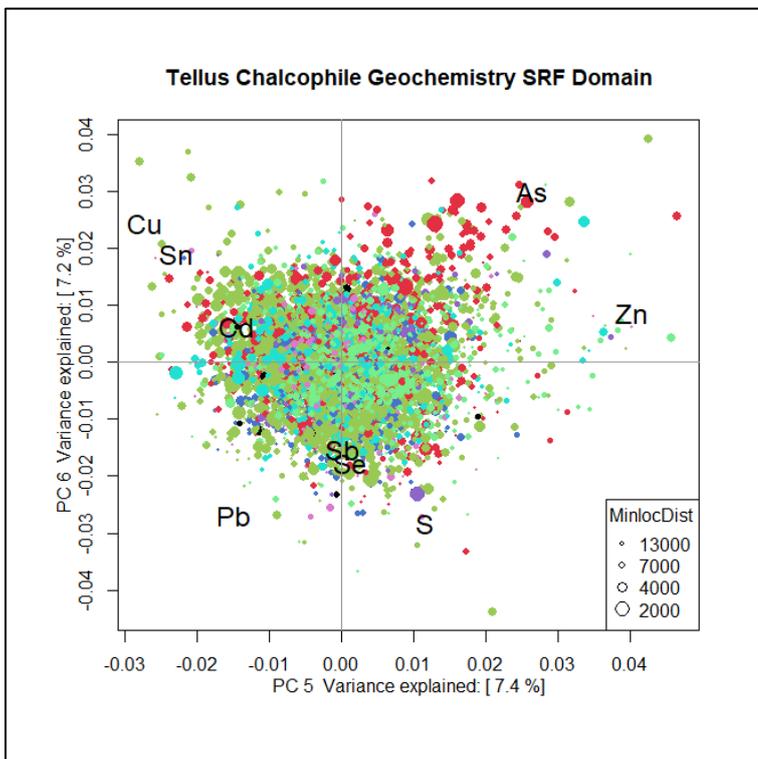


Figure D.27 PC5 v PC6 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.



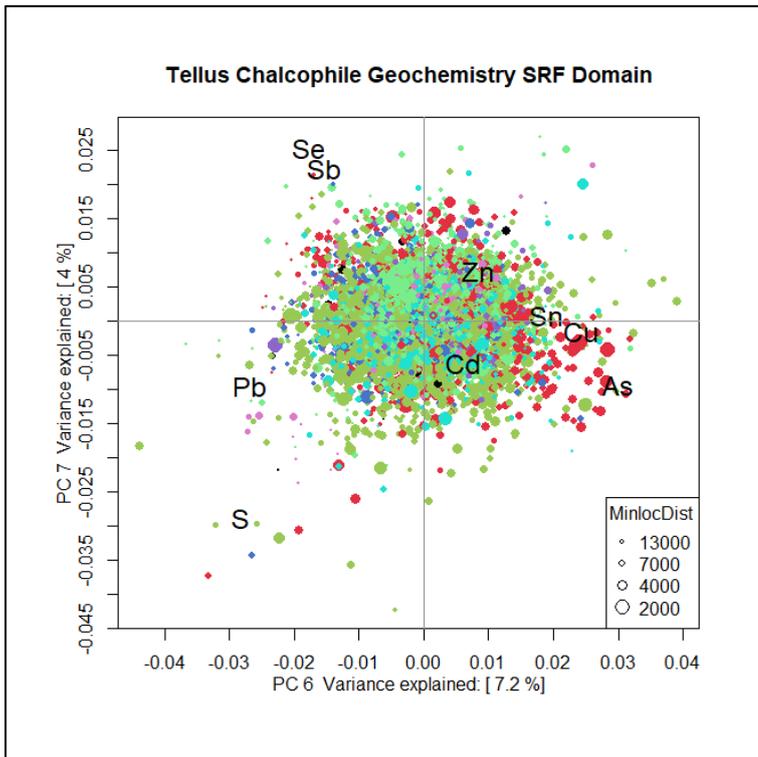


Figure D.28 PC6 v PC7 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.

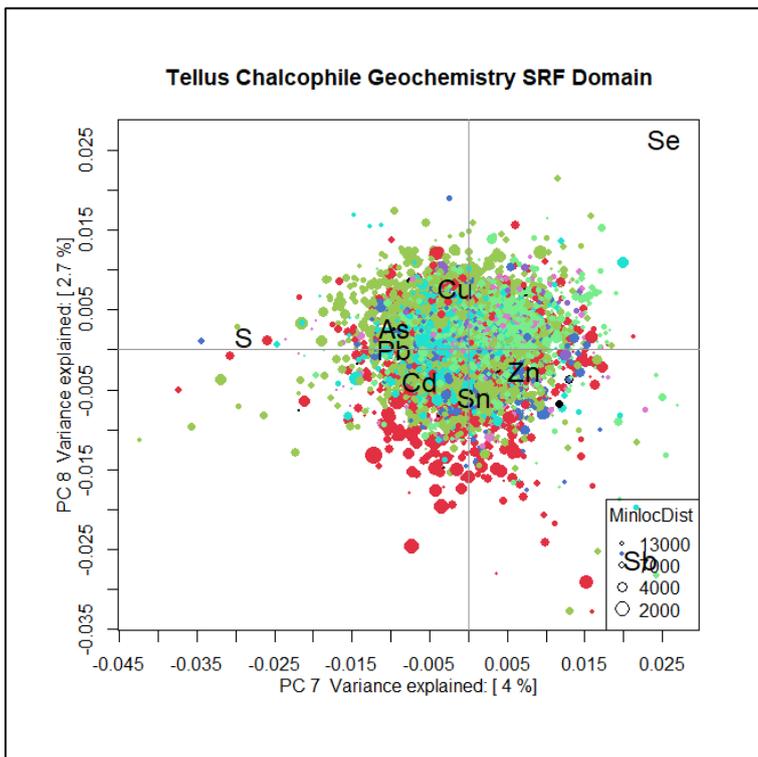


Figure D.29 PC7 v PC8 biplot of chalcophile elements for stream water classified by Teagasc subsoil type and by proximity to nearest recorded mineral locality. Legend as for Figures D.16 – D22.

D.6 Maps of Principal Components

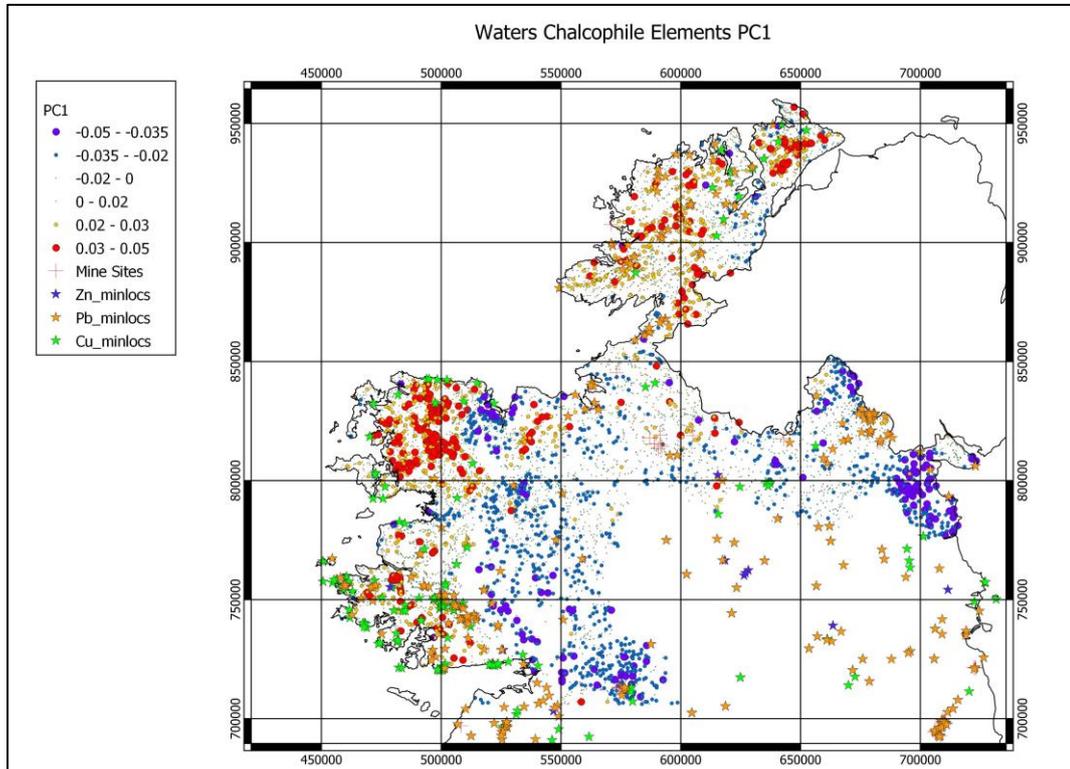


Figure D.30 Map of PC1 showing mineral localities



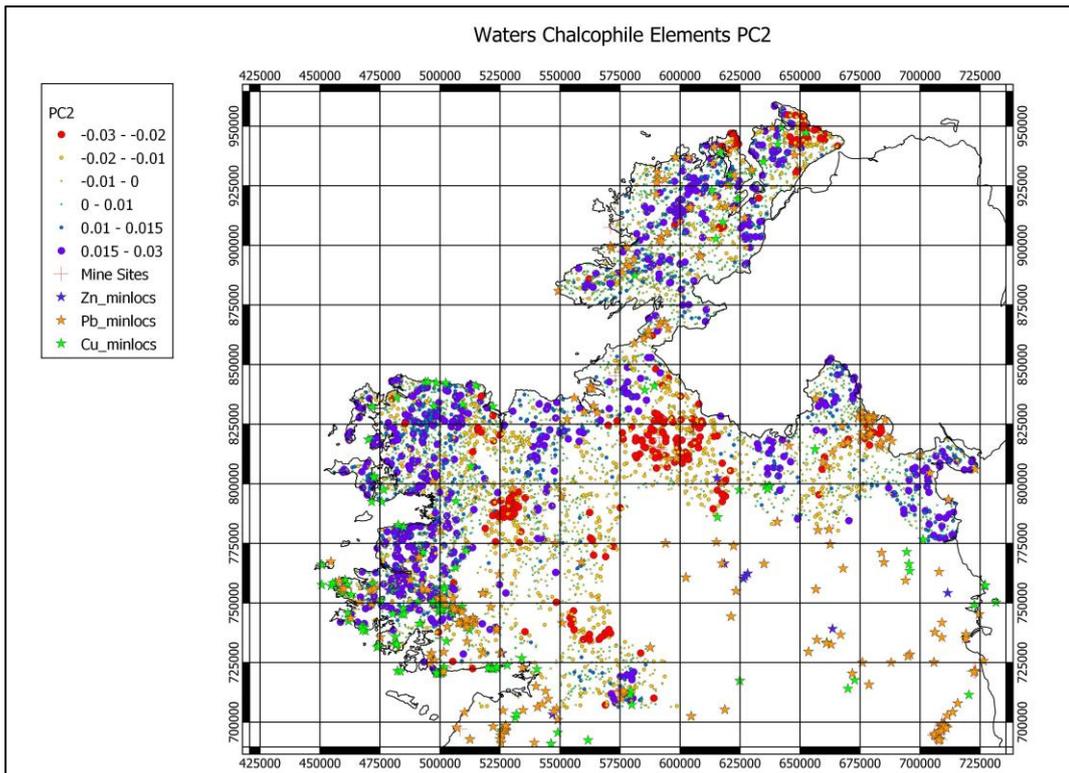


Figure D.31 Map of PC 2 showing mineral localities

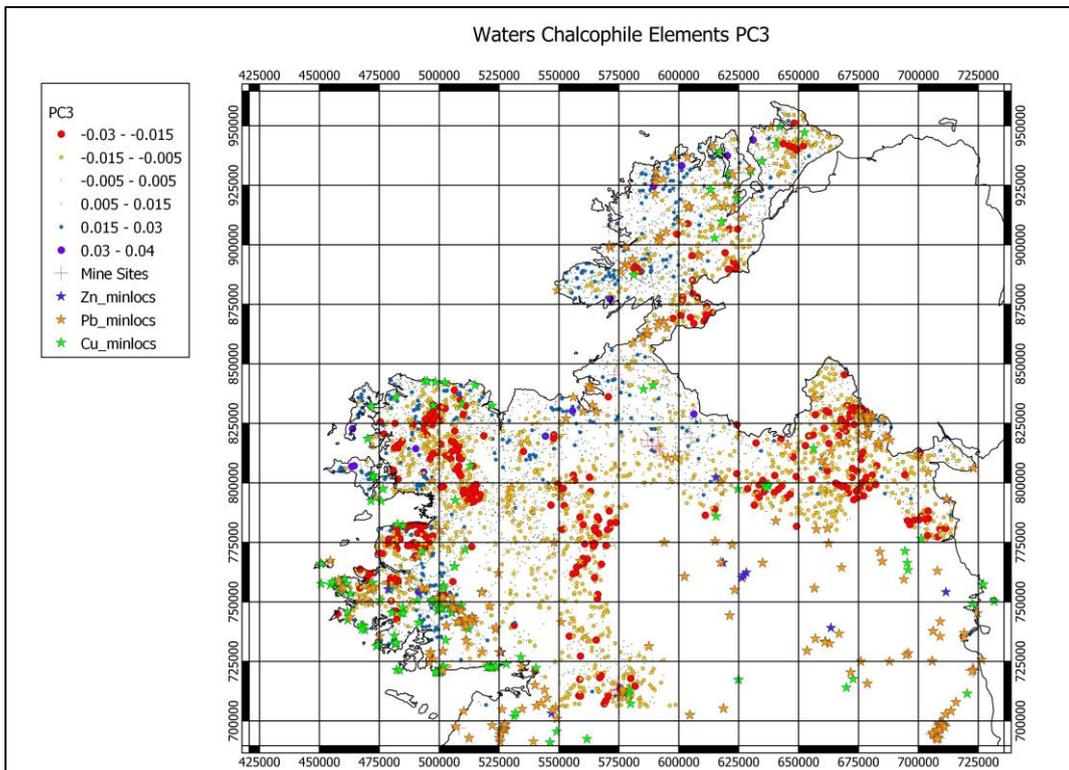


Figure D.32 Map of PC 3 showing mineral localities



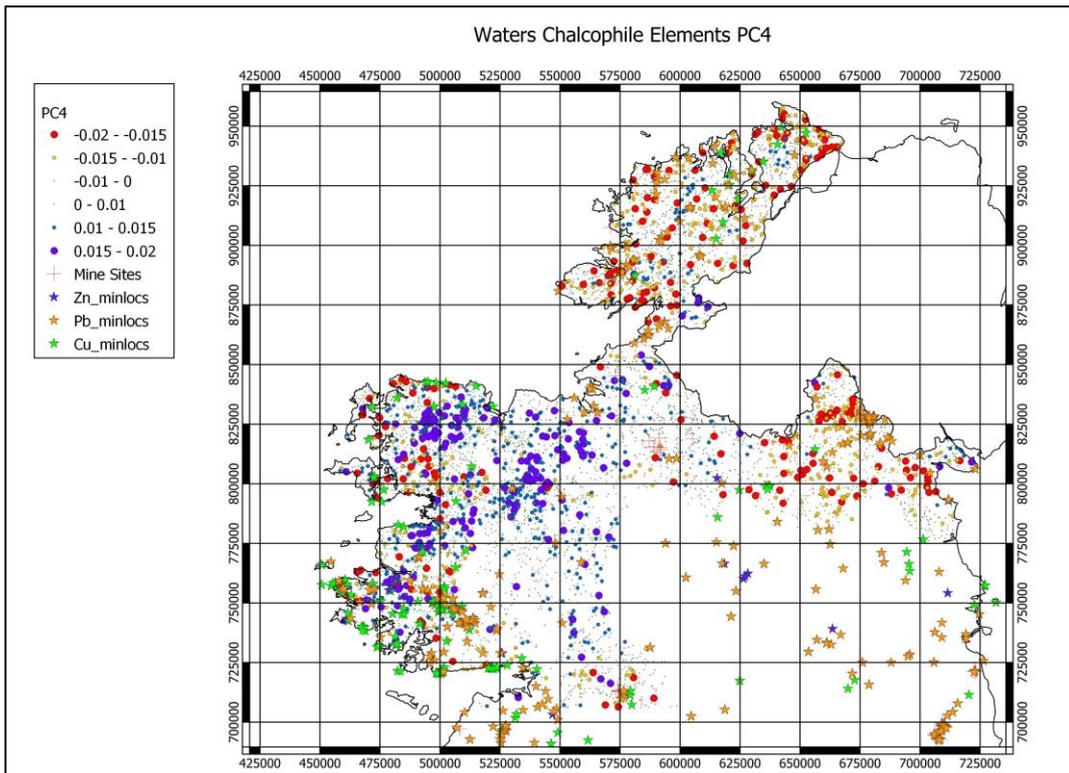


Figure D.33 Map of PC 4 showing mineral localities

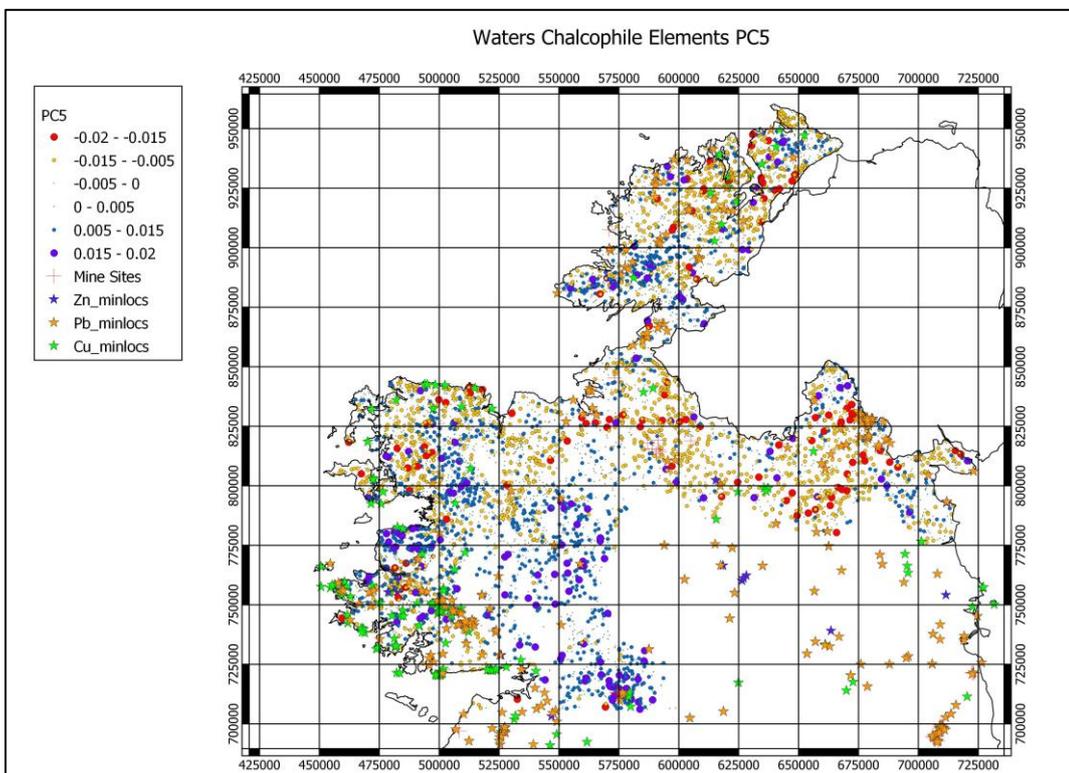


Figure D.34 Map of PC 5 showing mineral localities



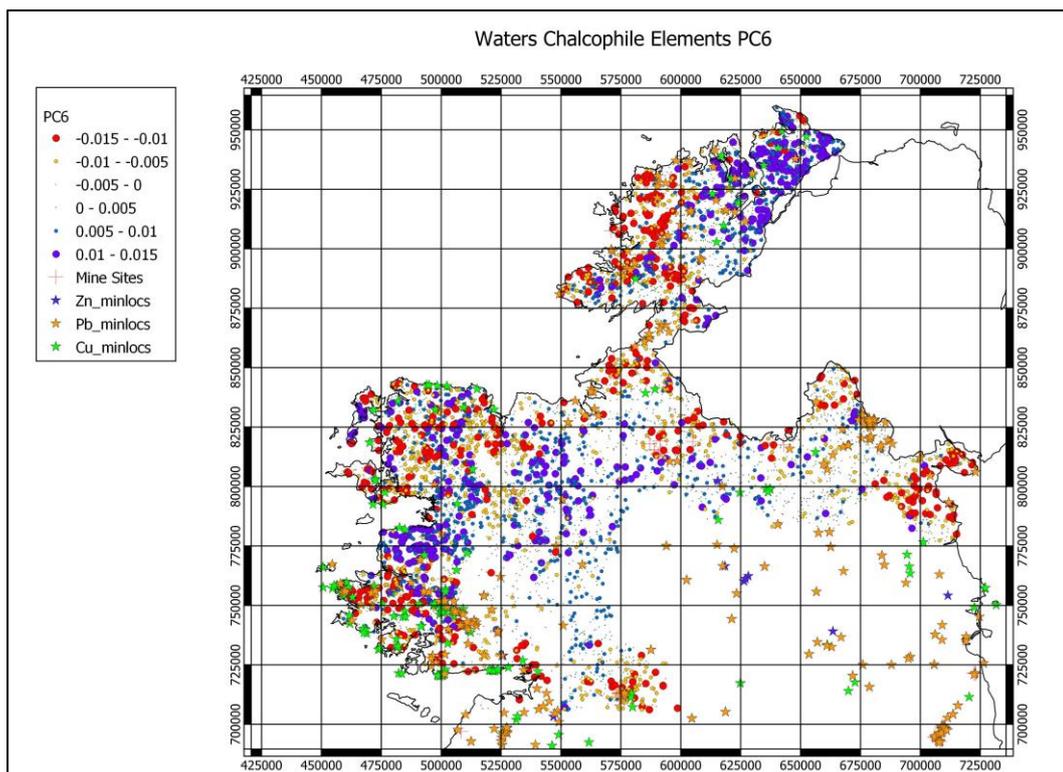


Figure D.35 Map of PC 6 showing mineral localities

D.7 Maps of Proximity of Stream Water Sampling Sites to Mineral Occurrences

The maps below (Figures D.36 to D.38) provide a measure of proximity of the stream water sampling sites to the known base metal occurrences in the MinLocs database. MinLocs includes sites ranging from minor shows of mineralization to major prospects and abandoned mines. The maps demonstrate the potential to model mineral potential based on existing databases.



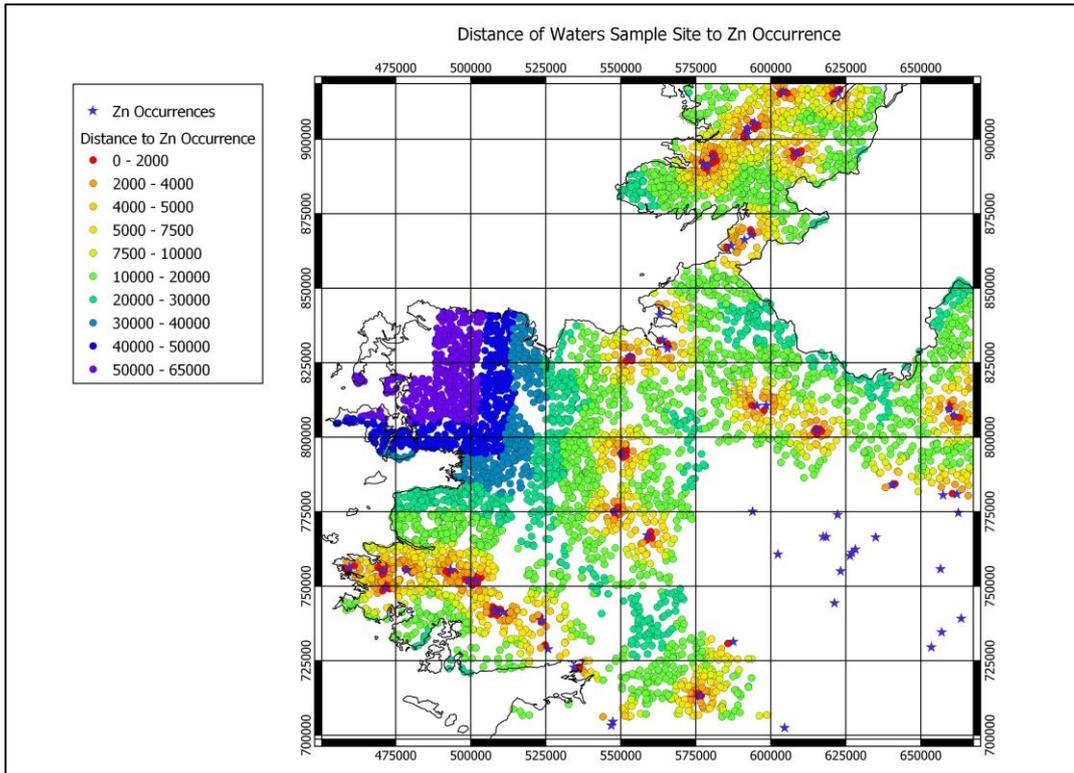


Figure D.36 Distance of water sample sites from Zn occurrences (MinLocs)

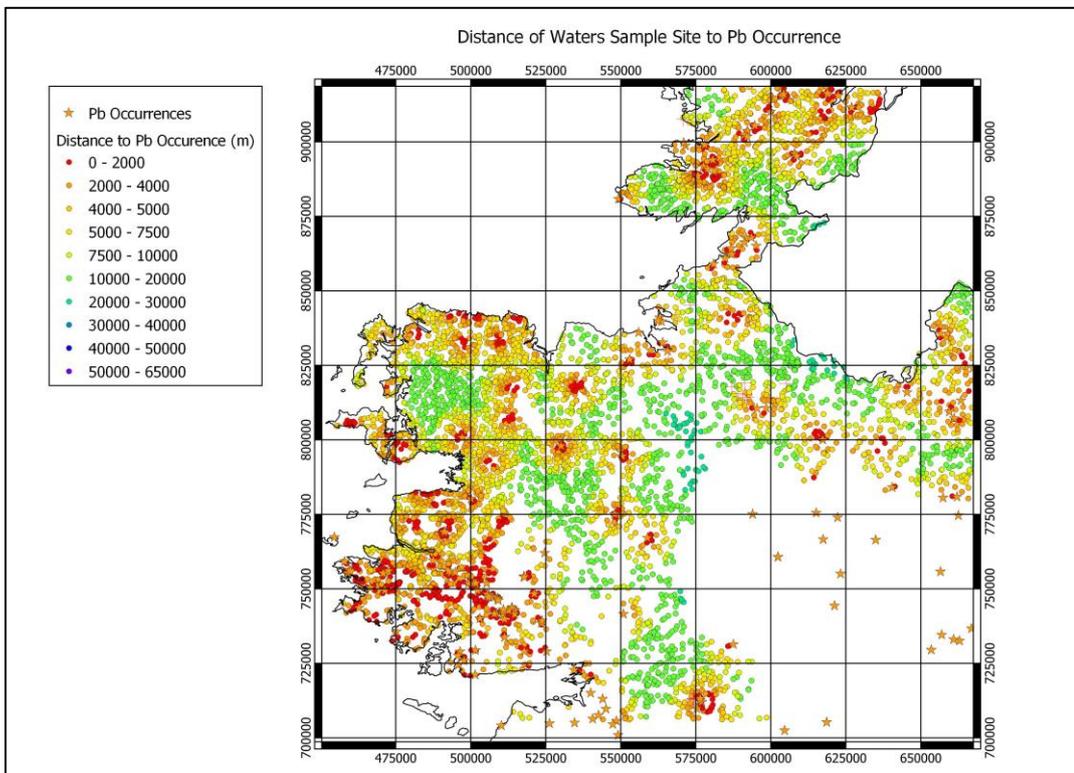


Figure D.37 Distance of water sample sites from Pb occurrences (MinLocs)



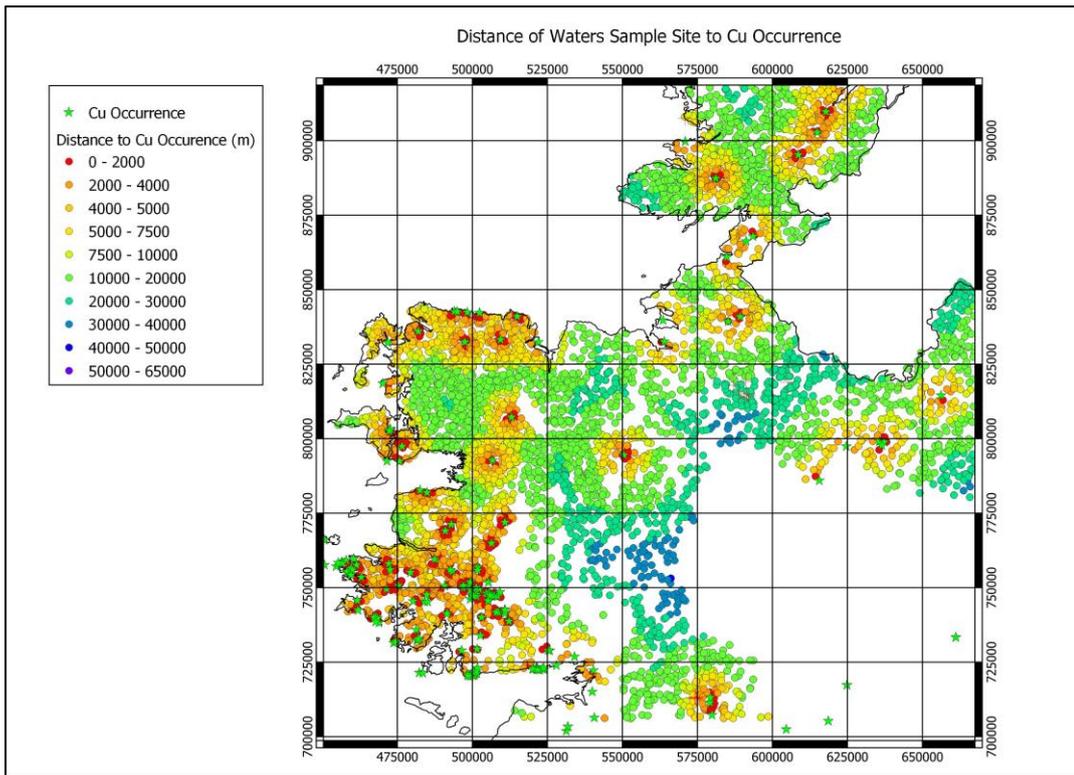


Figure D.38 Distance of water sample sites from Cu occurrences (MinLocs)



Appendix E: Priority Substances in Tellus stream waters

The accompanying figures and maps display the results of statistical analysis of Tellus stream water data for ten priority substances in order to (i) characterize the background variability and (ii) to identify the elements for which Tellus samples exceed EQS concentrations and do not fit within reasonable geochemical background.

The ten elements are F, As, Be, Cd, Cr, Cu, Ni, Pb, Sb and Zn. Details of preprocessing and data treatment are contained in Section 4.1.

Also included are Q-Q plots for four priority substances As, Cr, Cu and Zn, classified by SRF Domain (see section 3.3.1 for discussion). These Q-Q plots were used to estimate background concentrations of some elements in Tellus stream water classified by SRF Domain (Table 3.2).

E.1 EDA analysis

Exploratory Data Analysis (EDA) and Quantile-Quantile plots have been prepared for the 10 elements and are shown in the following figures, E1 to E.20. Table E.1 shows the data for each element.

Element	Quantile													
	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.98	0.99	
F	5.25	6.01	7.12	8.33	16.00	25.00	38.00	55.00	81.00	132.00	169.00	207.00	241.66	
As	0.09	0.12	0.17	0.22	0.29	0.36	0.44	0.53	0.67	0.97	1.38	2.16	3.33	
Be	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.04	0.06	
Cd	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.04	0.08	0.11	
Cr	0.06	0.08	0.10	0.13	0.15	0.18	0.22	0.27	0.33	0.44	0.58	0.81	1.00	
Cu	0.09	0.14	0.23	0.32	0.44	0.58	0.75	0.98	1.35	1.95	2.77	4.38	6.01	
Ni	0.09	0.13	0.21	0.31	0.50	0.76	1.12	1.51	2.06	3.05	4.07	5.79	7.09	
Pb	0.01	0.01	0.01	0.02	0.04	0.06	0.08	0.11	0.16	0.27	0.41	0.65	0.90	
Sb	0.01	0.01	0.02	0.02	0.02	0.03	0.04	0.05	0.07	0.10	0.14	0.20	0.26	
Zn	0.30	0.44	0.65	0.85	1.06	1.31	1.61	2.02	2.66	4.07	6.56	21.74	34.23	

Table E.1 Quantile values for each of the ten priority elements

The EDA plots show the distribution of the data for each element. Histograms and density traces can reveal the presence of multiple populations, e.g. the bimodal distribution of F. On the Quantile-Quantile plots discontinuities may be used to estimate threshold values, e.g. background concentrations of elements in a given population. Background concentrations for priority elements are discussed in Section 3.

Table E.2 summarizes possible threshold levels for each element, estimated using break points in the Quantile-Quantile plots. The 98th percentile values are included for comparison.

	F	As	Be	Cd	Cr	Cu	Ni	Pb	Sb	Zn
Q-Q	316	10	0.12	0.32	1.78	10	10	1.5	0.32	6.5 / 3.4
98 th %ile	207	2.16	0.04	0.08	0.81	4.38	5.79	0.65	0.20	21.7



Table E.2 Threshold values for each of the ten priority elements based on (i) breaks in Q-Q plots and (ii) 98th percentile value

Principal component biplots of the combined group of the priority elements are shown in Figures E.21 to E.30.

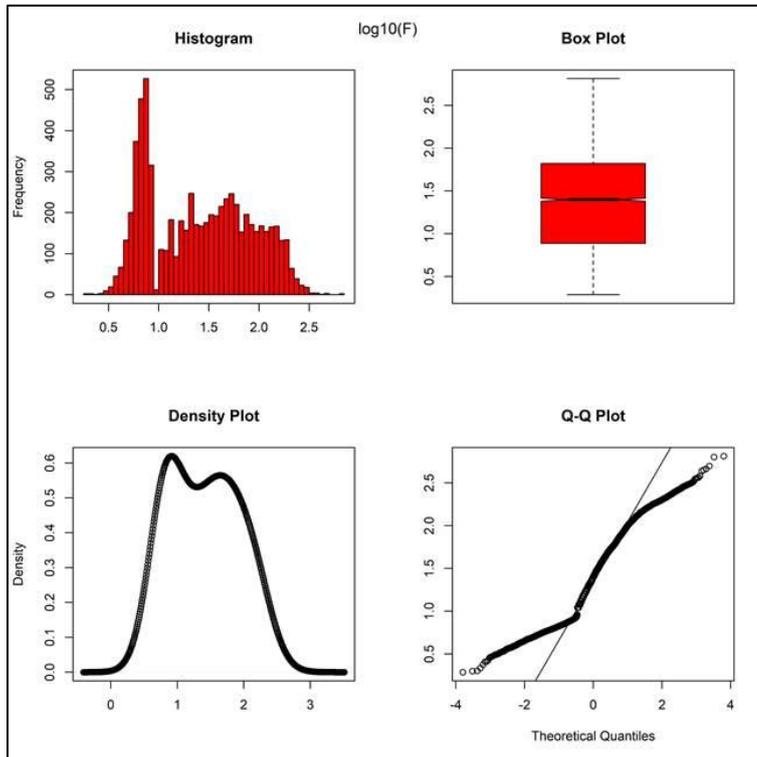


Figure E.1 EDA plots for F



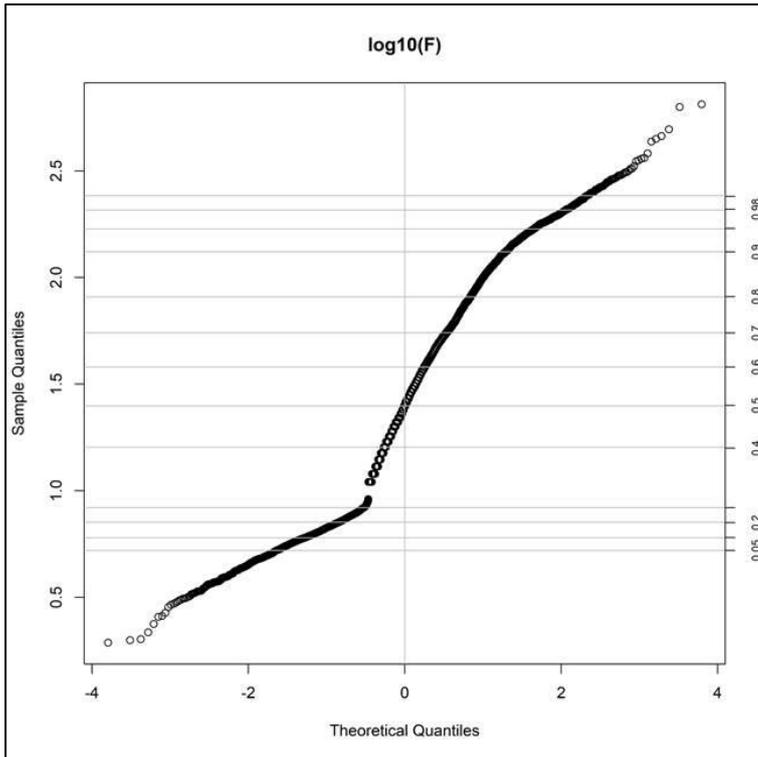


Figure E.2 Quantile-Quantile plot, F.

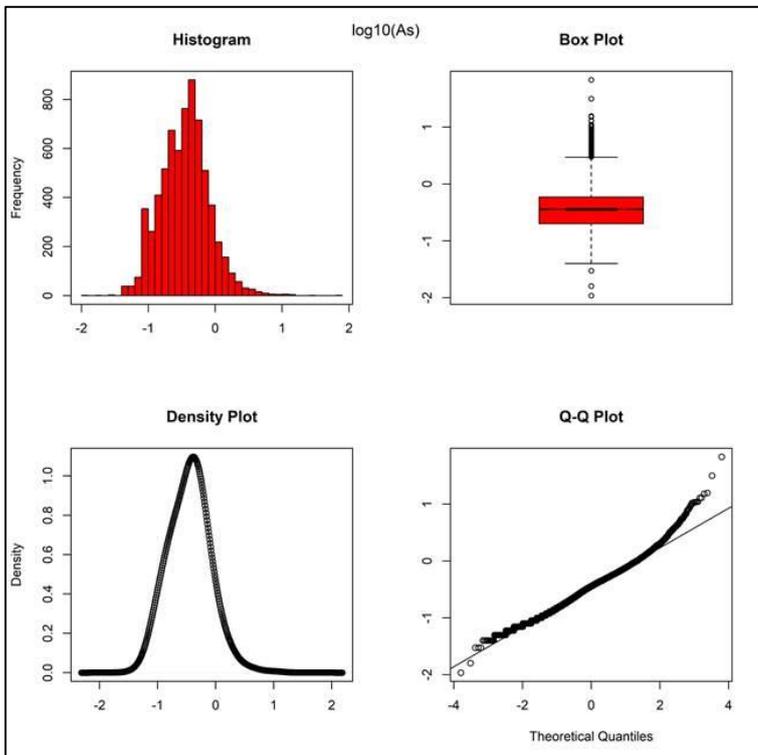


Figure E.3 EDA plots for As



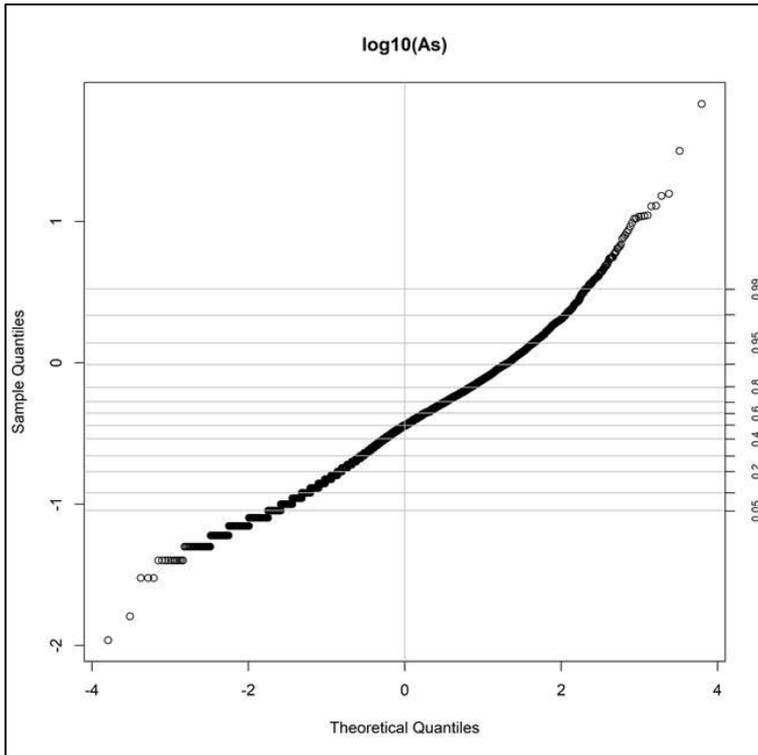


Figure E.4 Quantile-Quantile plot, As.

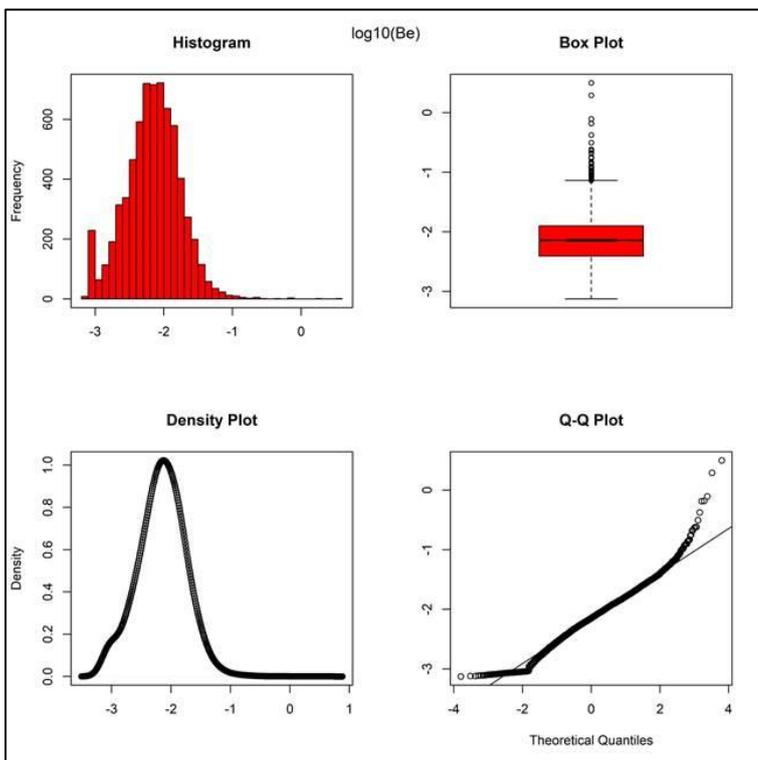


Figure E.5 EDA plots for Be



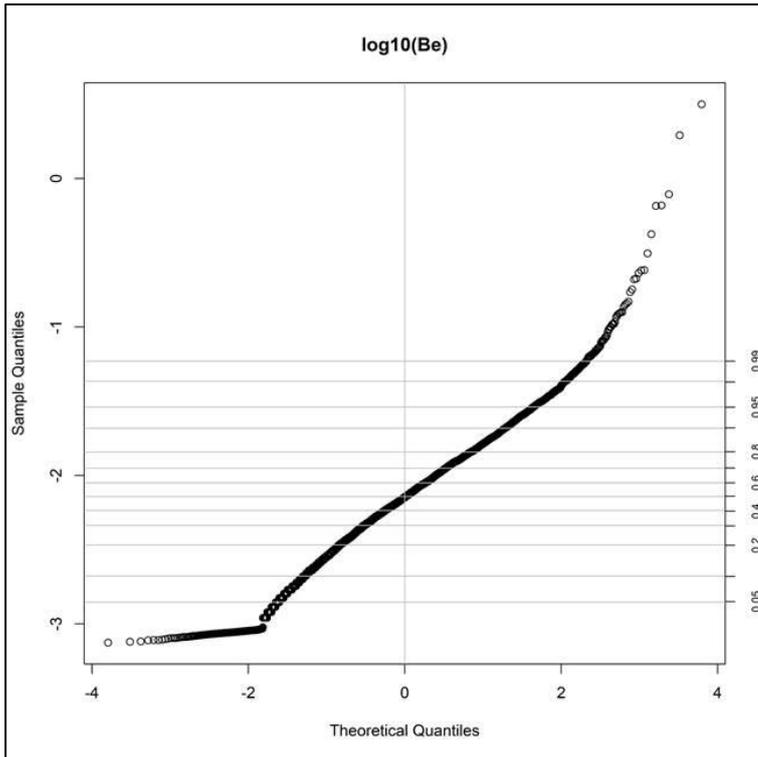


Figure E.6 Quantile-Quantile plot, Be.

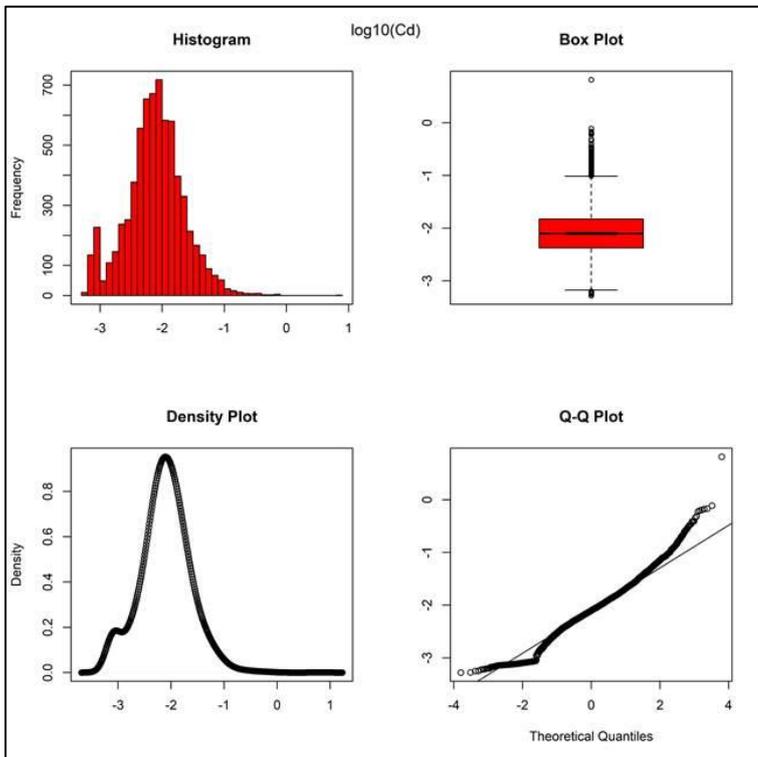


Figure E.7 EDA plots for Cd



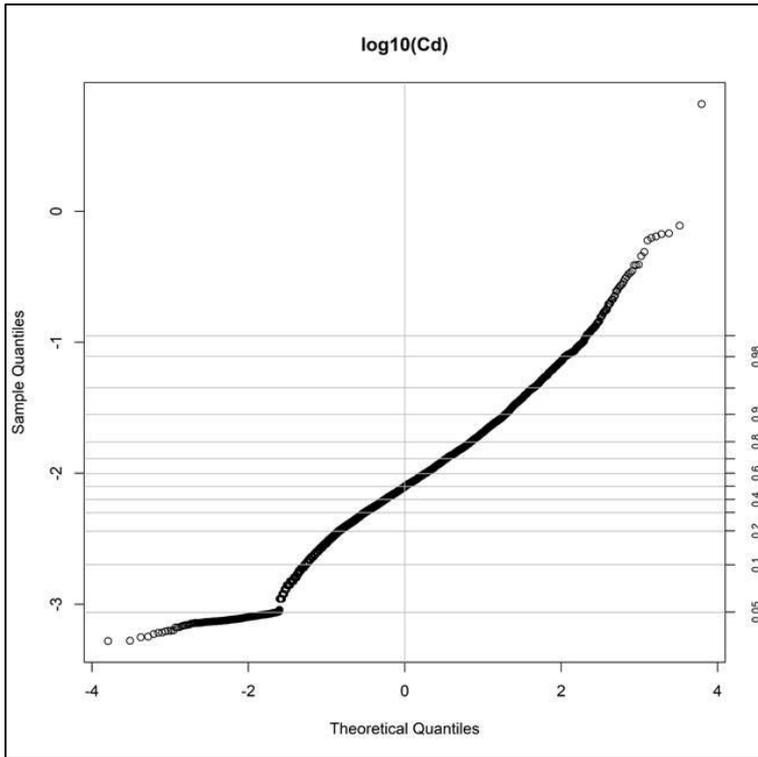


Figure E.8 Quantile-Quantile plot, Cd.

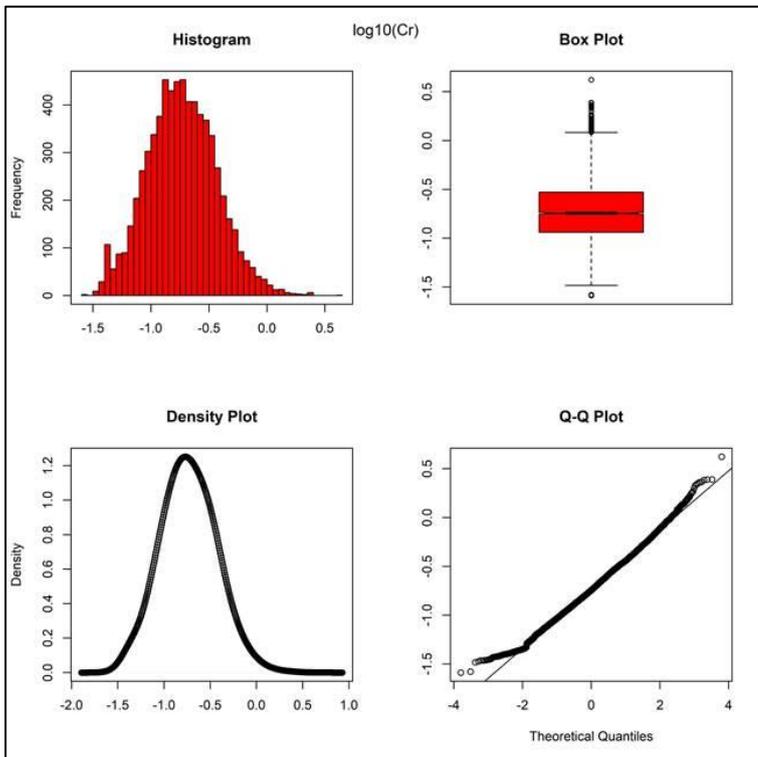


Figure E.9 EDA plots for Cr



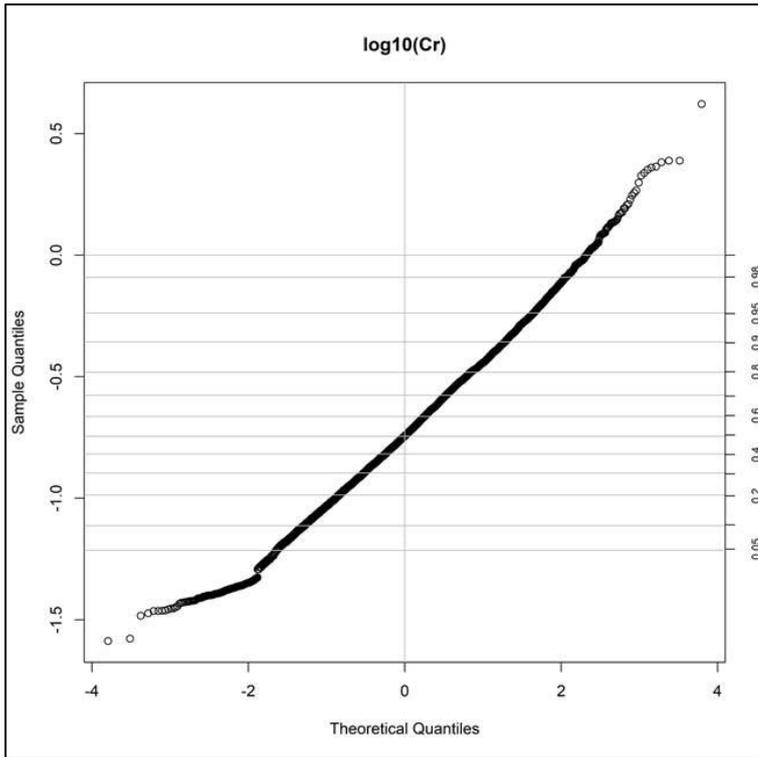


Figure E.10 Quantile-Quantile plot, Cr.

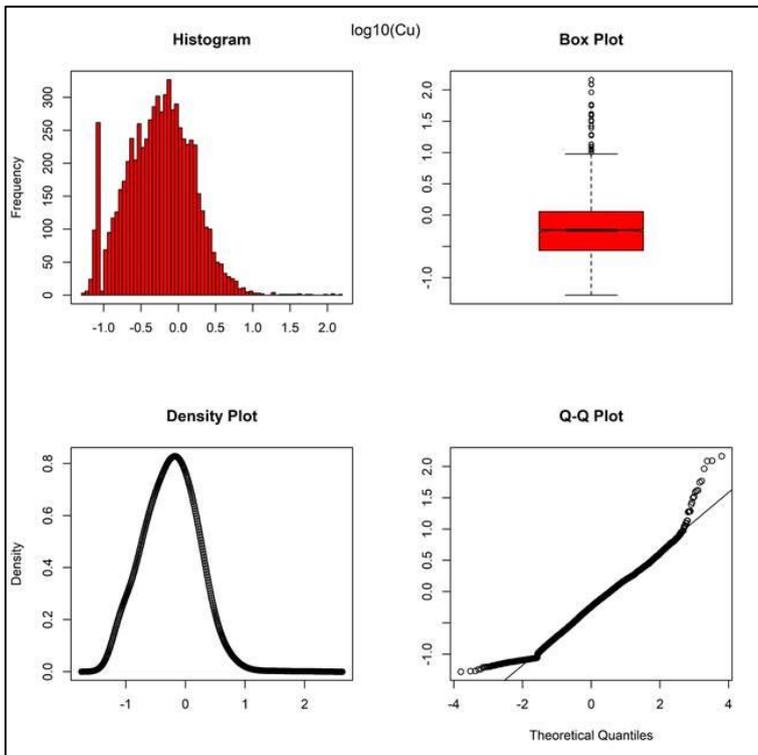


Figure E.11 EDA plots for Cu



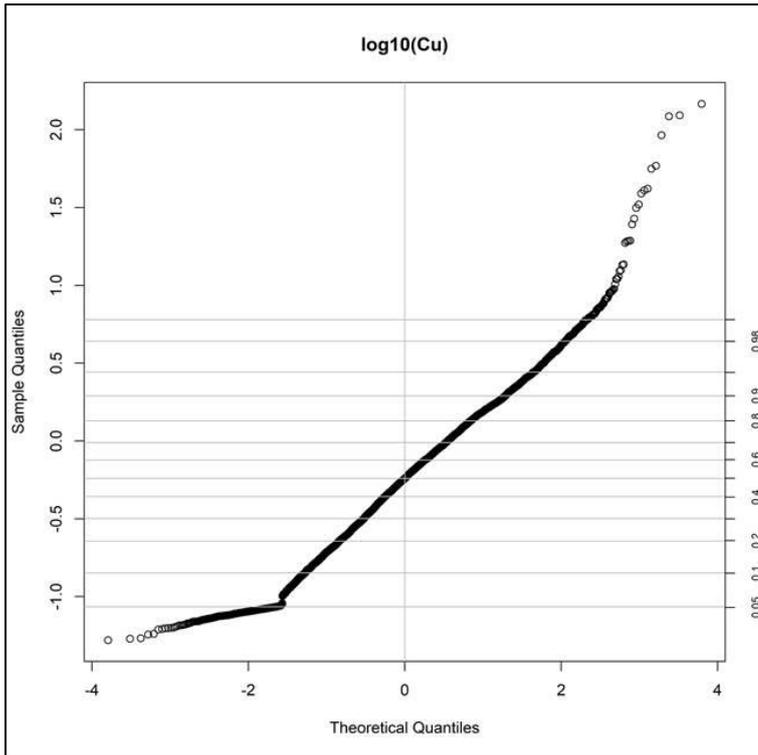


Figure E.12 Quantile-Quantile plot, Cu.

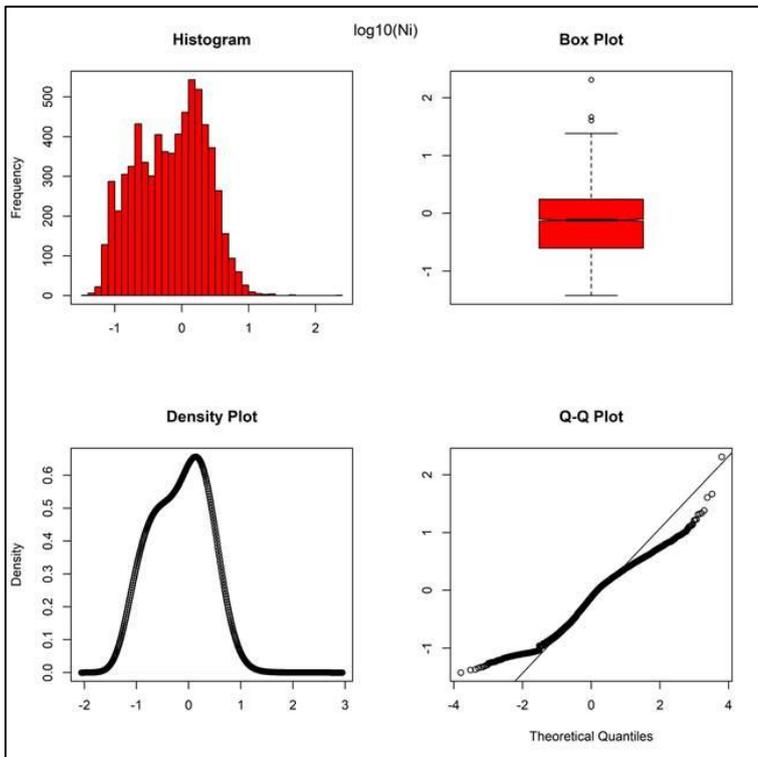


Figure E.13 EDA plots for Ni



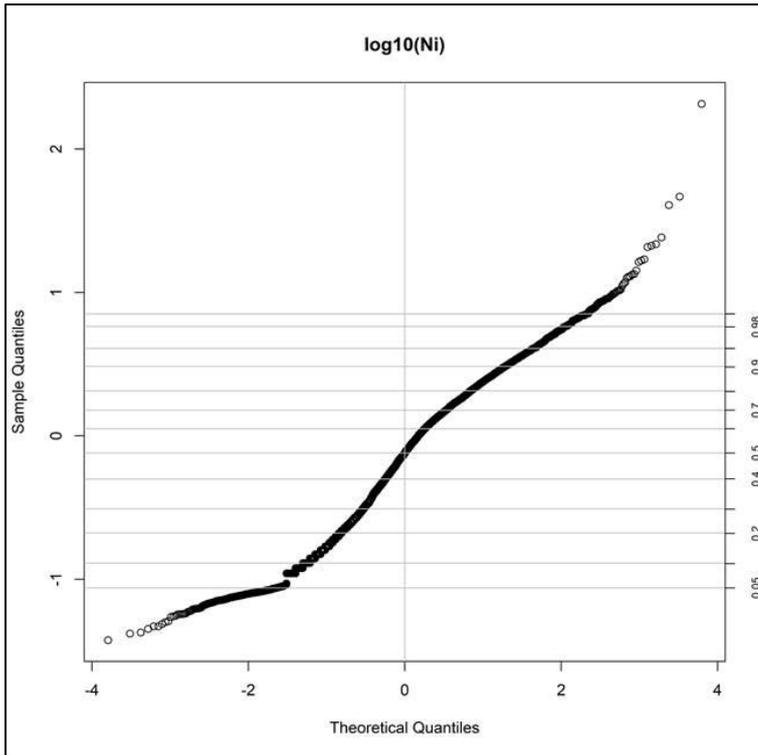


Figure E.14 Quantile-Quantile plot, Ni.

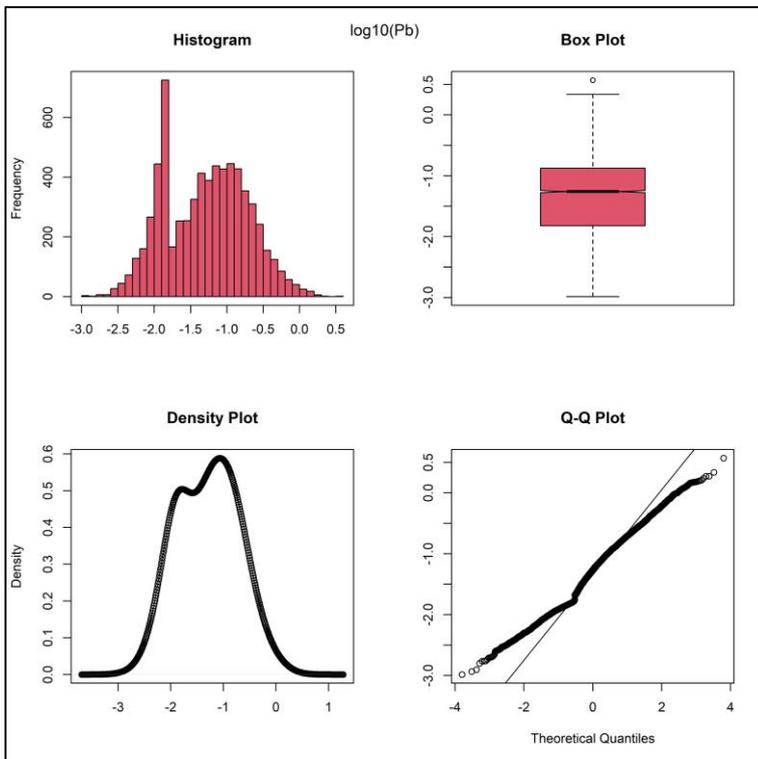


Figure E.15 EDA plots for Pb



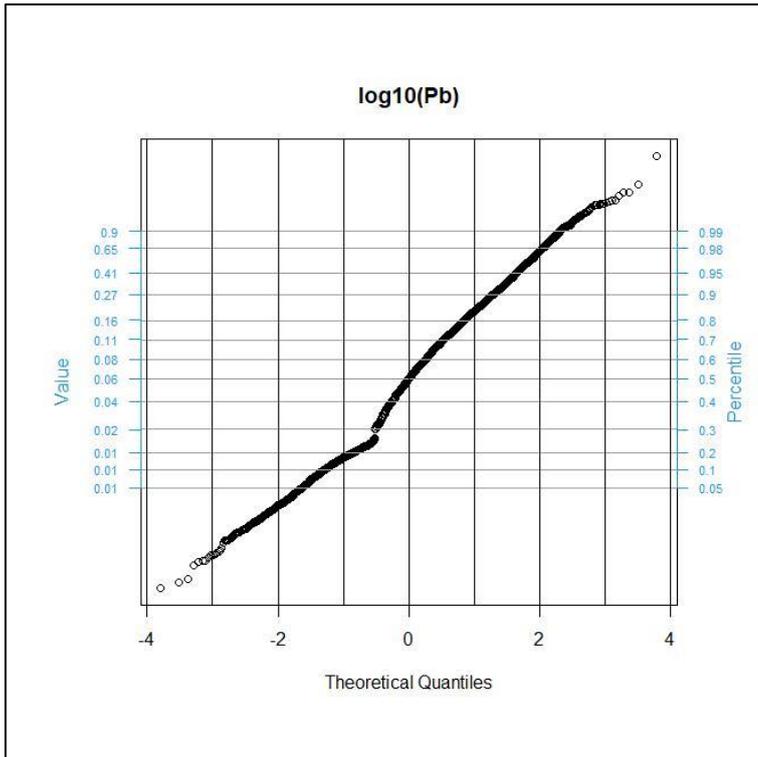


Figure E.16 Quantile-Quantile plot, Pb.

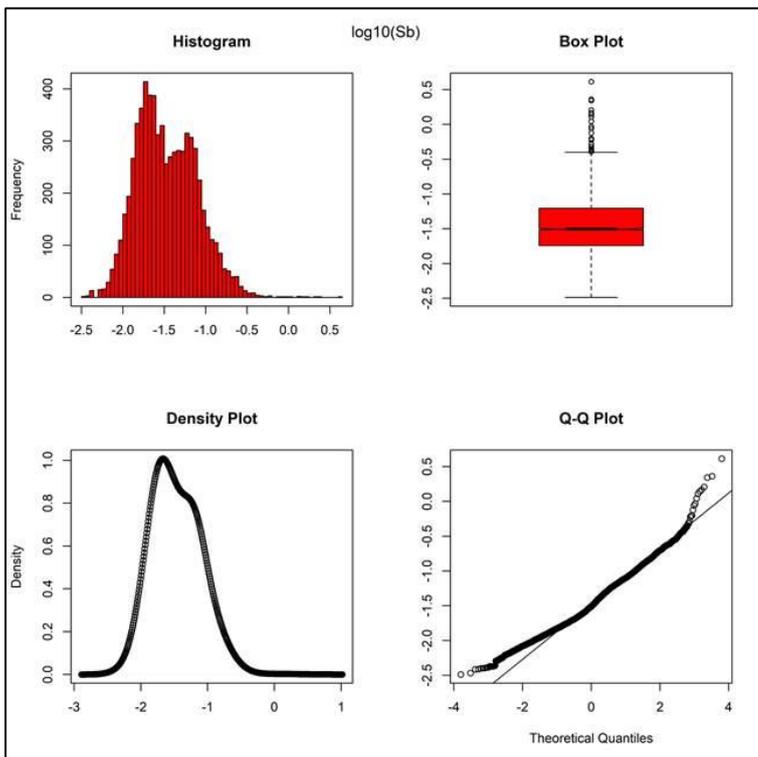


Figure E.17 EDA plots for Sb



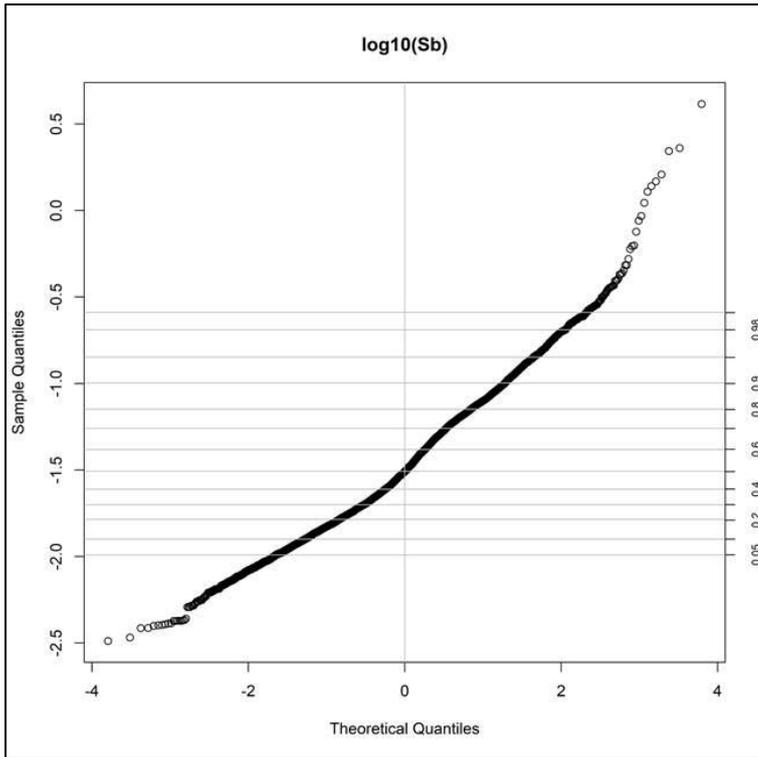


Figure E.18 Quantile-Quantile plot, Sb.

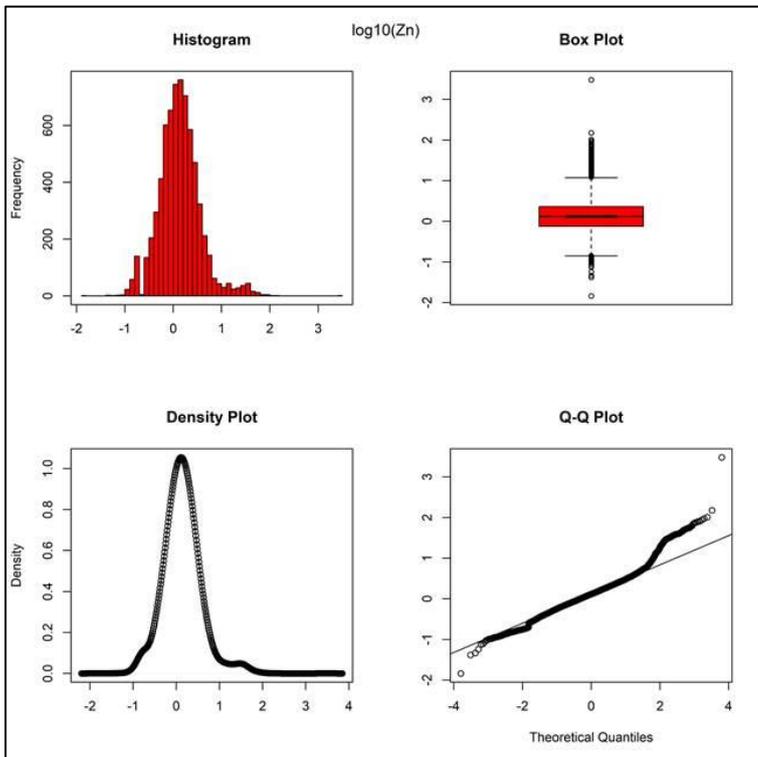


Figure E.19 EDA plots for Zn



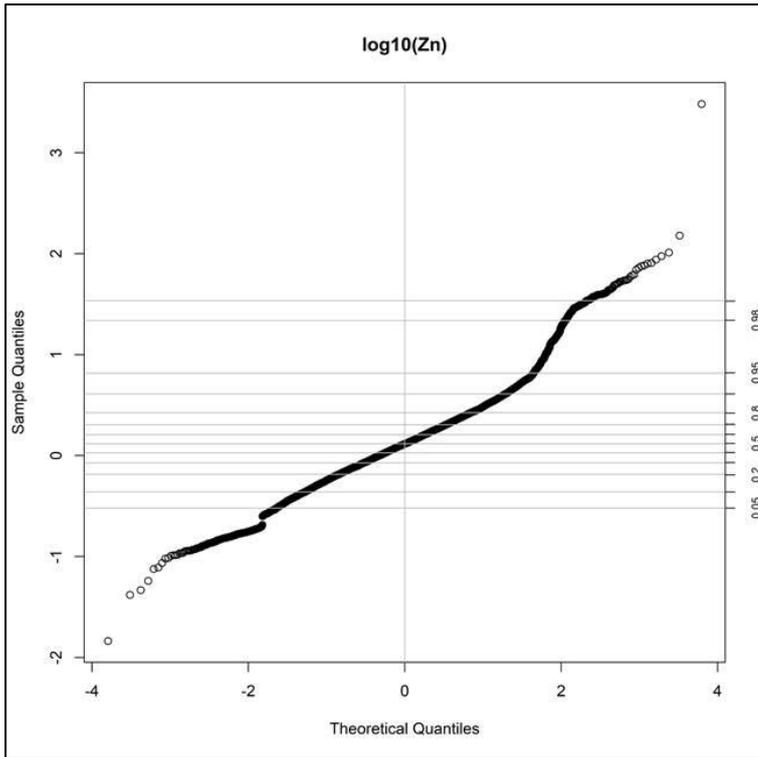


Figure E.20 Quantile-Quantile plot, Zn.

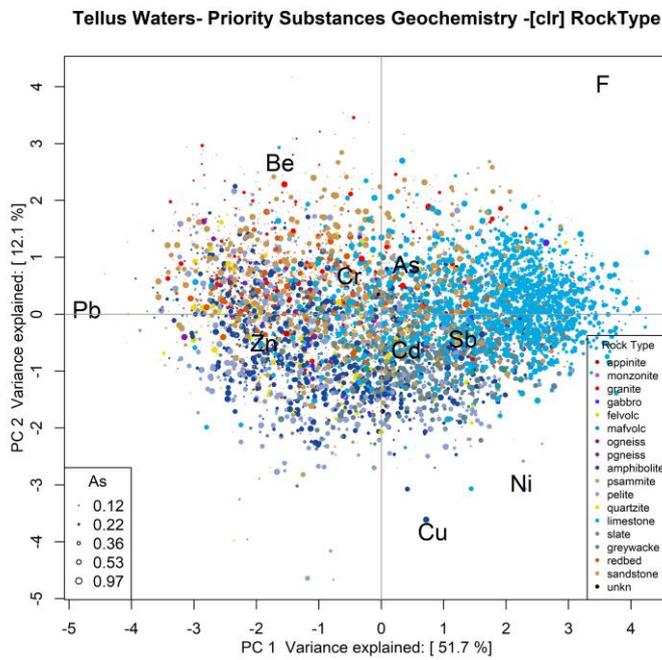


Figure E.21 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for As.



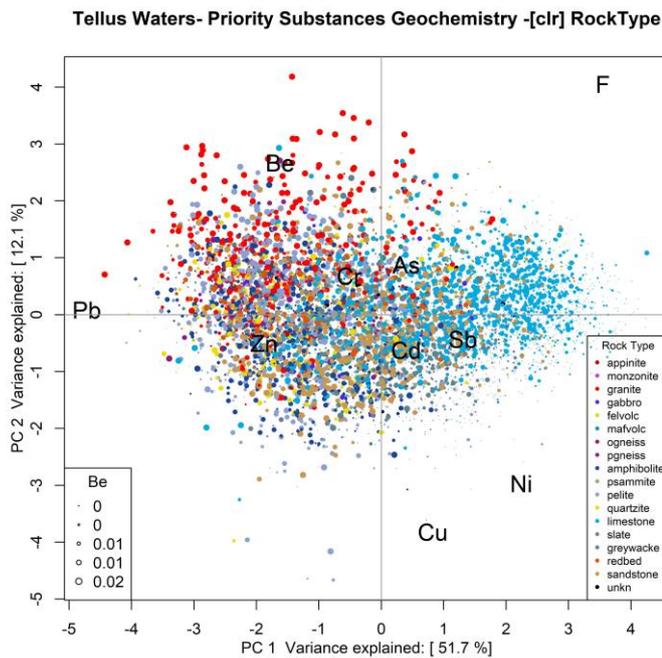


Figure E.22 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Be.

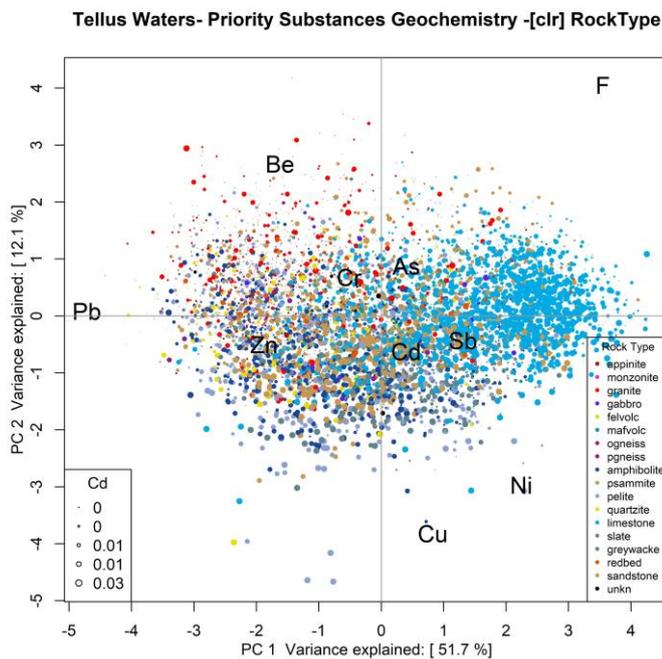


Figure E.23 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Cd.



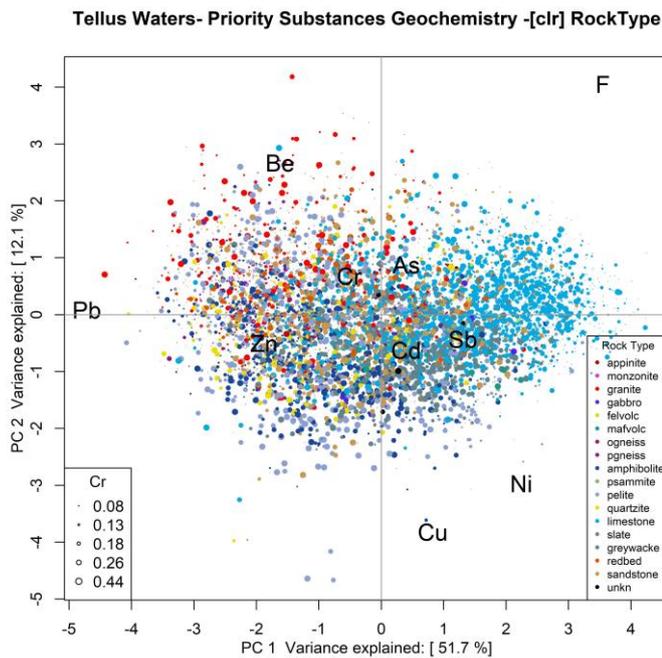


Figure E.24 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Cr.

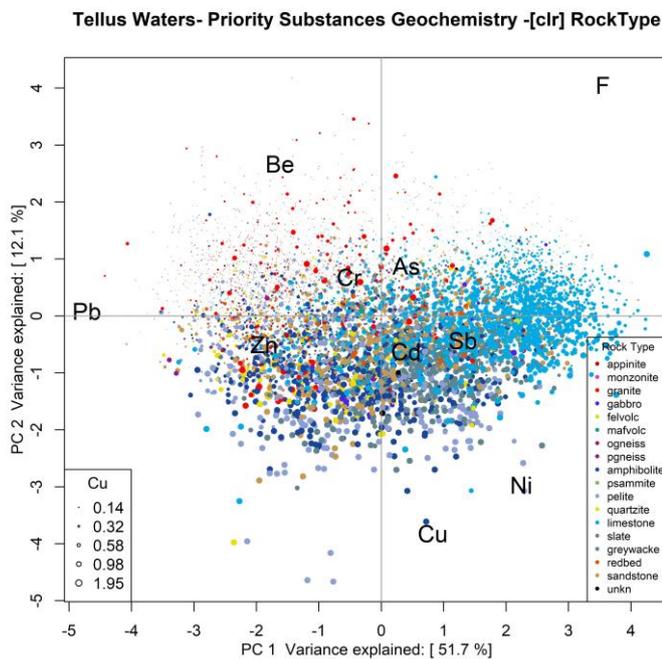


Figure E.25 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Cu.



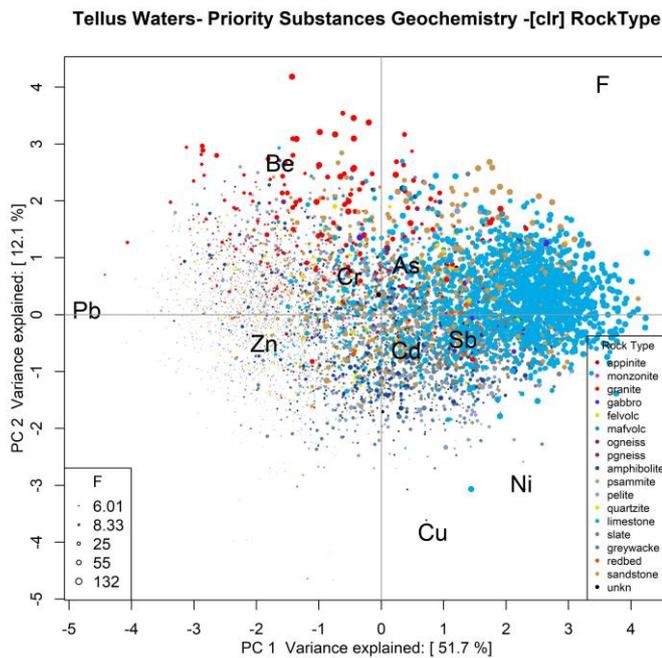


Figure E.26 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for F.

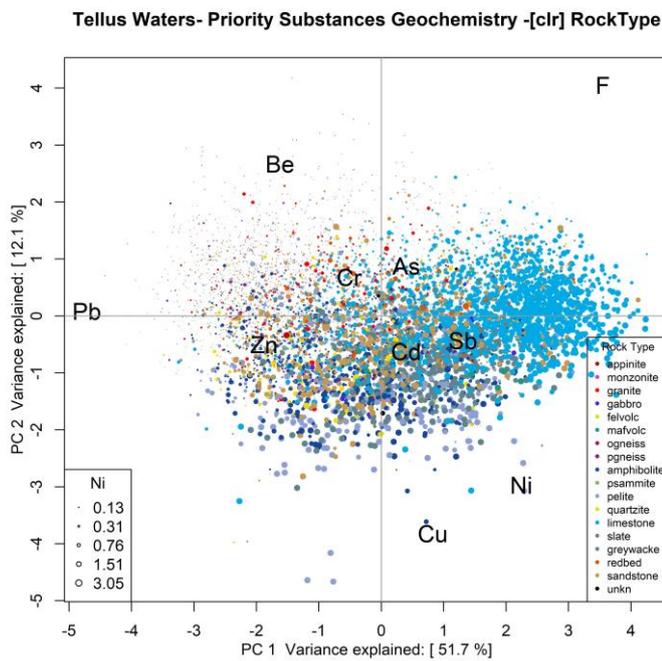


Figure E.27 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Ni.



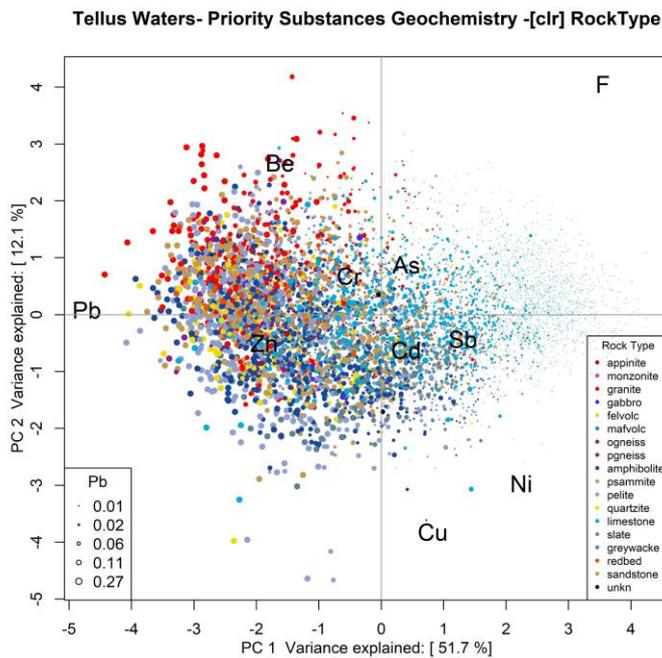


Figure E.28 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Pb.

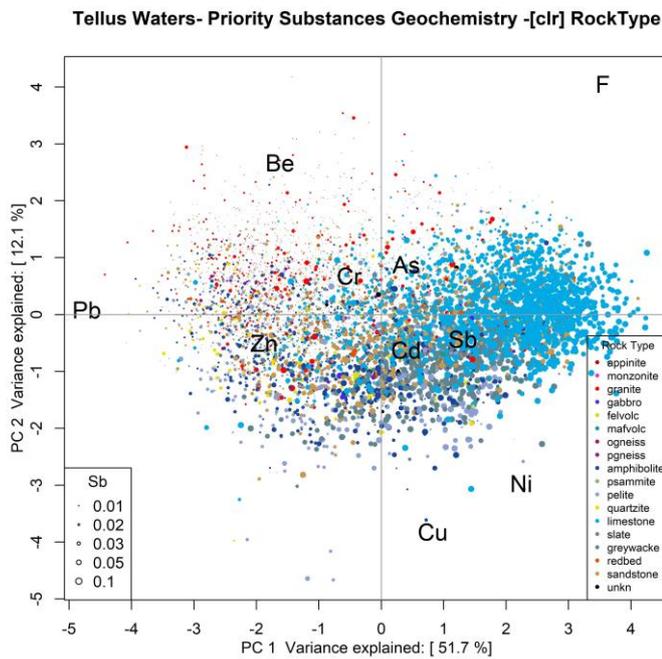


Figure E.29 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Sb.



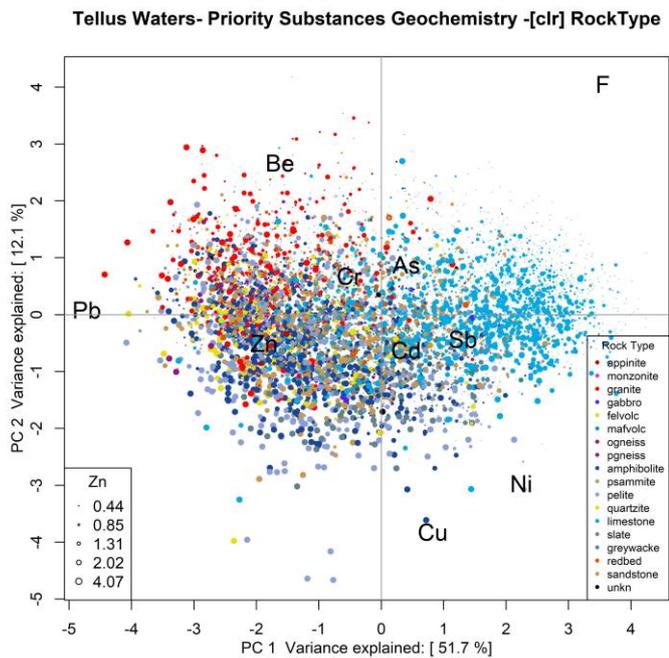


Figure E.30 – Biplot of PC1vPC2 coded by lithology with symbol sizes scaled for Zn.



E.2 Q-Q plots for As, Cr, Cu and Zn in stream water, classified by SRF Domain

Domain 1

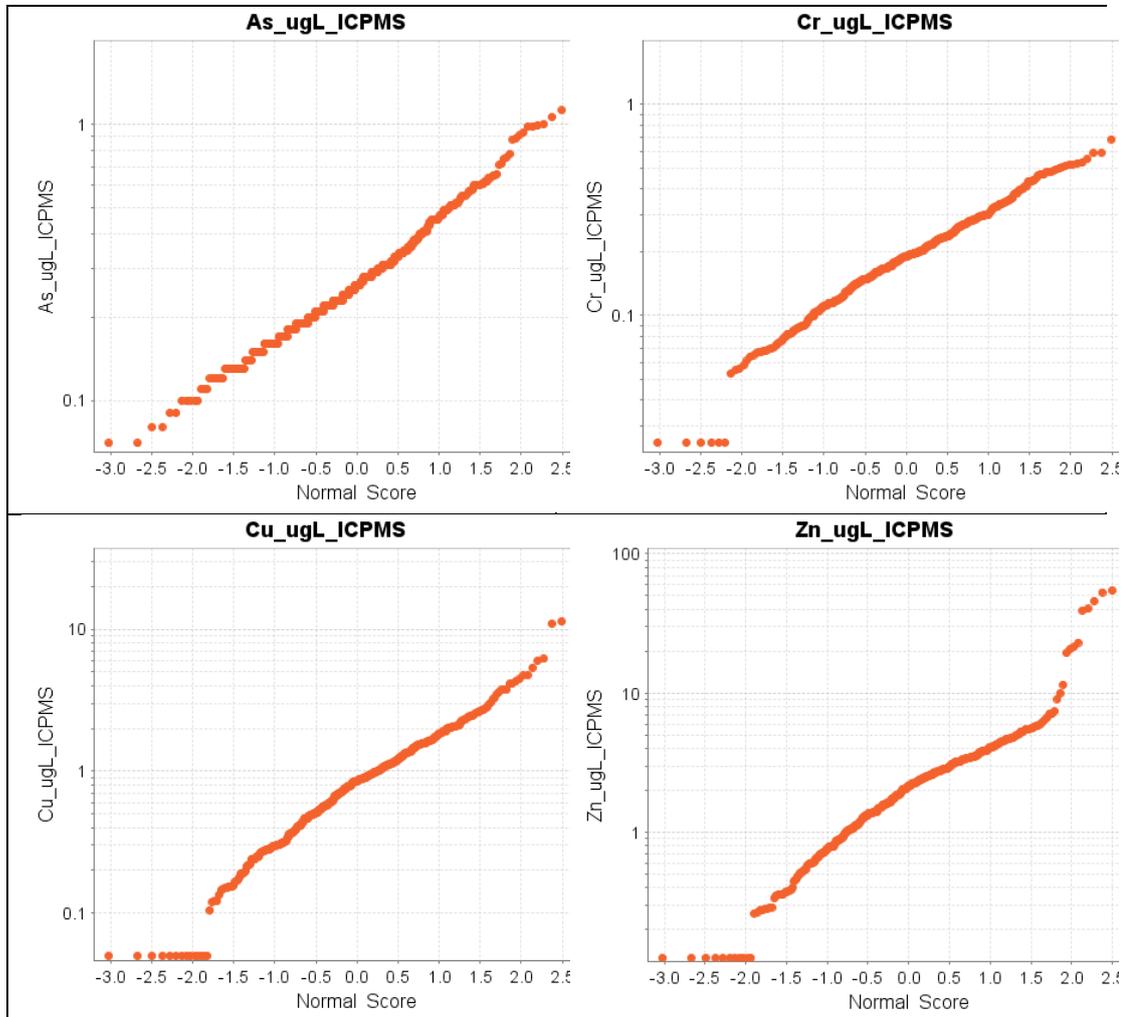


Figure E.31 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 1.



Domain 2

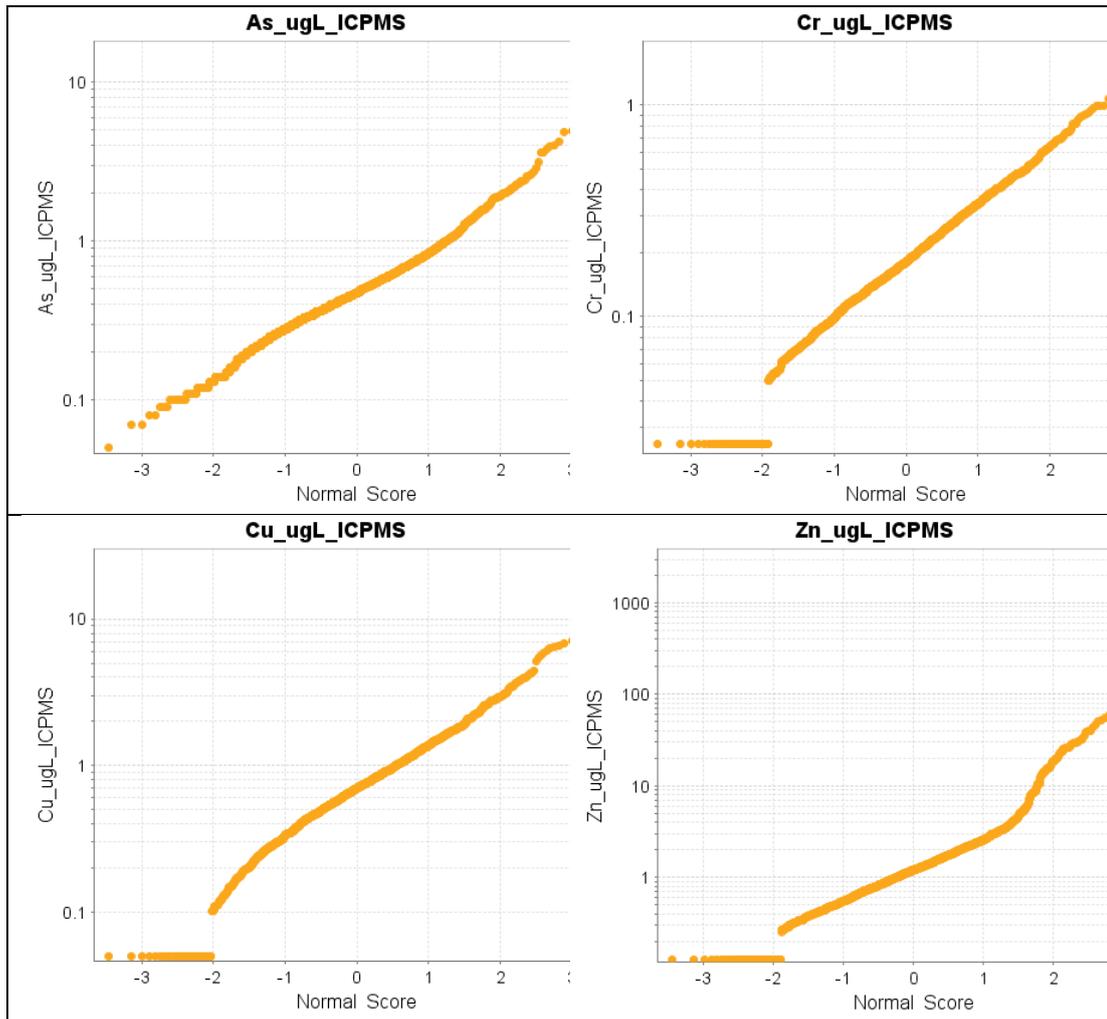


Figure E.32 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 2.



Domain 3

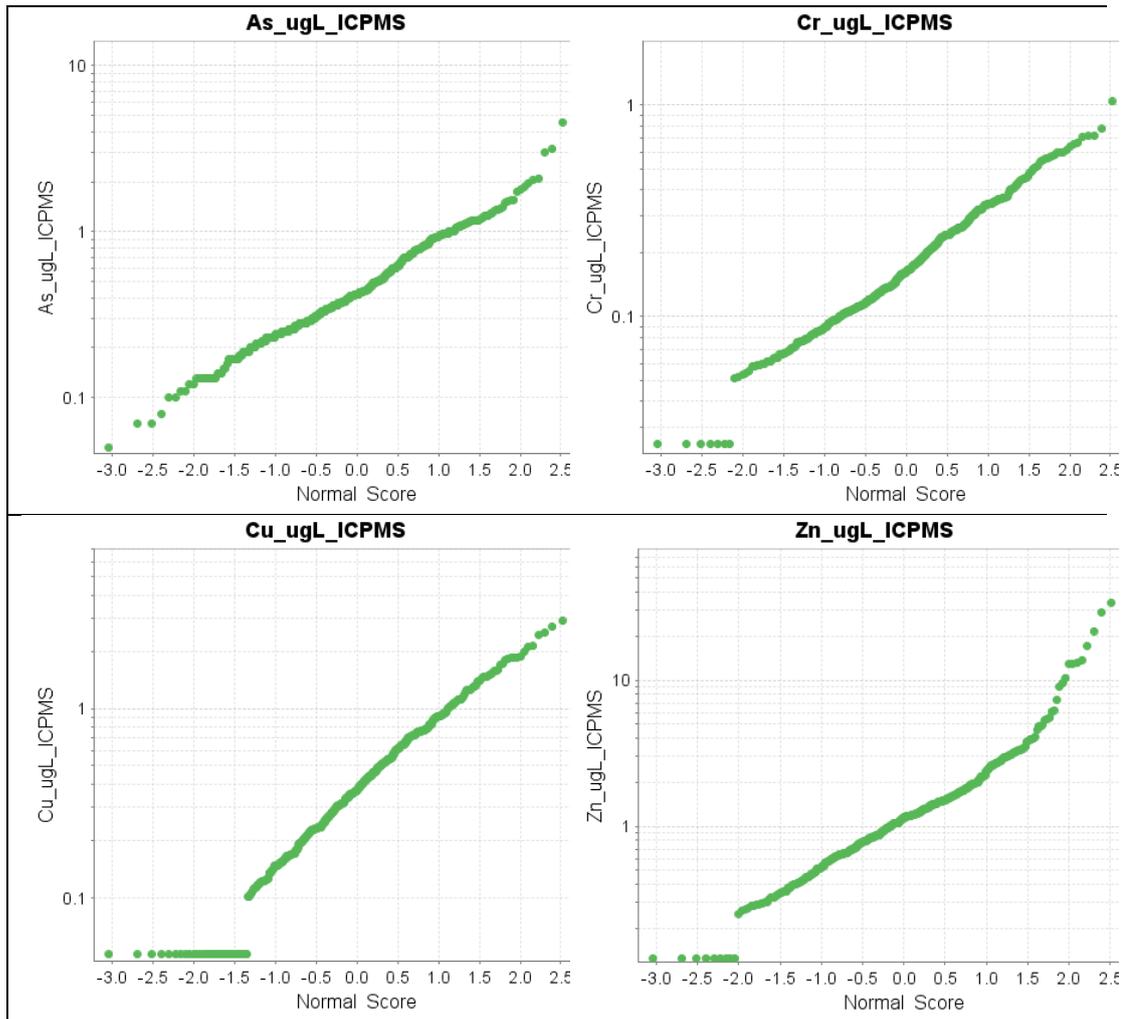


Figure E.33 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 3.



Domain 4

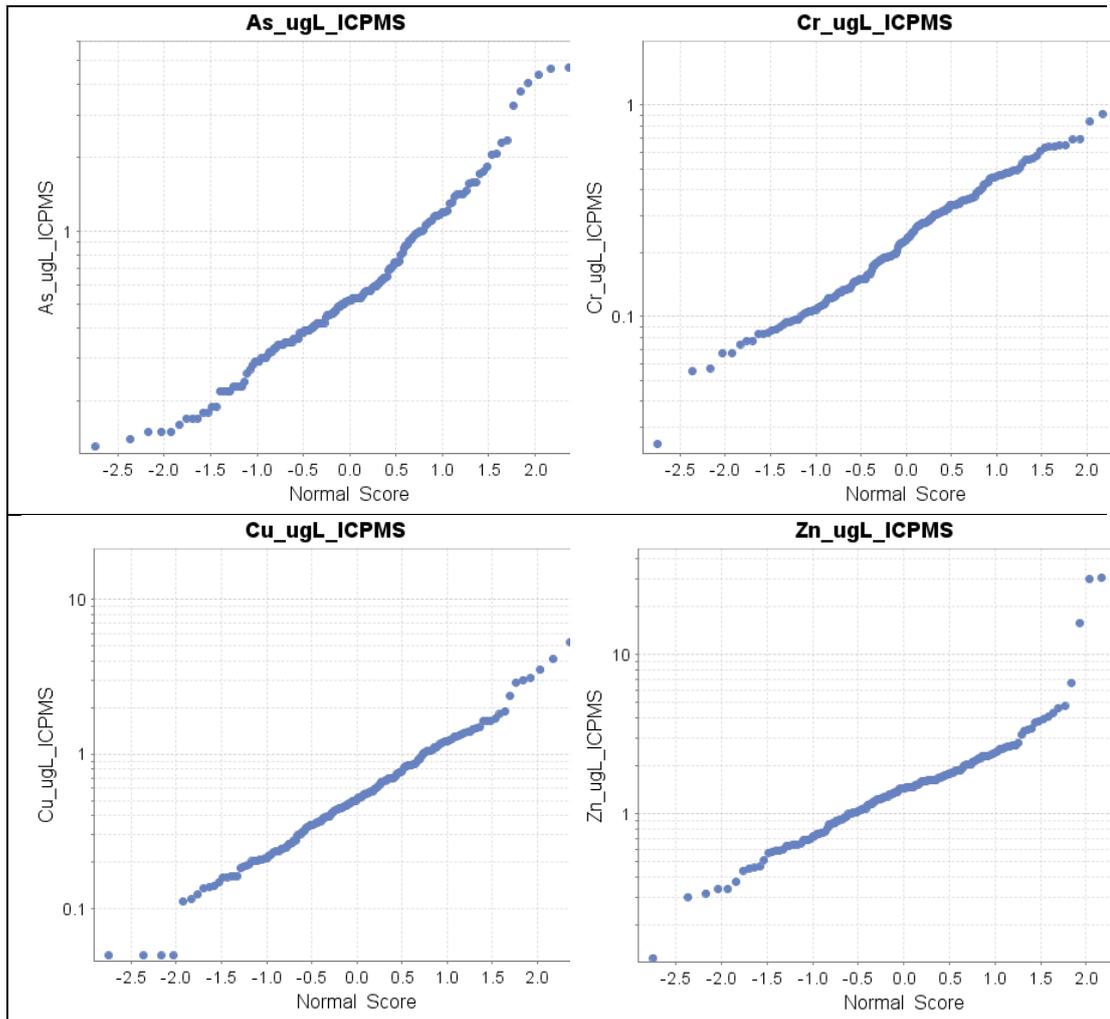


Figure E.34 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 4.



Domain 5

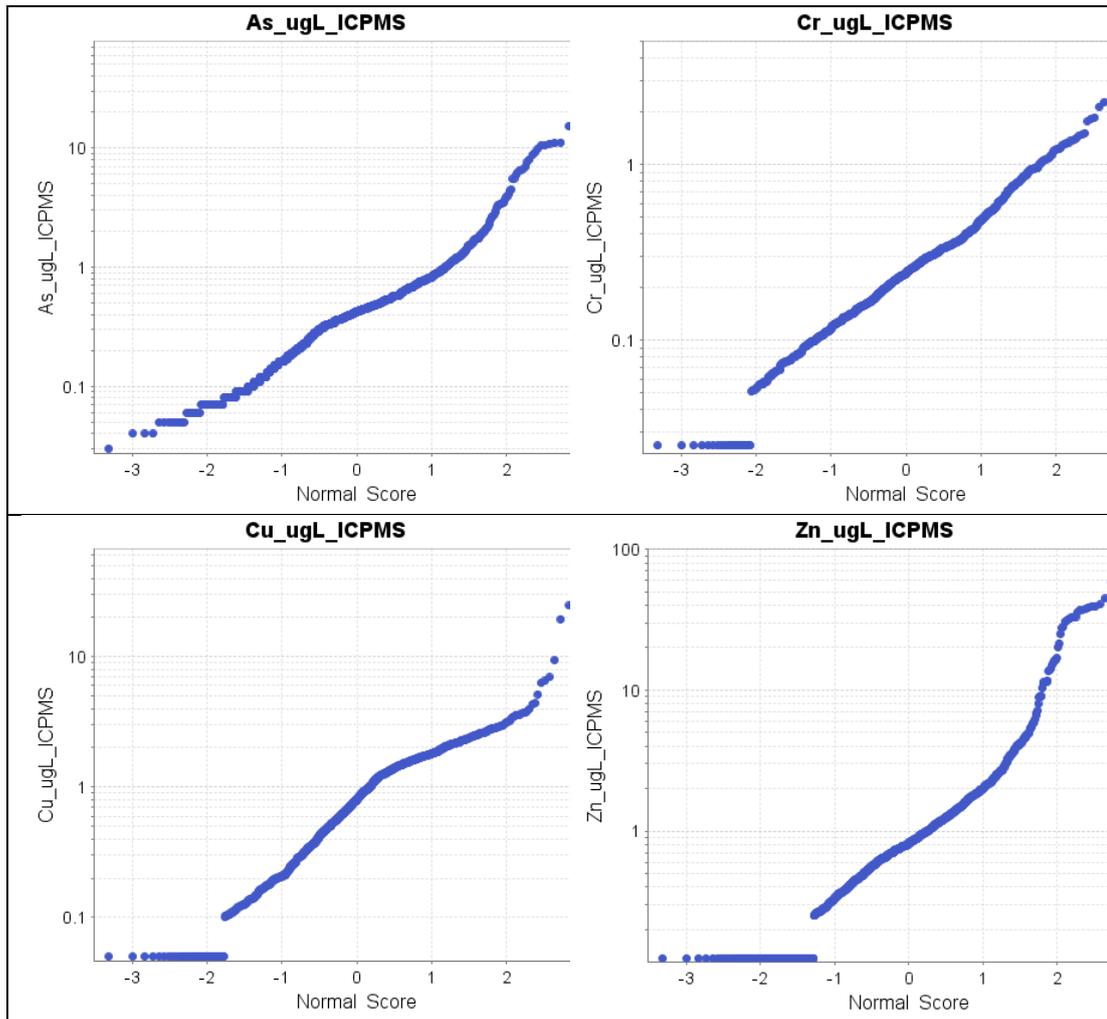


Figure E.35 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 5.



Domain 6

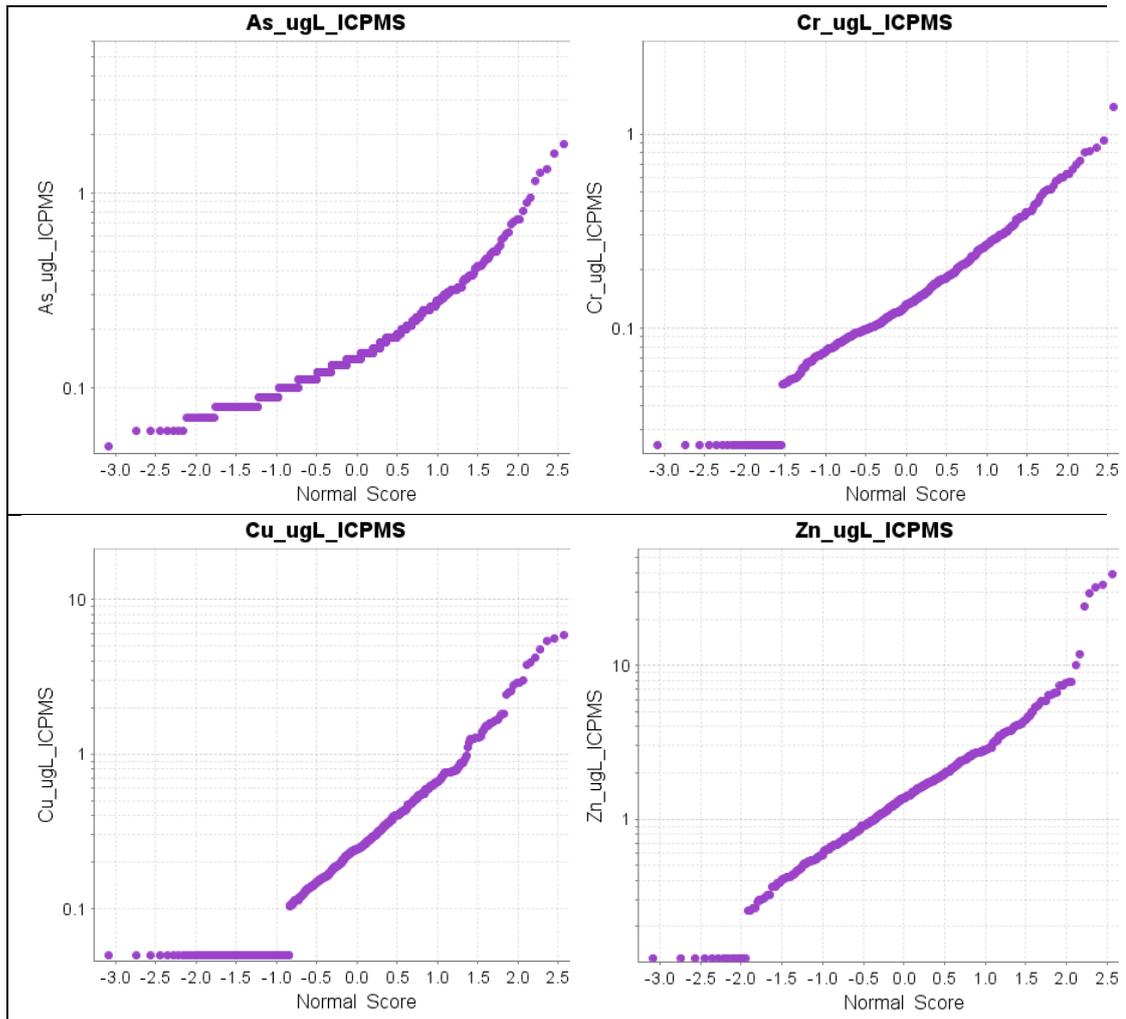


Figure E.36 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 6.



Domain 7

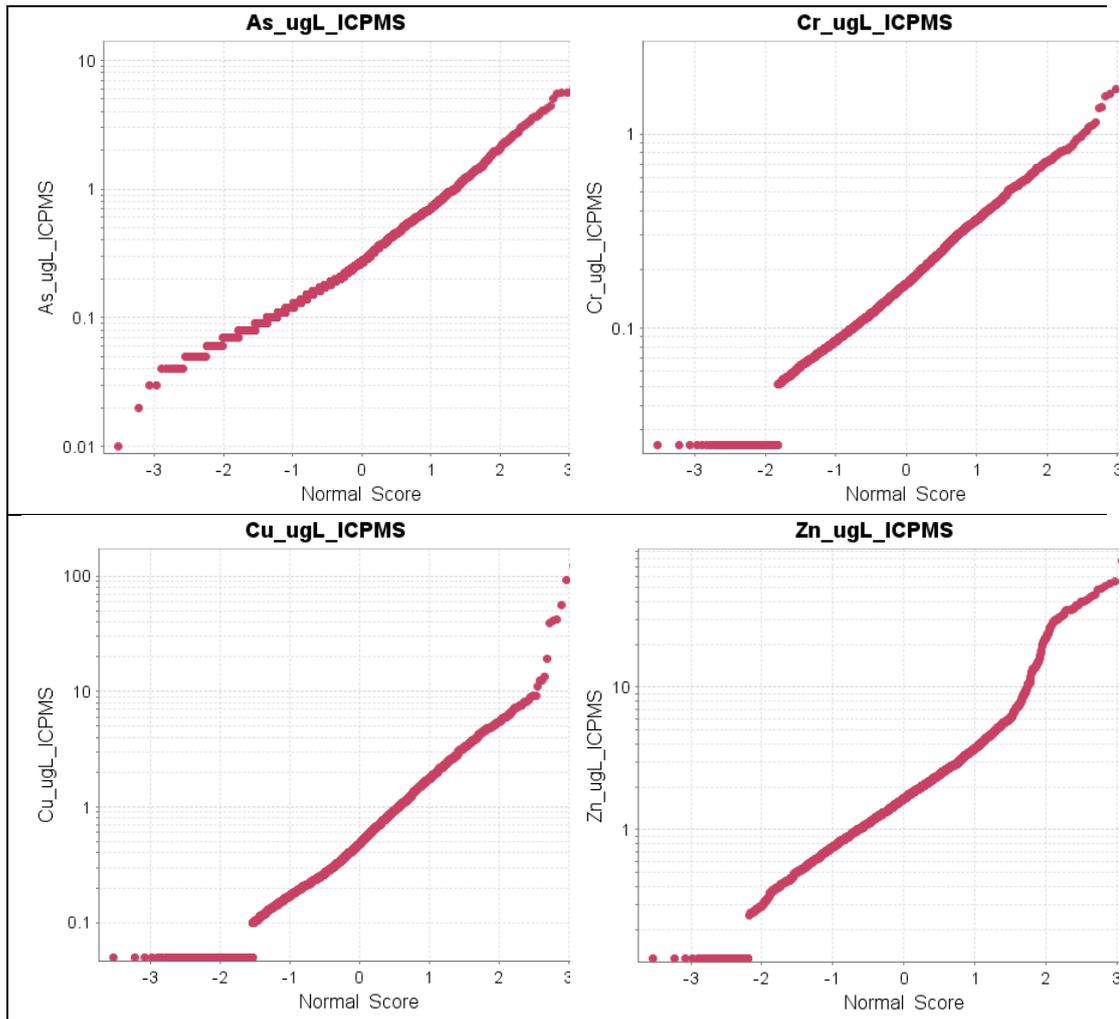


Figure E.37 – Q-Q plots for As, Cr, Cu and Zn for SRF Domain 7.

