

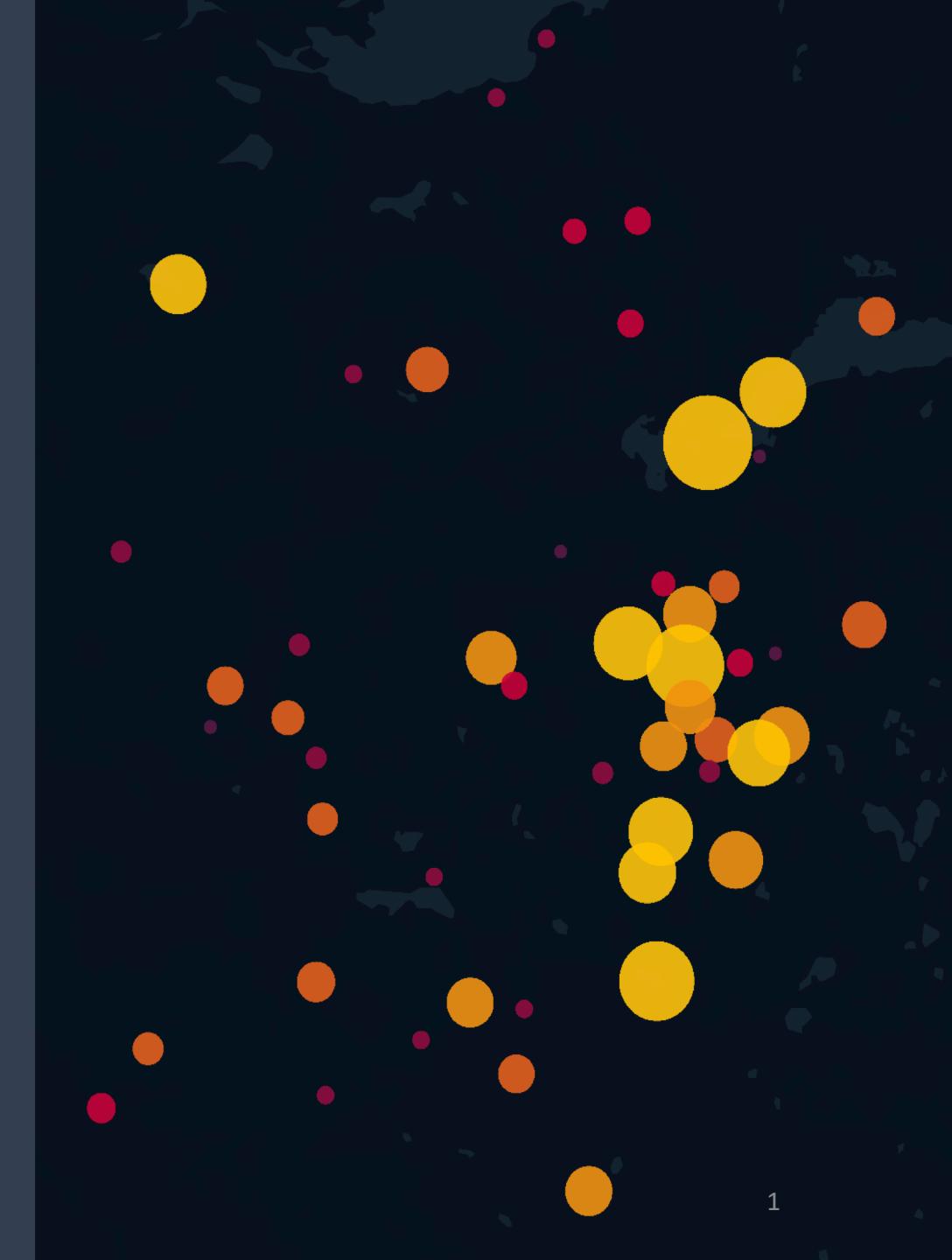
Can soil geochemistry predict the mercury concentration in fish?

Kristin M. Eccles

kristin.eccles@utoronto.ca

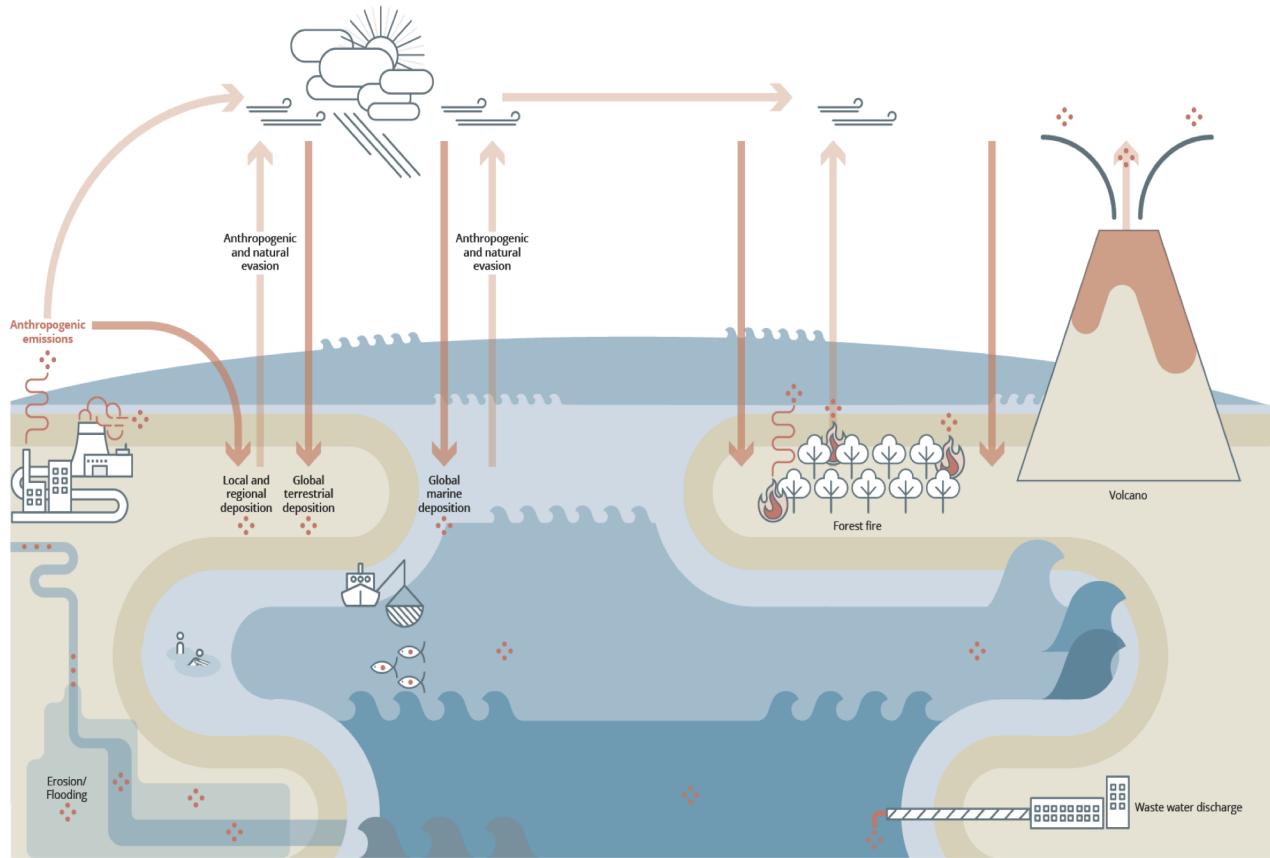


@kristineccles



Background/ Context

- Mercury is considered by WHO as one of the top ten chemicals or groups of chemicals of major public health concern
- Toxic effects:
 - Nervous system
 - Endocrine system
 - Immune system
 - Cardiovascular system
 - Gastrointestinal tract
- Top predators have the highest risk of exposure do to bioaccumulation and biomagnification
- Humans are primarily exposure though a fish diet



Research Question:

Since mercury biomagnification and accumulation is controlled by environmental factors can we use soil geochemistry to predict fish tissue Hg in the United States?



PERIODIC TABLE OF ELEMENTS

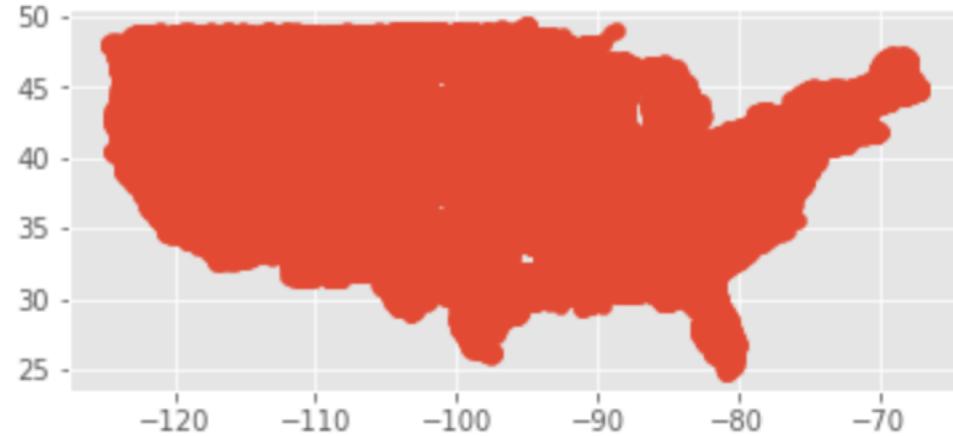
PubChem

1	H	Hydrogen	2	He	Helium
3	Li	Lithium	4	Be	Beryllium
11	Na	Sodium	12	Mg	Magnesium
19	K	Potassium	20	Ca	Calcium
37	Rb	Rubidium	38	Sr	Strontrium
55	Cs	Cesium	56	Ba	Barium
87	Fr	Francium	88	Ra	Radium
21	Sc	Scandium	22	Ti	Titanium
39	Y	Yttrium	40	Zr	Zirconium
72	Hf	Hafnium	73	Ta	Tantalum
104	Rf	Rutherfordium	105	Db	Dubium
57	La	Lanthanum	58	Ce	Cerium
89	Ac	Actinium	90	Th	Thorium
91	Pa	Protactinium	92	U	Uranium
93	Np	Neptunium	94	Pu	Plutonium
95	Am	Americium	96	Cm	Curium
97	Bk	Berkelium	98	Cf	Californium
99	Es	Einsteinium	100	Fm	Fermium
101	Md	Mendelevium	102	No	Nobelium
103	Lr	Lawrencium			
5	B	Boron	6	C	Carbon
13	Al	Aluminum	14	Si	Silicon
31	Ga	Gallium	32	Ge	Germanium
49	In	Indium	51	Sb	Sb
79	Hg	Mercury	81	Tl	Thallium
80	Pt	Platinum	82	Pb	Lead
81	Os	Osmium	83	Bi	Bismuth
82	Pt	Pt	84	Po	Po
83	Ir	Iridium	85	At	At
84	Pb	Pb	86	Rn	Rn
85	Bi	Bismuth	114	Fl	Fl
86	Po	Po	115	Mc	Mc
87	At	At	116	Lv	Lv
88	Rn	Rn	117	Ts	Ts
113	Nh	Nihonium	118	Og	Oganesson
114	Fl	Fl			
115	Mc	Mc			
116	Lv	Lv			
117	Ts	Ts			
118	Og	Oganesson			
60	Nd	Neodymium	61	Pm	Promethium
62	Sm	Samarium	63	Eu	Europium
64	Gd	Gadolinium	65	Tb	Terbium
66	Dy	Dysprosium	67	Ho	Holmium
68	Er	Erbium	69	Tm	Thulium
70	Yb	Ytterbium	71	Lu	Lutetium

Data

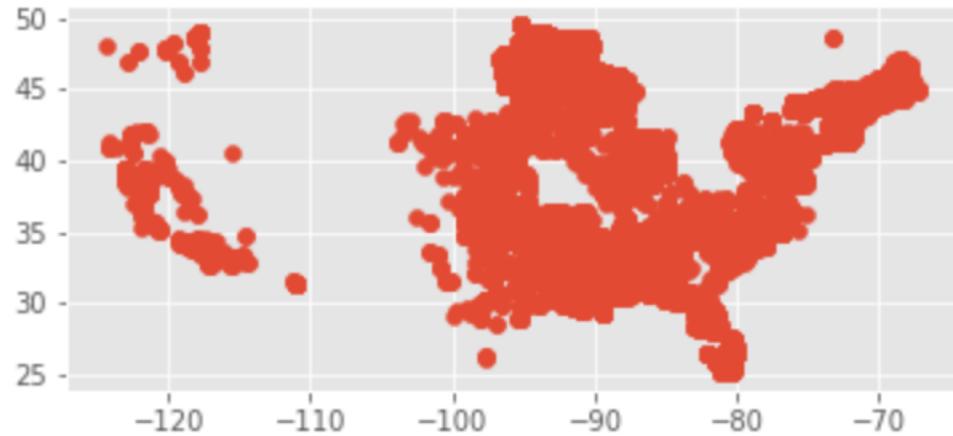
- Soil Geochemistry: National Geochemical Survey

- n= 64,756



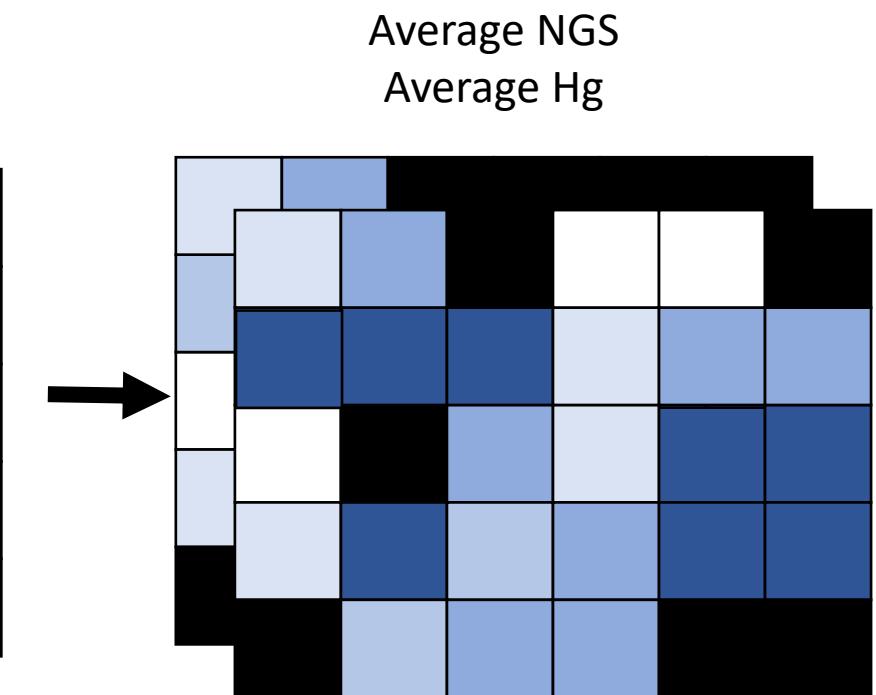
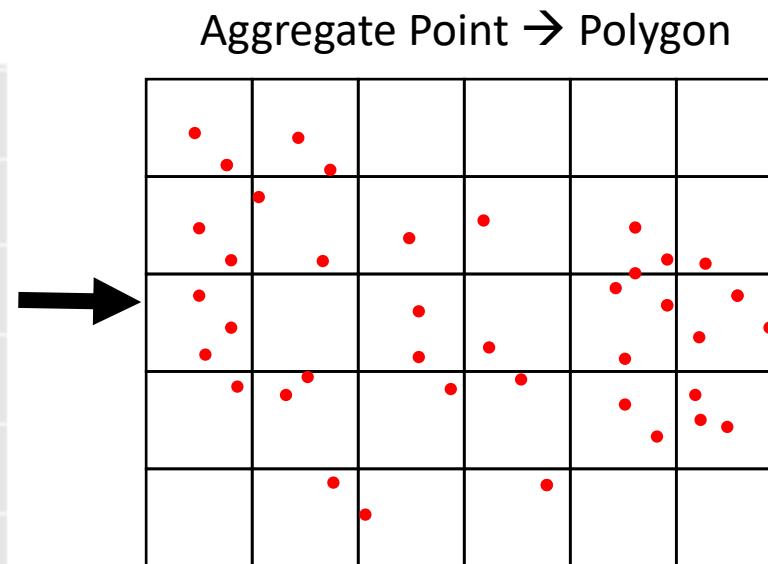
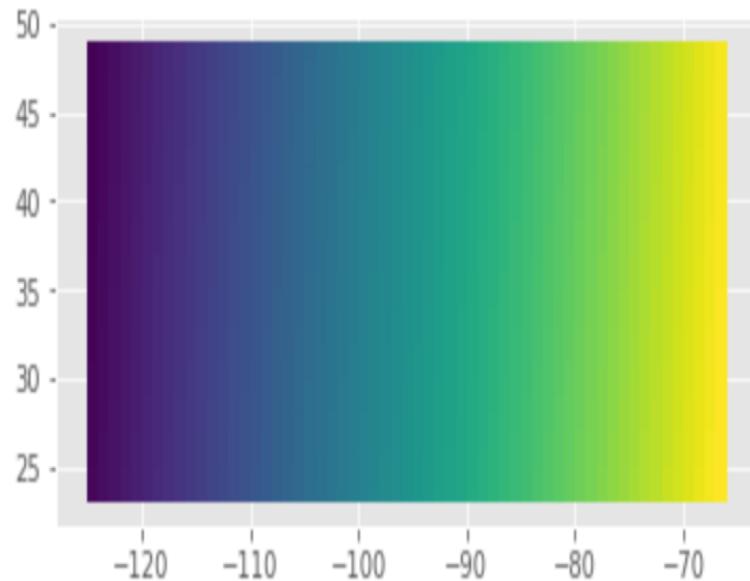
- Fish Tissue Hg: EPA Fish Hg database

- n= 19,705



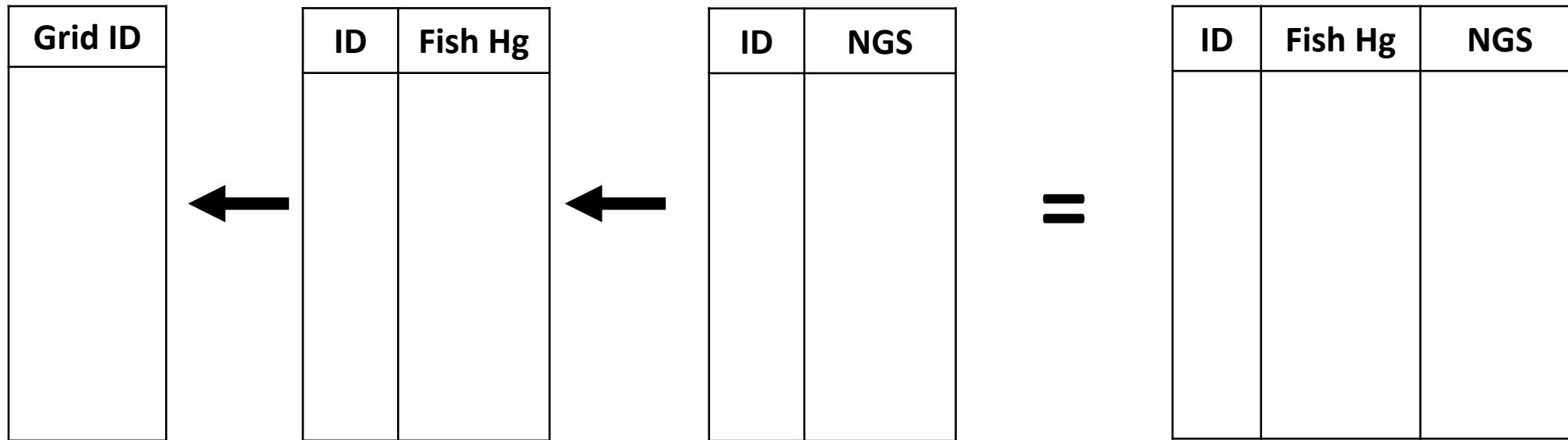
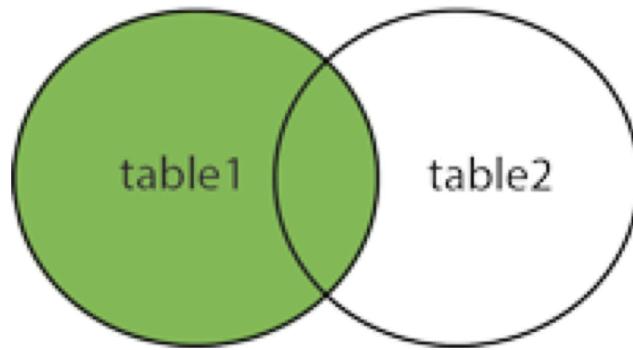
Data Processing

- Generate a grid
 - $n= 1534$



- Merge Data frames

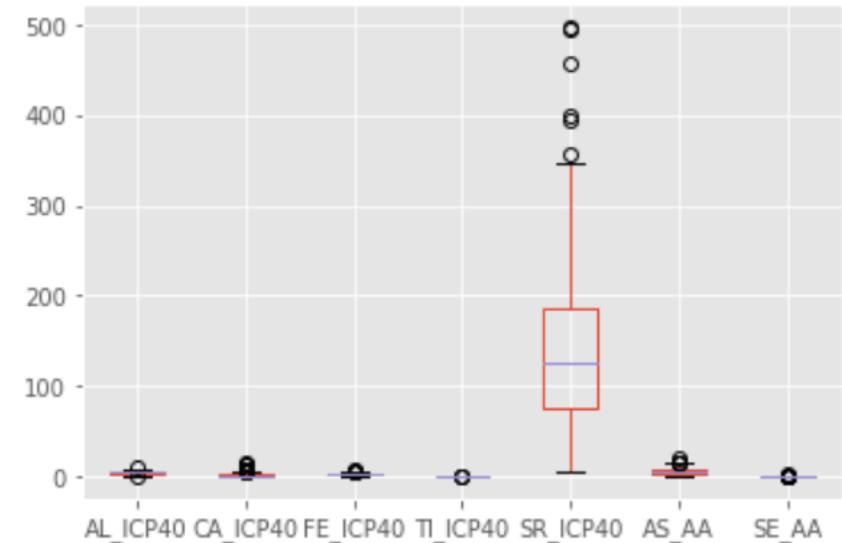
LEFT JOIN



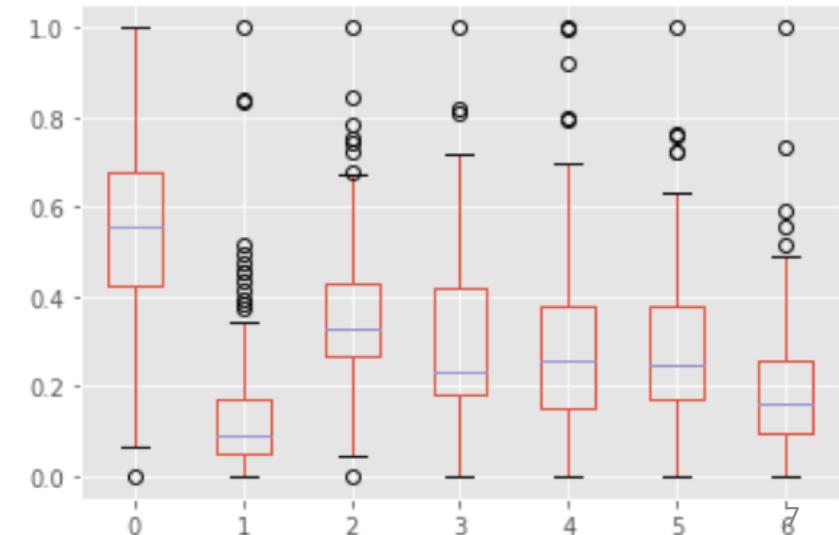
Scrub the Data

- Clean the dataset
 - Remove missing values (n=214)
- (Automated) Variable selection
 - Remove highly correlated variables ($r > \pm 0.80$)
 - Remove variables where the p-value of the correlation with the dependent variable is more than 0.20
 - Normalize the data (this method is more appropriate for ecotoxicology data)

Before Normalization



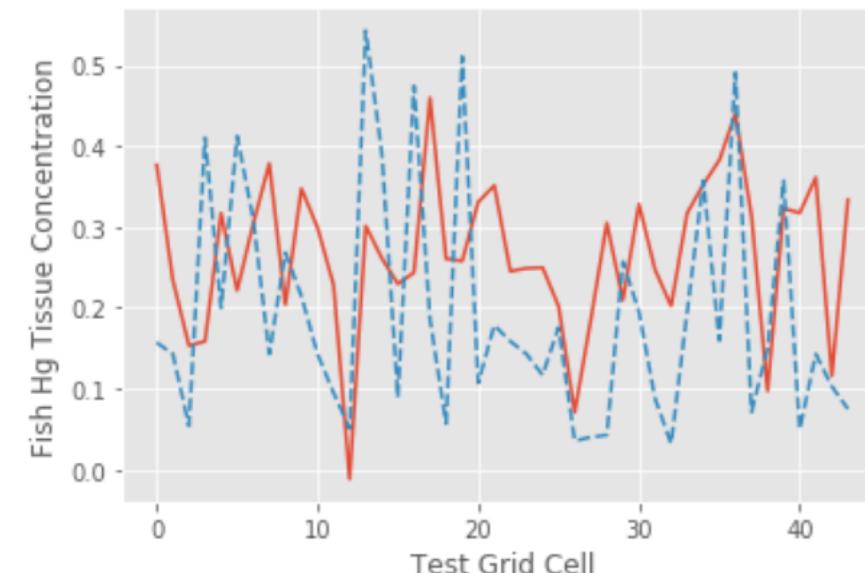
After Normalization



Model

- Ridge regression Results
 - The root mean square error = 0.18
 - The mean absolute error = 0.15
 - R² = 0.11
- Model coefficients are:
 - X1: AL_ICP40 -0.018
 - X2: CA_ICP40 -0.034
 - X3: FE_ICP40 -0.23
 - X4: TI_ICP40 0.032
 - X5: SR_ICP40 -0.16
 - X6: AS_AA -0.12
 - X7: SE_AA -0.36

OLS Regression Results						
Dep. Variable:	hg_mean	R-squared (uncentered):	0.327			
Model:	OLS	Adj. R-squared (uncentered):	0.291			
Method:	Least Squares	F-statistic:	9.092			
Date:	Tue, 03 Dec 2019	Prob (F-statistic):	4.02e-09			
Time:	16:32:10	Log-Likelihood:	-49.698			
No. Observations:	138	AIC:	113.4			
Df Residuals:	131	BIC:	133.9			
Df Model:	7					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
x1	0.8173	0.300	2.726	0.007	0.224	1.410
x2	0.3742	0.221	1.695	0.093	-0.063	0.811
x3	-1.0023	0.388	-2.582	0.011	-1.770	-0.234
x4	0.7521	0.193	3.904	0.000	0.371	1.133
x5	-0.5005	0.271	-1.850	0.067	-1.036	0.035
x6	0.1678	0.251	0.670	0.504	-0.328	0.663
x7	-0.2506	0.227	-1.103	0.272	-0.700	0.199
Omnibus:	157.709	Durbin-Watson:	2.143			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4002.281			
Skew:	4.216	Prob(JB):	0.00			
Kurtosis:	27.999	Cond. No.	14.4			



Untransformed data (but may violate OLS regression assumptions)

	AL_ICP40	CA_ICP40	FE_ICP40	TI_ICP40	SR_ICP40	AS_AA	SE_AA
AL_ICP40	1.000000	0.105773	0.742391	0.115026	0.736606	0.236462	0.079761
CA_ICP40	0.105773	1.000000	0.064647	-0.259561	0.379239	0.211274	0.227949
FE_ICP40	0.742391	0.064647	1.000000	0.431655	0.421571	0.247446	0.010555
TI_ICP40	0.115026	-0.259561	0.431655	1.000000	0.018363	-0.345852	-0.262499
SR_ICP40	0.736606	0.379239	0.421571	0.018363	1.000000	0.035092	0.080518
AS_AA	0.236462	0.211274	0.247446	-0.345852	0.035092	1.000000	0.464423
SE_AA	0.079761	0.227949	0.010555	-0.262499	0.080518	0.464423	1.000000

OLS Regression Results									
Dep. Variable:	hg_mean	R-squared (uncentered):	0.408						
Model:	OLS	Adj. R-squared (uncentered):	0.388						
Method:	Least Squares	F-statistic:	20.66						
Date:	Tue, 03 Dec 2019	Prob (F-statistic):	4.96e-21						
Time:	16:27:58	Log-Likelihood:	-35.355						
No. Observations:	217	AIC:	84.71						
Df Residuals:	210	BIC:	108.4						
Df Model:	7								
Covariance Type:	nonrobust								
	coef	std err	t	P> t	[0.025	0.975]			
AL_ICP40	0.0802	0.024	3.285	0.001	0.032	0.128			
CA_ICP40	0.0205	0.011	1.813	0.071	-0.002	0.043			
FE_ICP40	-0.1394	0.035	-3.995	0.000	-0.208	-0.071			
TI_ICP40	0.6279	0.113	5.533	0.000	0.404	0.852			
SR_ICP40	-0.0009	0.000	-2.299	0.022	-0.002	-0.000			
AS_AA	0.0169	0.008	2.131	0.034	0.001	0.033			
SE_AA	-0.2576	0.105	-2.461	0.015	-0.464	-0.051			
Omnibus:	262.458	Durbin-Watson:	1.430						
Prob(Omnibus):	0.000	Jarque-Bera (JB):	13980.961						
Skew:	5.031	Prob(JB):	0.00						
Kurtosis:	41.013	Cond. No.	1.04e+03						

Conclusions

- At this scale we were able to significantly model the relationship between environmental soil geochemistry and fish tissue Hg
- Overall good predictive power but still unexplained variance
- The order of operations matters
 - Difference between fields
- Future questions:
 - How does the grid location change the results?
 - How does the type of grid chosen affect these results? Would a honeycomb grid have been better?
 - How does the size of the grid change the results?