Draft Report

Proposal: Mapping and modeling soil toxic metal concentration based on remote sensing and environmental variables.

Contamination has been an anthropogenic environmental issue contributed by industrial activities and human disturbance. Compared to the air and water, soil is less monitored. However, remote sensing provides a new opportunity to model and detect the toxic metal concentration in soil over time with the incorporation of geographic and environmental attributes.

This study is to build a heavy metal spatial concentration model using variables (multi-spectral satellite images, geographic and environmental raster) and previous soil sample data to help calibration and validation.

Literature review:

In the last 15 years, there are much research to explore the potential of using remote sensing and spectral indices to model and map heavy metals in topsoil. Melendez-Pastor et al. studied and proved the significant role of remote sensing in soil toxic metals' detection. Especially the reflectance spectral information of minerals in soils **(clay mineral ratio)** could be used to detect heavy metals (Wu, YZ; Malley, D.F.; Ben-Dor, E.).

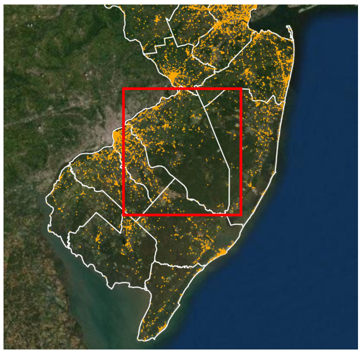
Choe et al.'s research showed significant correlations between **Pb, Zn, Cu and As** and different spectral absorption features (e.g., peak depth, area, asymmetry and band ratio).

Peng et al. (Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data) proposed the use of Landsat 8 imagery from January or February to extract spectral indices (including **band2-7, BCI, NDVI, Brightness, TVI**), in combination with auxiliary data like proximity to road to build a model using **Cubist tool** for estimating the heavy metal concentrations in soils. They mainly focus on **arsenic (As), chromium (Cr), nickel (Ni), copper (Cu), lead (Pb) and zinc (Zn)**, among which Cu works best.

Liu used Sentinel-2A multispectral imagery to investigate the stress exerted by heavy metals in soils on crops. Yun Yang et al. (Estimating the heavy metal concentrations in topsoil in the Daxigou mining area, China, using multispectral satellite imagery) used Landsat 8 Imagery with elevation to estimate **Cu, Pb, and As** concentrations in topsoil via proposed **GA–BP model**. They find significant correlations between the concentrations of Cu and **B2, B3, B4, CMR, MNDWI, Greenness, EVI, and NDVI**, and between Pb and **B2, B3, and B4, EVI and brightness factors**.

FarhadMirzaei et al. (Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2) applied Sentinel-2 satellite (**band 2-12**) to model the concentration of Pb, Cu, and Ni using linear regression, multivariate, and step-by-step method.

Study area:



Camden and Burlington county in New Jersey.

Camden and Burlington county are suburban areas where there is large amount of pervious soil surface, a certain amount of landfill facilities, and relatively homogenized land use. And sufficient soil sample data conducted by NJDEP previously also make it a great study area to research.

Data table:

		2001 (for model)	**2016** (for model)	**2022** (for spatial visualization)
Soil Sample	**Soil Sample Survey Data (80% for calibration, 20% for validation)**	1.NJDEP: characterization of ambient levels of selected metals and cpahs in new jersey soil	2.Site Remediation and Waste Management Program (SRWMP), PAH data	
Potential Variables	**Landcover (to identify pervious surface)**	3.NLCD 2000(National Land Cover Database)	NLCD 2015(National Land Cover Database)	NLCD 2021(National Land Cover Database)
	Elevation, Slope, Aspect		Same DEM Data	
	Distance to disturbance (to roads, landfills)	Road data, Landfill data (filter by active year)	Road data, Landfill data (filter by active year)	Road data, Landfill data (filter by active year)
	Precipitation/Temperature	4. NOAA 2001	NOAA 2016	NOAA 2022
	Satellite images (11 bands and multi-spectral images)	5. USGS (Landsat 8) 2001-02	USGS (Landsat 8) 2016-01	USGS (Landsat 8) 2016-01
Model Results	**Heavy Metal Spatial Concentration**	Spatial concentration of antimony, arsenic, barium, cadmium, calcium, copper, lead, manganese, potassium, selenium, sodium, vanadium, and zinc How they change over time?		

Note: the data source and exploration (datasets 1-5) is in the reference part of this proposal

Potential Spectral Indices (based on Literature review):

-Band2- band7

- **BCI** (biophysical composition index) = $\frac{(H+I)/(2-V)}{(H+I)/(2+V)}$

-**NDVI** (normalized differential vegetation index) = $\frac{B5-B4}{B5+B4}$

-TVI (transformed vegetation index) = $\left(\frac{B5-B4}{B5+B4}+0.5\right)^{1/2} \times 100$

-Brightness = $0.3561 \times B2 + 0.3972 \times B3 + 0.3904 \times B4 + 0.6966 \times B5 + B6 \times 0.2286 + B7 \times 0.1596$

-CMR (clay mineral ratio) = B6/B7

-**MNDWI** (improved normalised water index) **=** (B3−B6)/(B3+B6)

-EVI (enhanced vegetation index) = 2.5×(B5−B4)/(B5+6×B4−7.5×B2+1)

-Greenness = −0.294×B2−0.243×B3−0.542×B4+0.728×B5+0.071×B6−0.161×B7

Method:

1) Spectral Indices and Environmental Variables

Spectral indices, which are generated from bands of satellite image, could be used to reflect the soil properties associated with heavy metals distribution. The research is to use Landsat-7 image in 2001 winter, Landsat-8 images in 2016 and 2022 winter. (According to the articles, winter satellite image can better reflect soil conditions)

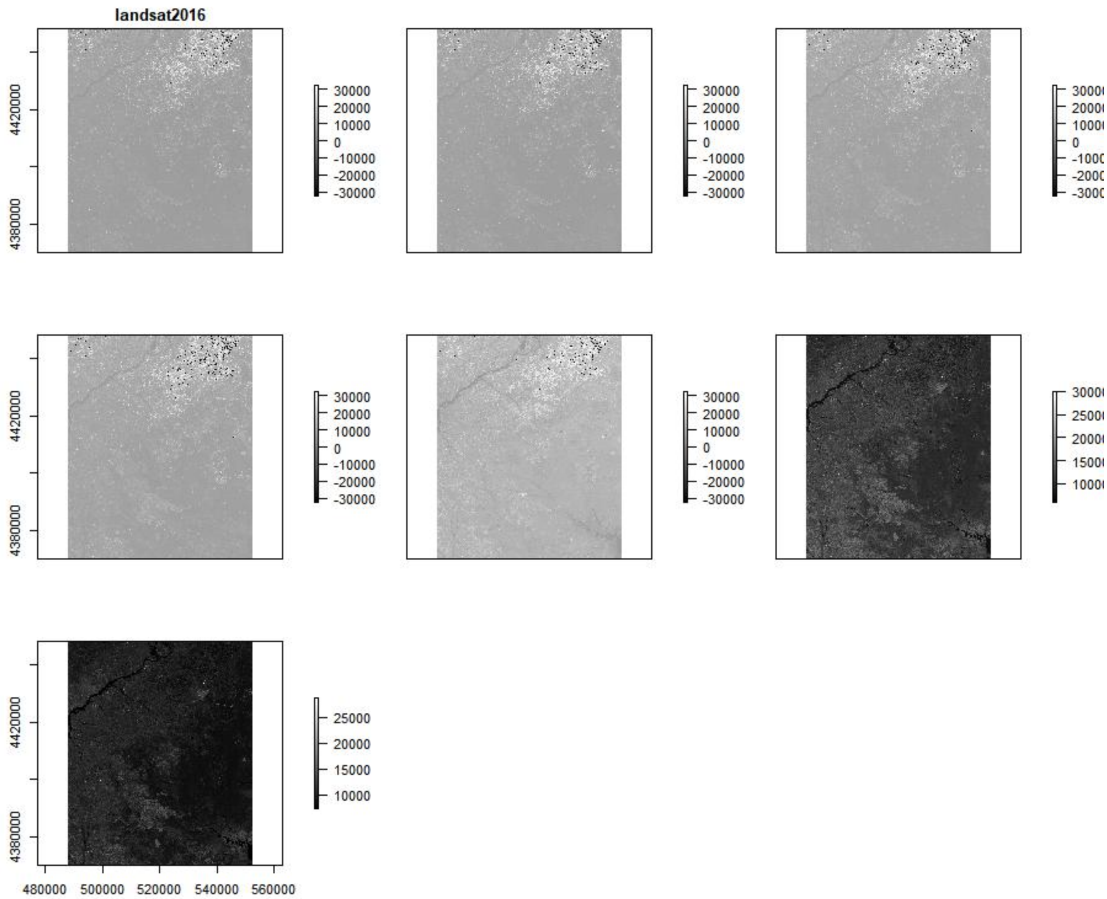
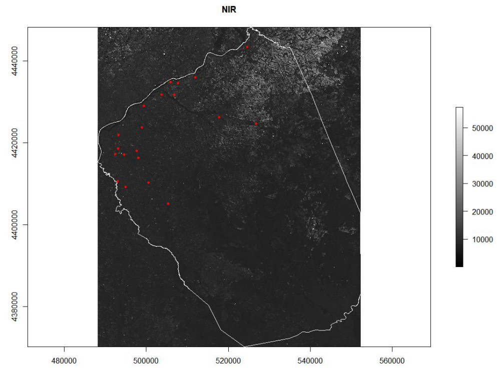


Fig1. 2016 Landsat-8 band1-band



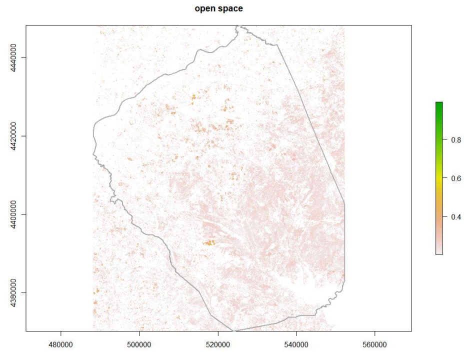


Fig2. study areas and sample points in 2016

Fig3. NDVI map

Besides that, environmental variables like precipitation, elevation, aspect, slope, water flow and anthropogenic variables like distance to human disturbance, roads and brownfield can also impact the soil and vegetation conditions and can be included as potential variables in the model. (attached in the Datasets sample)

2) Modeling and Calibration Using Previous Soil Sample Data

NJDEP PAH surveys conducted in 2001 and 2016 have covered the significant chemical data of the soil samples taken from different sites. (in study area: 24 sample location with 2 depth and 22 metals measurement data covered in the 2016 soil survey; 14 sample location with 22 metals measurement data covered in the 2001 survey)

These metal sample data could be used for regression analysis to determine the most related variables and help the model calibration and validation. (Cubist tool or GA–BP model are used in relevant research)

With the 2016 Landsat bands data extracted from sample points, the initial analysis is to study the correlation between bands or indices and corresponding heavy metal amount, which is the Arsenic. After remove the outlier, the result shows there is a correlation between Arsenic amount and the band5, band3, band4, band2.

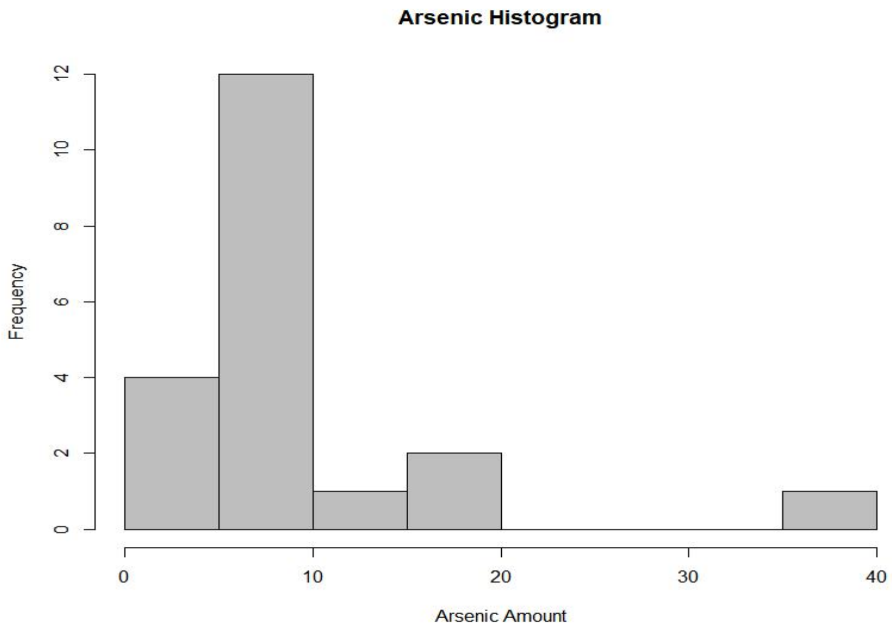


Table1. Arsenic Histogram showing a potential outlier

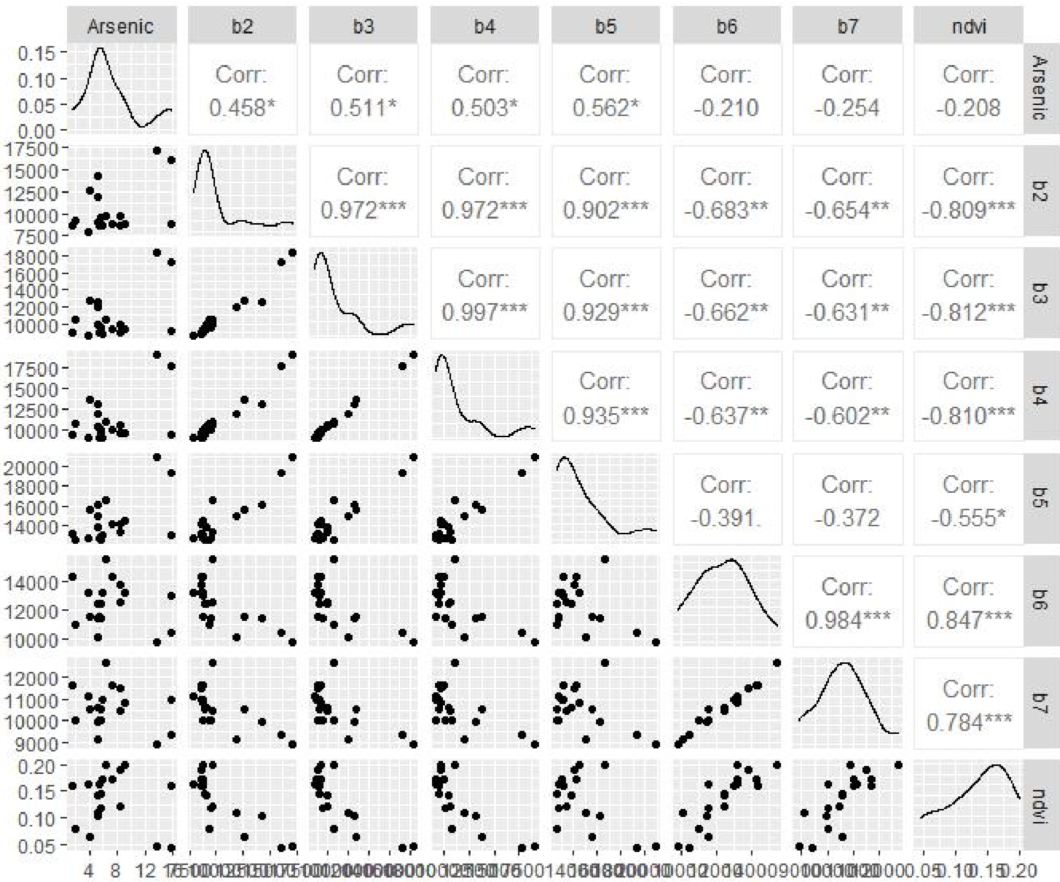


Table2. Correlation between different bands and Arsenic concentration

3) How will this be used

With the model built, once a user inputs the latest parameters including satellite image bands of the whole study area, precipitation, temperature and other real-time data, the model could be used to visualize the spatial distribution of toxic metal concentration in soil and to identify areas in a higher risk of soil contamination and its change over time.

Deliverable:

The final deliverable of this study can be a research paper.

Following questions:

-Different metal might have correlation with different contributors. It could be necessary to select several metals into categories according to other reference articles. Use regression to identify related variables for each category.

-How to choose a specific method to calibrate the model?

Reference:

1) Related Literature:

Wu, Y.Z.; Chen, J.; Ji, J.F.; Gong, P.; Liao, Q.L.; Tian, Q.J.; Ma, H.R. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. Soil Sci. Soc. Am. J. 2007, 71, 918–926.

Estimating the heavy metal concentrations in topsoil in the Daxigou mining area, China, using multispectral satellite imagery: https://www.nature.com/articles/s41598-021-91103-8

Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data: https://www.mdpi.com/2072-4292/8/12/1003/htm

Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2: https://www.sciencedirect.com/science/article/pii/S1110982321000223#f0025

2) Datasets sample

1. NJDEP: characterization of ambient levels of selected metals and cpahs in new jersey soil, https://www.nj.gov/dep/dsr/publications/Characterization%20of%20Ambient%20Levels%20of%20Selected%20Metals%20and%20cPAHs%20in%20NJ%20Soils_Year%20Three_Highlands,%20Valley%20and%20Ridge,%20and%20Coastal%20Plain%20_Full%20Report.pdf
2. Site Remediation and Waste Management Program (SRWMP), https://www.nj.gov/dep/dsr/health/statistics-metals-soil.pdf (Raw data in /RawData/PAH_sample_2017/)

TABLE 11 SUMMARY OF COASTAL PLAIN RURAL SOIL DATA										
	Sample ID:	NJDEP NRDC Cleanup Criteria	NJDEP RDC Cleanup Criteria	CP-64	CP-65	CP-66	CP-67	CP-68	CP-69	CP-70A
Analyte	Date:	03-May-1999	03-May-1999	09-Jun-2001	09-Jun-2001	09-Jun-2001	09-Jun-2001	09-Jun-2001	30-Nov-2000	30-Nov-2000
Metals										
Silver		4100	110	0.18 U	0.19 U	0.20 U	0.19 U	0.18 U	0.079 U	0.093 U
Aluminum		NA	NA	1620	4230	4390	1620	5900	5250	9560
Arsenic		20	20	0.59 B	3.0	2.4	1.1 B	1.5	4.1	8.2
Barium		47000	700	6.0 B	16.8 B	14.9 B	6.7 B	7.8 B	0.14 U	43.1
Beryllium		2	2	0.056 U	0.19 B	0.083 B	0.060 U	0.057 B	0.023 U	0.027 U
Calcium		NA	NA	60.4 B	70.4 B	106 B	65.3 B	59.4 B	3.6 U	4.3 U
Cadmium		100	39	0.068 U	0.071 U	0.077 U	0.072 U	0.066 U	0.034 U	0.040 U
Cobalt		NA	NA	0.48 B	2.2 B	0.91 B	0.47 B	0.61 B	0.057 U	0.067 U
Chromium		NA	120000	2.8	5.7	5.0	2.6	4.9	5.4	12.4
Copper		600	600	3.5	4.1	4.0	2.6 B	3.4	10.1	15.2 J
Iron		NA	NA	1790	4970	3810	1860	3120	4760	11700
Mercury		270	14	0.047	0.044	0.061	0.033 B	0.053	0.10 J	0.31 J
Potassium		NA	NA	71.7 B	135 B	228 B	127 B	130 B	3.8 U	4.4 U
Magnesium		NA	NA	94.4 B	250 B	324 B	126 B	224 B	1.3 U	943
Manganese		NA	NA	12.2 J	252 J	18.4 J	11.2 J	10.6 J	14.6 J	171 J
Sodium		NA	NA	66.3 B	65.6 B	75.8 B	51 B	76 B	9.8 U	12 U
Nickel		2400	250	0.81 B	2.6 B	2.5 B	0.92 B	2.2 B	0.15 U	8.1
Lead		600	400	13.9	20.6	18	16.2	14.1	36.6	250
Antimony		340	14	0.34 U L	0.35 U L	0.41 BL	0.36 U L	0.33 U L	0.19 U L	0.23 U L
Selenium		3100	63	0.39 U	0.41 U	0.45 U	0.42 U	0.38 U	1.1	0.41 U
Thallium		2	2	0.71 U	0.74 U	0.81 U	0.76 U	0.69 U	0.41 U	0.48 U
Vanadium		7100	370	8.0	11.1	13.7	8.1	12.8	20.8	22.5
Zinc		1500	1500	3.8	9.3	9.2	3.9	6.4	11.6	44

Sample ID	Municipality	County	Population Density (2010)	Distance to nearest KCSL (ft)	Area Type	Soil Type	Sample Depth	Aluminum	Antimony	Arsenic	Barium	Beryllium	Cadmium
CAMD01	OAKLYN BORO	CAMDEN	5769	658	Open	Urban land	shallow	8670	1.70	8.4	69.5	0.46	0.630
CAMD02	BERLIN BORO	CAMDEN	2102	1468	Open	Mullica sandy loam	shallow	3360	0.00	1.6	6.7	0.00	0.072
CAMD03	CHERRY HILL TWP	CAMDEN	2939	1780	Open	Fluvaquents	shallow	3520	0.89	8.4	11.3	0.41	0.120
CAMD06	GLOUCESTER TWP	CAMDEN	2776	4560	Open	Tinton sand	shallow	1970	0.39	7.4	10.2	0.12	0.000
CAMD07	HADDON TWP	CAMDEN	5215	1903	Open	Freehold-Downer-Urban land complex	shallow	8140	1.20	13.6	79.5	0.62	0.710
CAMD07	HADDON TWP	CAMDEN	5215	1903	Open	Freehold-Downer-Urban land complex	shallow	9000	1.60	17.3	79.6	0.68	0.640
CAMD07	HADDON TWP	CAMDEN	5215	1903	Open	Freehold-Downer-Urban land complex	shallow	6050	1.30	8.7	74.4	0.52	0.950
CAMD08	VOORHEES TWP	CAMDEN	2507	3269	Forested	Buddtown-Deptford fine sandy loams	shallow	6720	0.00	5.8	22.9	0.38	0.160
CAMD11	CAMDEN CITY	CAMDEN	7394	3040	Open	Urban land	shallow	3800	1.60	6.3	56.6	0.36	0.260
CAMD12	COLLINGSWOOD BORO	CAMDEN	7216	938	Open	Urban land	shallow	4040	1.90	5.2	30.0	0.24	0.096
CAMD13	HADDONFIELD BORO	CAMDEN	4082	2067	Forested	Freehold-Downer-Urban land complex	shallow	2500	0.32	5.4	15.9	0.31	0.150
CAMD14	RUNNEMEDE BORO	CAMDEN	4032	2652	Forested	Freehold-Downer-Urban land complex	shallow	4560	0.71	6.0	19.5	0.26	0.240

Table1,2: Soil sample data in 2001 and 2016

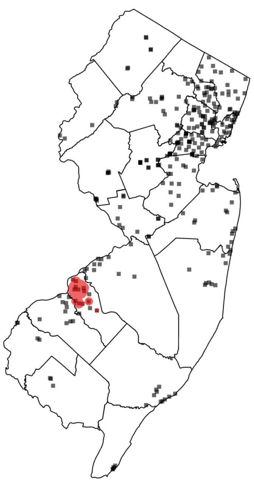


Fig.2016 soil sample data location

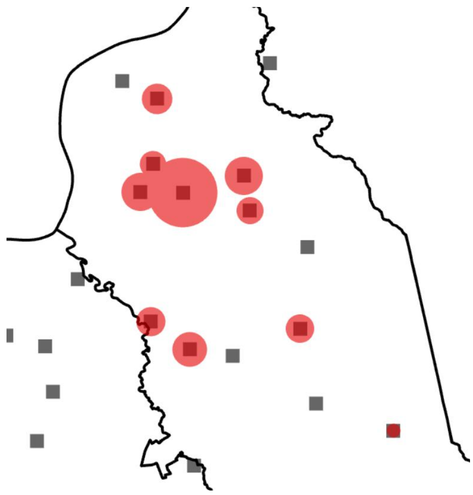
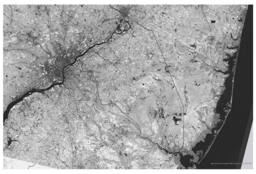
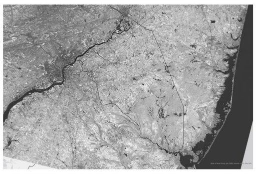
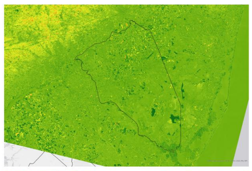
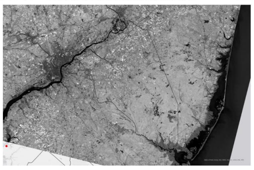
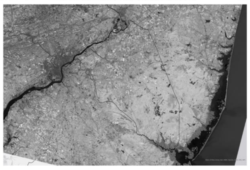
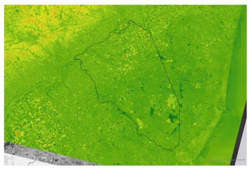


Fig. 2016 Arsenic Measurement from sample data

3. NLCD (National Land Cover Database)
4. Precipitation/ Temperature based on monitoring station (2000-2021, monthly)

https://www.ncei.noaa.gov/data/global-summary-of-the-month/

5. Landsat 8 | U.S. Geological Survey - USGS.gov, (Satellite images: in winter, less than 10%cloud cover) 2001, 2026, 2022

	2013	2021	
Clay mineral ratio = Band6/Band7			
normalised vegetation index = (Band5 − Band4)/(Ban5+Ban4)			
MNDWI = (B3−B6)/(B3+B6)	……	……	……
DVI = B5/B4	……	……	……

6. Environmental Variables: distance to disturbance