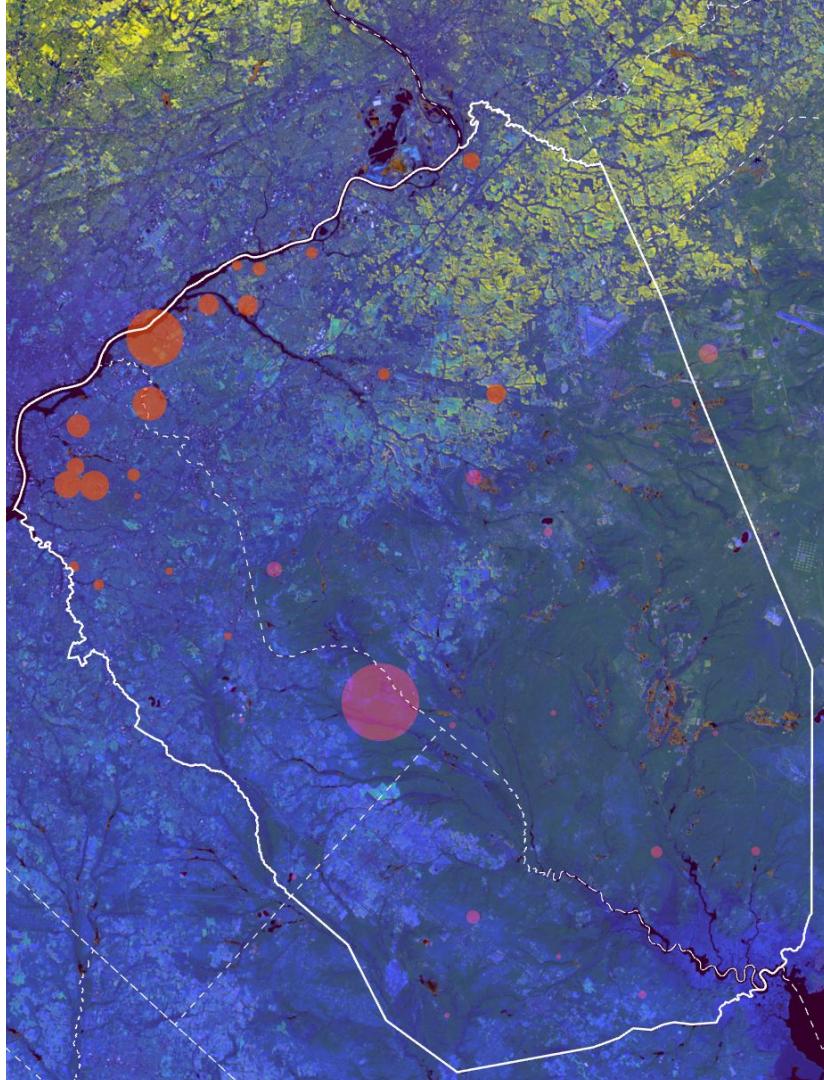


Mapping and modeling toxic metal concentration in soil

based on multi-spectral images
and environmental variables

Yuehui Gong

03/23/2022



A large white dump truck is shown from a low angle, its rear hatch open as it dumps a massive pile of trash onto a sprawling landfill. The ground is covered in a thick layer of discarded plastic bags, bottles, and other waste. In the background, a large flock of seagulls is captured in flight against a pale sky.

Human activities are great contributor to contamination and toxic metal concentration in the contaminated media (air, water, soil)

Soil is heavily contaminated and under-monitored

Traditional Practice

Soil Survey



New Opportunity

Published: 20 December 2009

Mineral composite assessment of Kelkit River Basin in Turkey by means of remote sensing

Hakan Mete Dogan 

Journal of Earth System Science 118, Article number: 701 (2009) | [Cite this article](#)

352 Accesses | 19 Citations | [Metrics](#)

Abstract

Utilizing remote sensing (RS) and geographic information systems (GIS) tools, mineral composite characteristics (ferrous minerals (FM), iron oxide (IO), and clay minerals (CM)) the Kelkit River Basin (15913.07 km^2) in Turkey were investigated and mapped. Mineral composite (MC) index maps were produced from three LANDSAT-ETM+ satellite images taken in 2000. Resulting MC index maps were summarized in nine classes by using 'natural breaks' classification method in GIS. Employing bi-variate correlation analysis, relationship among index maps were investigated. According to the results, FM and IO index maps show positive correlation, while CM index map is negatively correlated with FM and IO index ma

<https://link.springer.com/article/10.1007/s12040-009-0059-9>

Article | Open Access | Published: 03 June 2021

Estimating the heavy metal concentrations in topsoil in the Daxigou mining area, China, using multispectral satellite imagery

Yun Yang , Qinfang Cui, Peng Jia, Jinbao Liu & Han Bai

Scientific Reports 11, Article number: 11718 (2021) | [Cite this article](#)

962 Accesses | 1 Citations | [Metrics](#)

Abstract

A precise estimation of the heavy metal concentrations in soils using multispectral remote sensing technology is challenging. Herein, Landsat8 imagery, a digital elevation model, and geochemical data derived from soil samples are integrated to improve the accuracy of estimating the Cu, Pb, and As concentrations in topsoil, using the Daxigou mining area in Shaanxi Province, China, as a case study. The relationships between the three heavy metals and soil environmental factors were investigated. The optimal combination of factors associated with the elevated concentrations of each heavy metal was determined combining correlation analysis with collinearity tests. A back propagation network optimised using a genetic algorithm was trained with 80% of the data for samples and subsequently employed to estimate the heavy metal concentrations in the area. The validation results show that the

<https://www.nature.com/articles/s41598-021-91103-8>

Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2

Farhad Mirzaei , Yasser Abbasi , Teymour Sohrabi 

Show more 

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.ejrs.2021.03.002>

Under a Creative Commons license

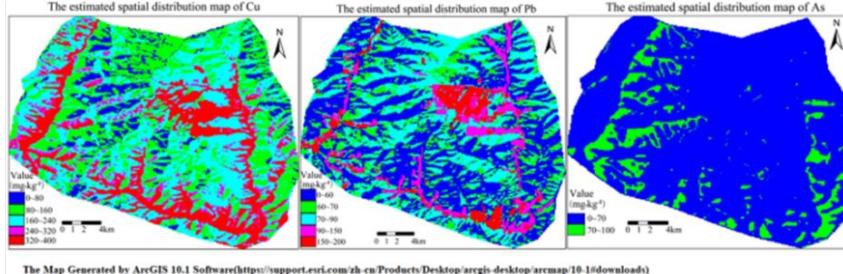
Get rights and content

 Open access

Abstract

The pollution of heavy metals is considered as one of the main problems of using wastewater for irrigation purposes. Frequent experimental measurements, time, and cost are essential for evaluating the pollution of heavy metals in a large area. Thus, using satellite images and establishing a relationship between the images and concentration of heavy metals can be regarded as a solution for estimating the

<https://www.sciencedirect.com/science/article/pii/S1110982321000223#f0025>



Published: 20 December 2009

Mineral composite assessment of Kelkit River Basin in Turkey by means of remote sensing

Hakan Mete Dogan 

Journal of Earth System Science 118, Article number: 701 (2009) | [Cite this article](#)

352 Accesses | 19 Citations | [Metrics](#)

Abstract

Utilizing remote sensing (RS) and geographic information systems (GIS) tools, mineral composite index maps of ferromanganese (FM), iron oxide (IO), and transition metals (TM) in the Kelkit River Basin, Turkey, were produced. Mineral composite index maps (MC) and mineral composite (MC) index maps were produced from three LANDSAT-ETM+ satellite images taken in 2000. The MC index maps were produced by using the breaks' classification method in GIS, employing bivariate correlation analysis, relationships among index maps were investigated. According to the results, FM and IO index maps show positive correlation, while CM index map is negatively correlated with FM and IO index ma

Remote sensing provides a new opportunity to model and detect the distribution of toxic metal in soil

Article | Open Access | Published: 03 June 2021

Estimating the heavy metal concentrations in topsoil in the Daxigou mining area, China, using multispectral satellite imagery

Yun Yang , Qinfang Cui, Peng Jia, Jinbao Liu & Han Bai

Scientific Reports 11, Article number: 11718 (2021) | [Cite this article](#)

962 Accesses | 1 Citations | [Metrics](#)

Abstract

A precise estimation of the heavy metal concentrations in soils using multispectral remote sensing technology is challenging. Herein, Landsat8 imagery, a digital elevation model, and a soil sample dataset from soil samples are used to estimate the heavy metal concentrations in the Daxigou mining area, Shaanxi Province, China, as a case study. The relationships between the three heavy metals and environmental factors were investigated. The optimal combination of factors was determined by the correlation coefficient. The relationship between the determined coefficients and concentration of heavy metals was analyzed by using a correlation analysis with collinearity tests. A back propagation network optimized using a genetic algorithm was trained with 80% of the data for samples and subsequently employed to estimate the heavy metal concentrations in the area. The validation results show that the

Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2

Farhad Mirzaei , Yasser Abbasi , Teymour Sohrabi 

Show more 

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.ejrs.2021.03.002>

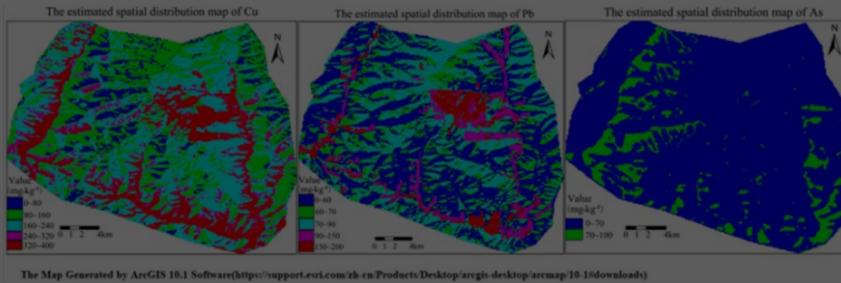
Get rights and content

Open access

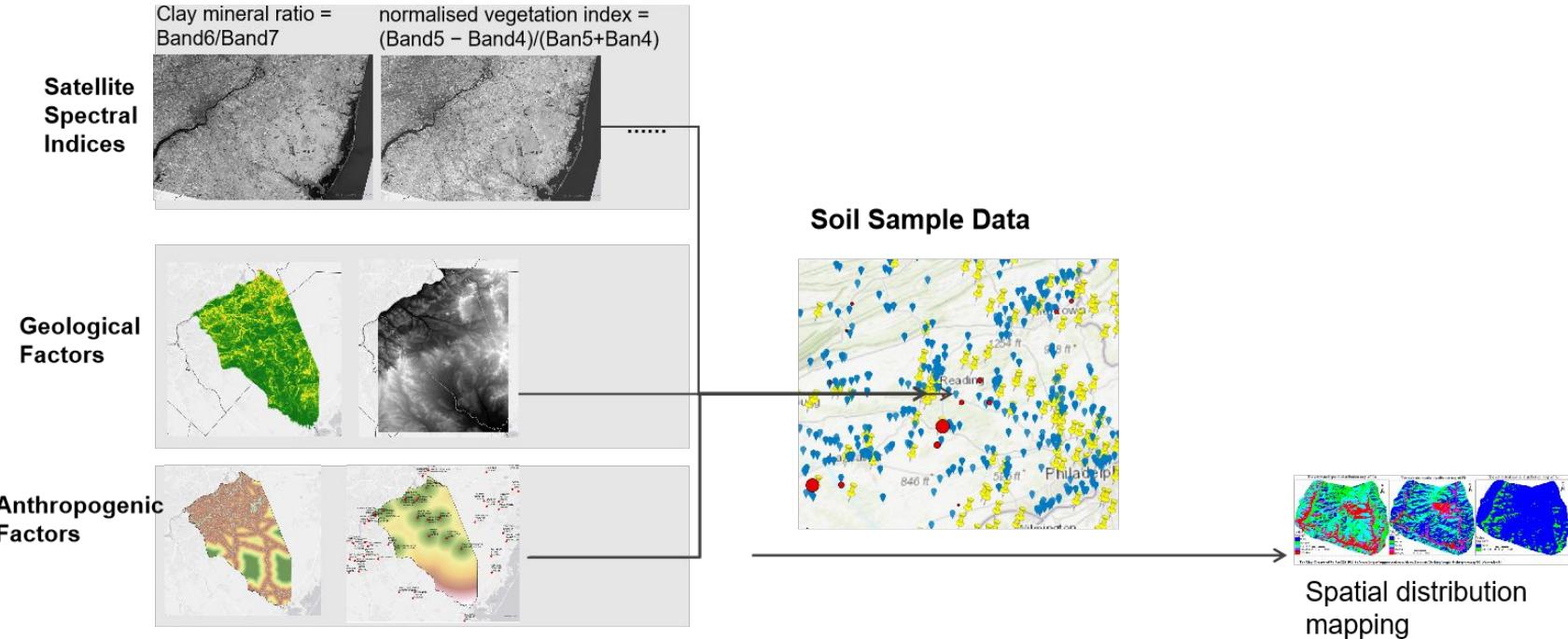
<https://link.springer.com/article/10.1007/s12040-009-0059-9>

<https://www.nature.com/articles/s41598-021-91103-8>

<https://www.sciencedirect.com/science/article/pii/S1110982321000223#f0025>

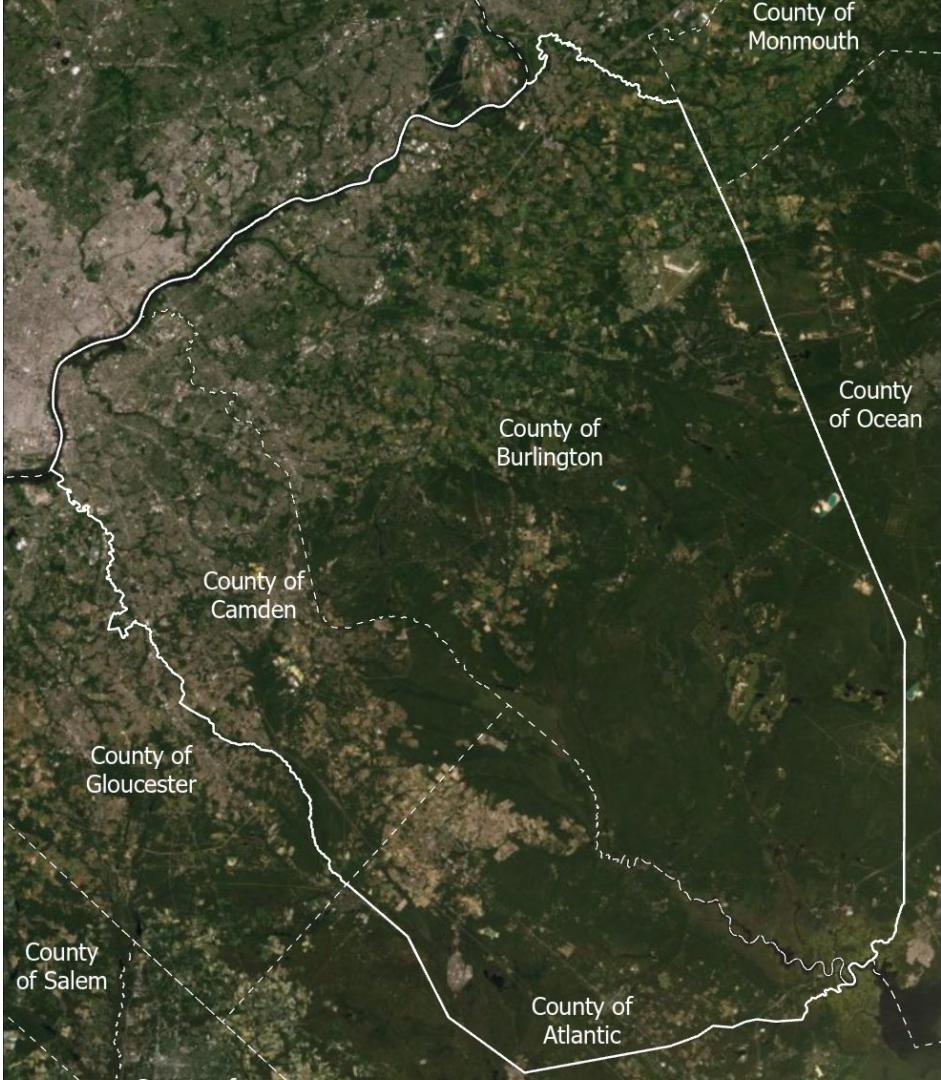
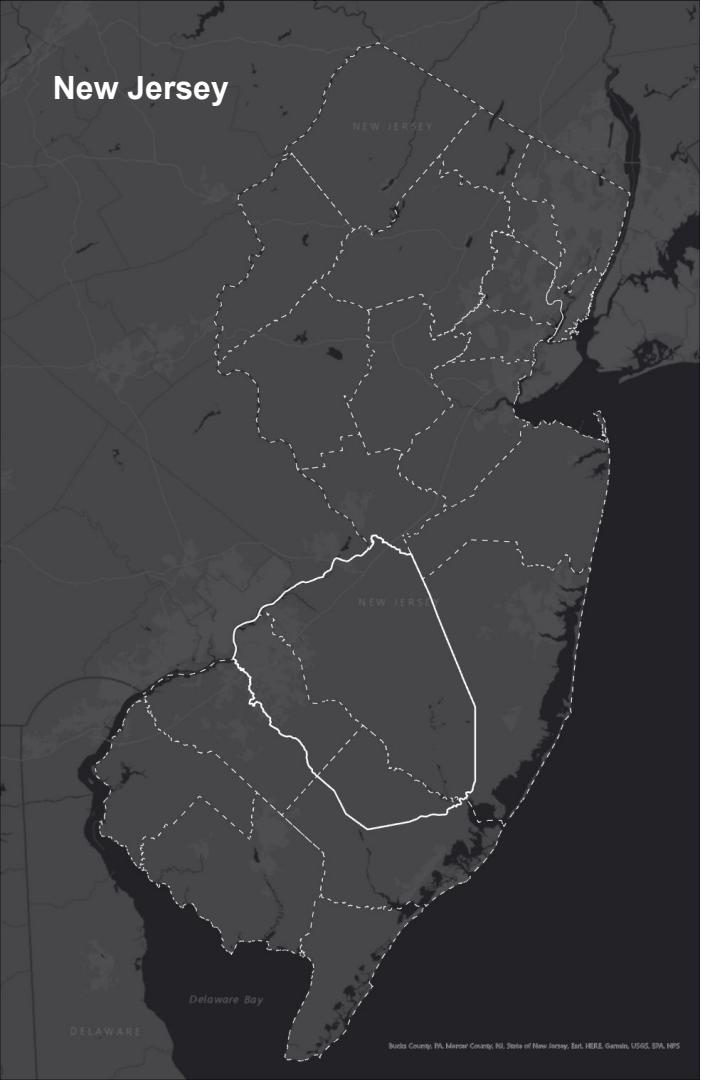


Methodology



Model Training and Testing

Study Area



Dataset

	2001 (Model Building)	2016 (Model Building)	2022 (for spatial visualization)
Training: 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	Land cover (to identify pervious surface)	3.NLCD 2000(National Land Cover Database)	NLCD 2015(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)
	Satellite images (11 bands and multi-spectral images)	5. USGS (Landsat 8) 2001-02	USGS (Landsat 8) 2016-01
Model Results	Heavy Metal Spatial Concentration	Spatial concentration of Arsenic, Lead, Copper, Chromium	

Dataset:

Soil survey sample data

2001 Survey

-23 metal

-NJDEP (New Jersey Department of Environmental Protection)

Study area comprised rural areas of the Valley and Ridge, Highlands, and Coastal Plain physiographic provinces.

SUMMARY OF COASTAL PLAIN RURAL SOIL DATA										
Sample ID:	NJDEP NRD/Cleanup Criteria	NJDEP RDC/Cleanup Criteria	CP-64	CP-65	CP-66	CP-67	CP-68	CP-69	CP-70A	CP-70B
Analyte	Date:	03-May-1999	03-May-1999	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	184.4 J	112.2 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

https://www.nj.gov/dep/dsr/publications/Characterization%20of%20Ambient%20Levels%20of%20Selected%20Metals%20and%20PAHs%20in%20NJ%20Soils_Year%20Three_HIGHLANDS,%20VALLEY%20and%20RIDGE,%20and%20Coastal%20Plain%20_Full%20Report.pdf

2016 Survey

-22 metal

-SRWMP (Site Remediation and Waste Management Program)

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	Fluvaequents	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

<https://www.nj.gov/dep/dsr/health/statistics-metals-soil.pdf>

Dataset:

Soil survey sample data

2001 Survey Samples

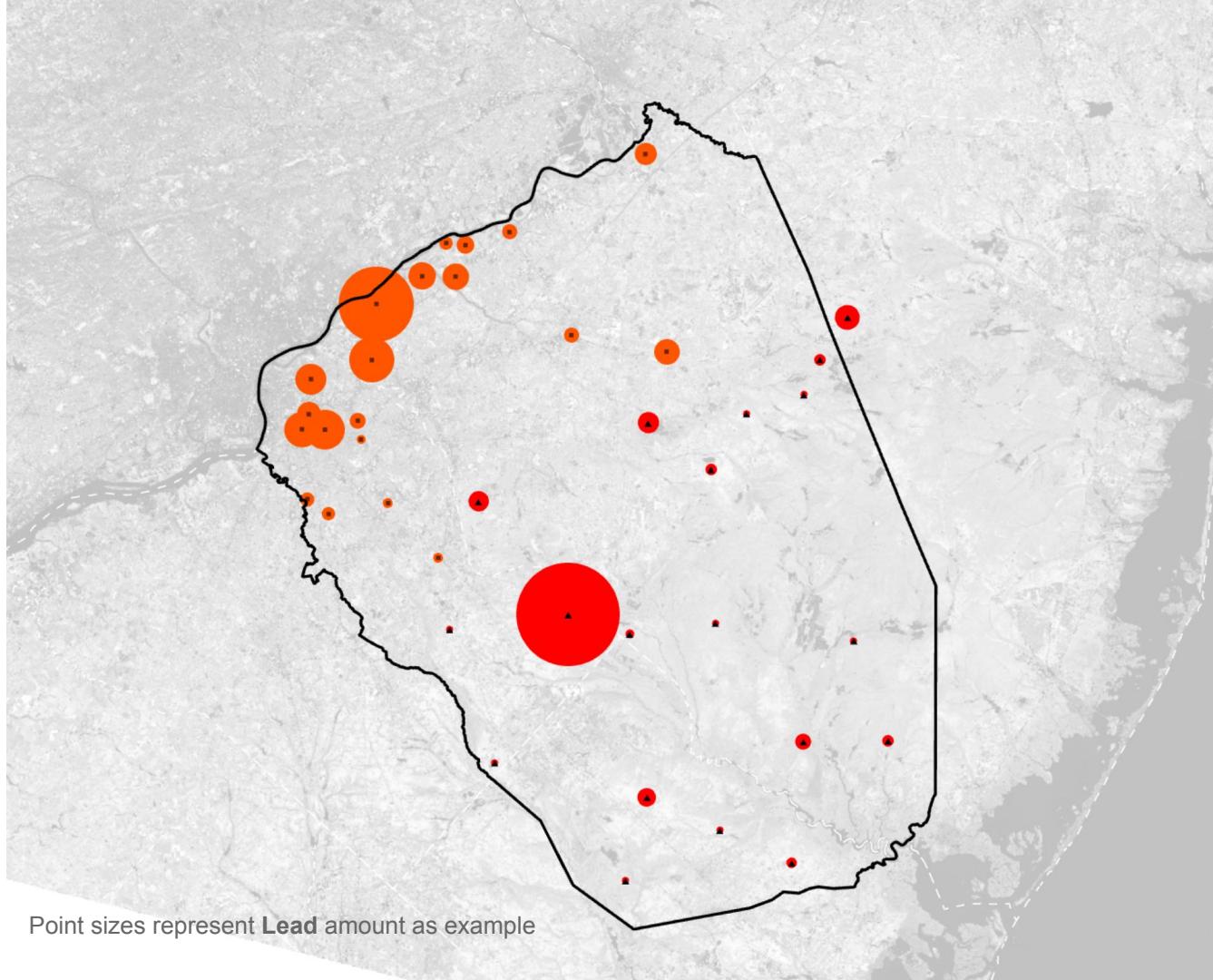


20 sample points

2016 Survey Samples



19 sample points



Point sizes represent Lead amount as example

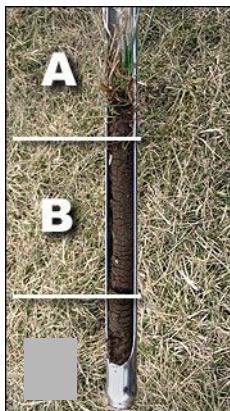
Dataset:

Soil survey sample data

Metal to study:

Arsenic, Copper, Lead, Chromium

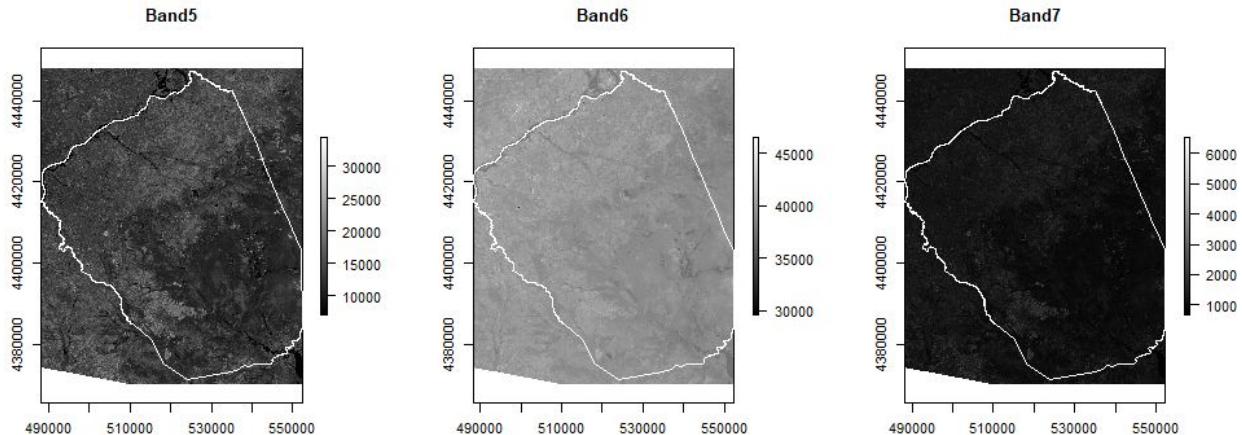
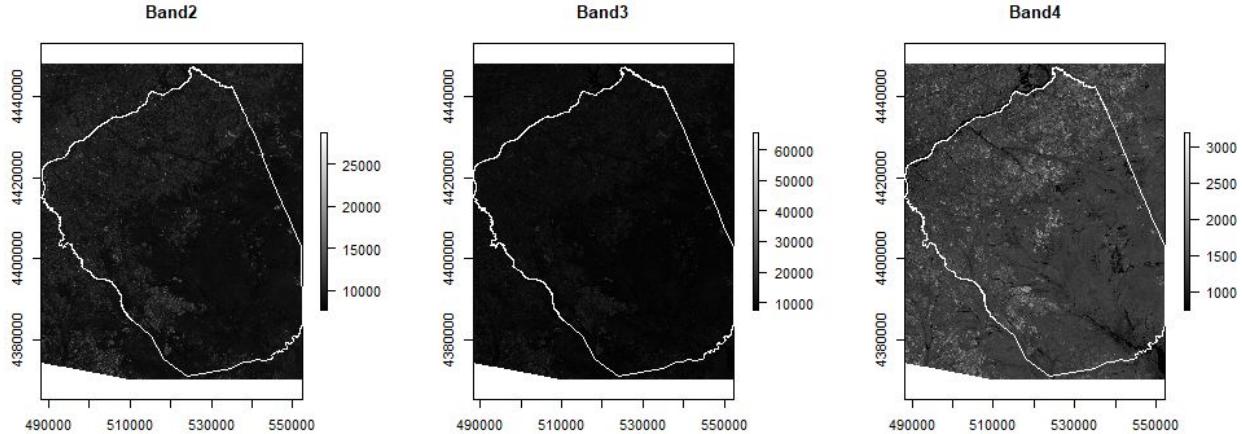
Metal concentrations were significantly higher in shallow soil samples, which help to detect metal in topsoil using remote sensing



Soil
sample
depth

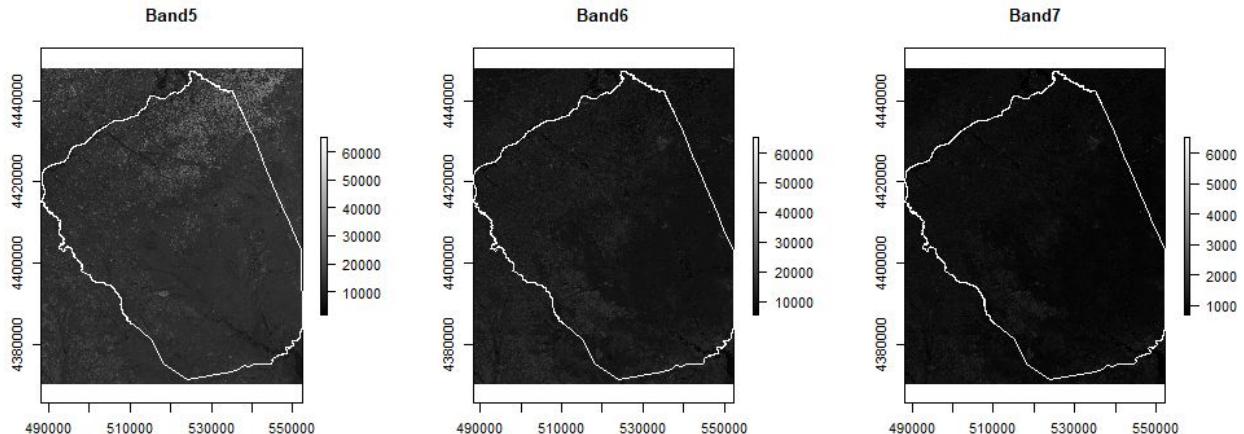
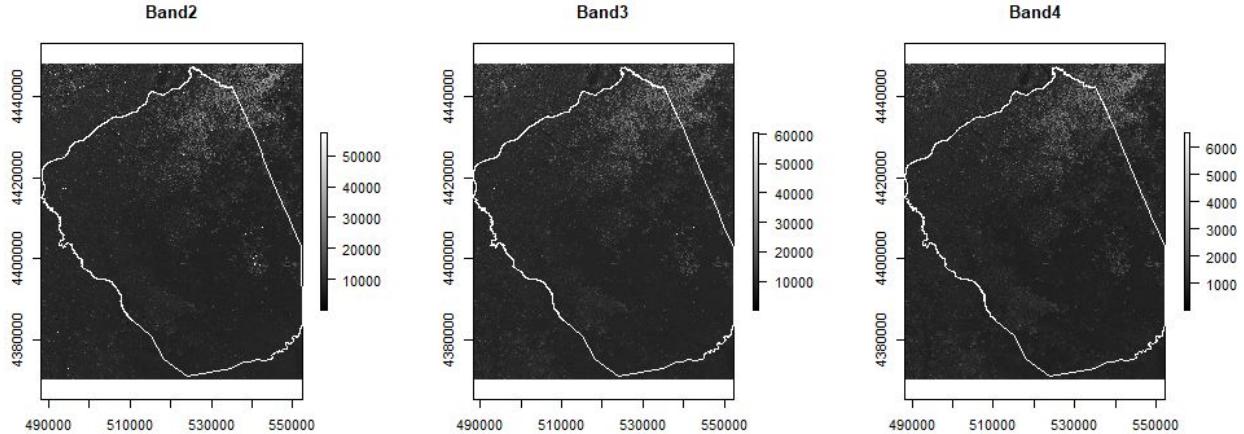
Dataset:

Satellite image:
Landsat 7 Band 1-7
(2002 Feb)



Dataset:

Satellite image:
Landsat 8 Band 2-7
(2016 Feb)



Dataset:

Satellite image: Spectral Indices

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

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

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

$$-\text{CMR} \text{ (clay mineral ratio)} = B6/B7$$

$$-\text{MNDWI} \text{ (improved normalised water index)} = (B3 - B6)/(B3 + B6)$$

$$-\text{EVI} \text{ (enhanced vegetation index)} = 2.5 \times (B5 - B4) / (B5 + 6 \times B4 - 7.5 \times B2 + 1)$$

$$-\text{Greenness} = -0.294 \times B2 - 0.243 \times B3 - 0.542 \times B4 + 0.728 \times B5 + 0.071 \times B6 - 0.161 \times B7$$

Literature reference:

Peng et al. (Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data):

Spectral indices: **band 2-7, BCI, NDVI, Brightness, TVI**

Model: **Cubist** tool

Metal: arsenic (As), chromium (Cr), nickel (Ni), copper (Cu), lead (Pb) and zinc (Zn), among which Cu works best

Yun Yang et al. (Estimating the heavy metal concentrations in topsoil in the Daxigou mining area, China, using multispectral satellite imagery)

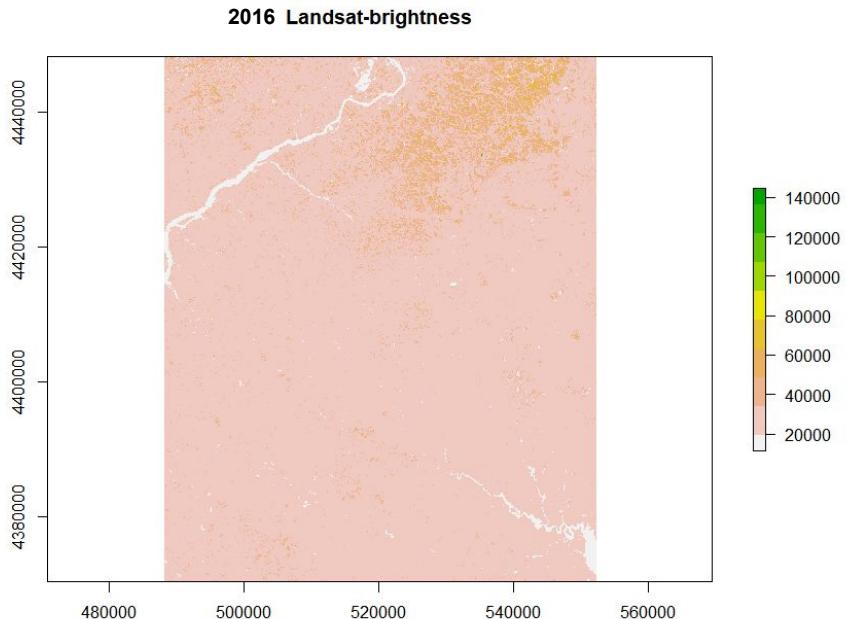
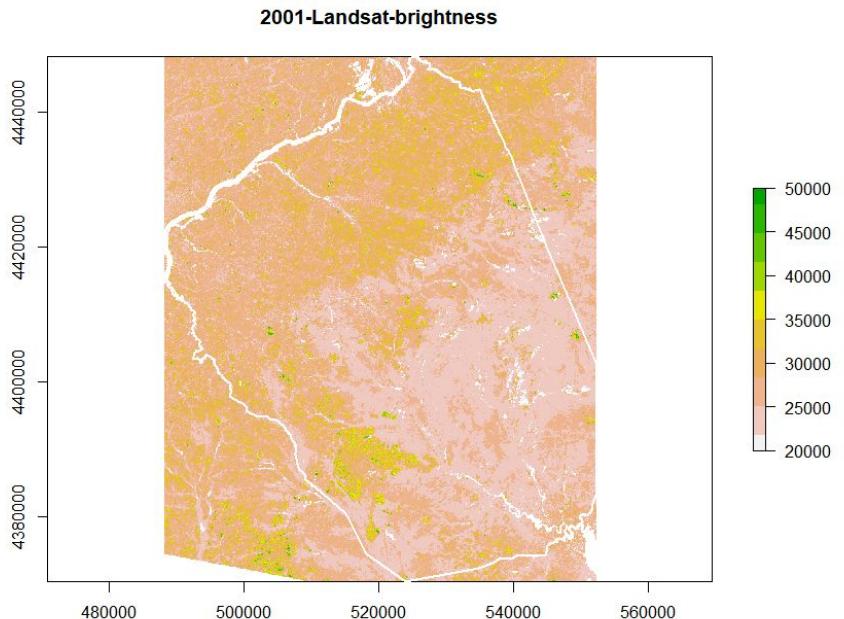
Metal: **Cu, Pb, and As**

Model: **GA-BP model**

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**

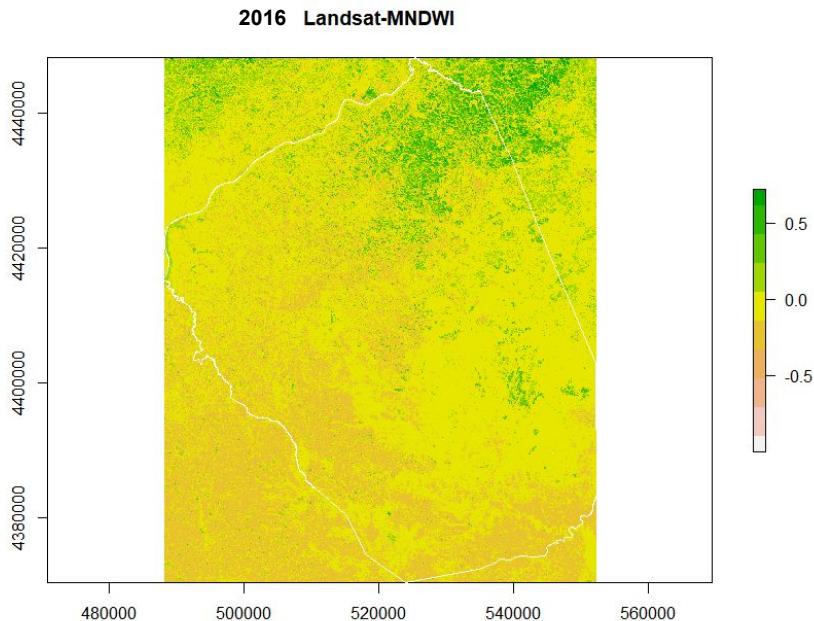
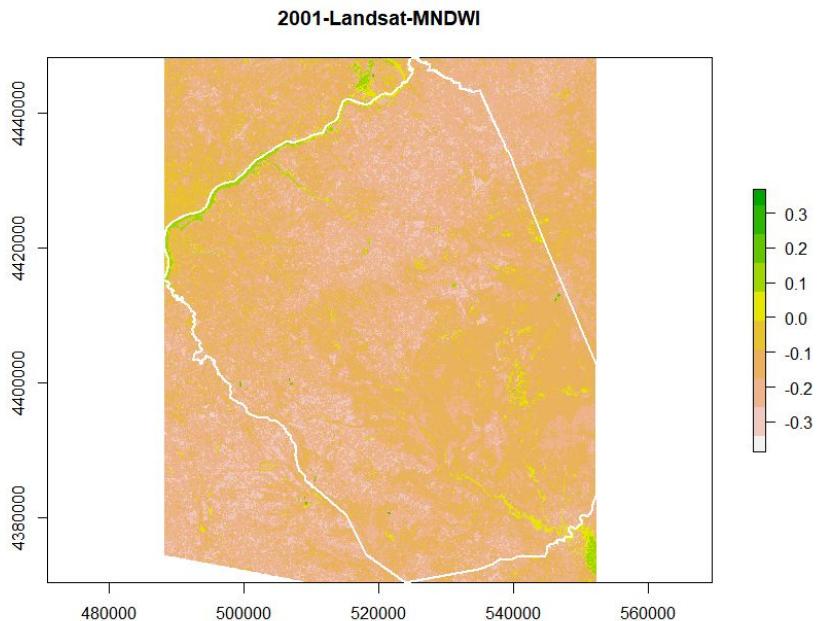
Dataset:

Satellite image: Spectral Indices



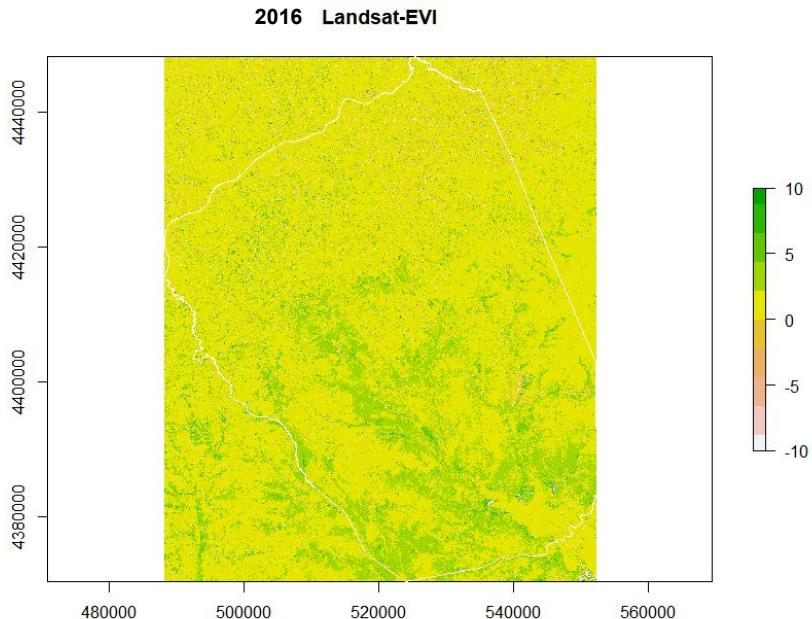
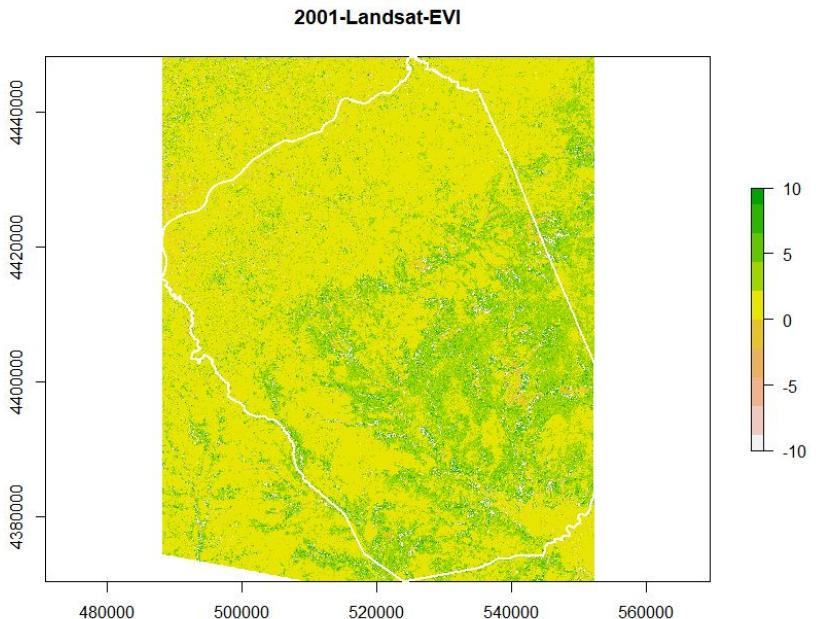
Dataset:

Satellite image: Spectral Indices



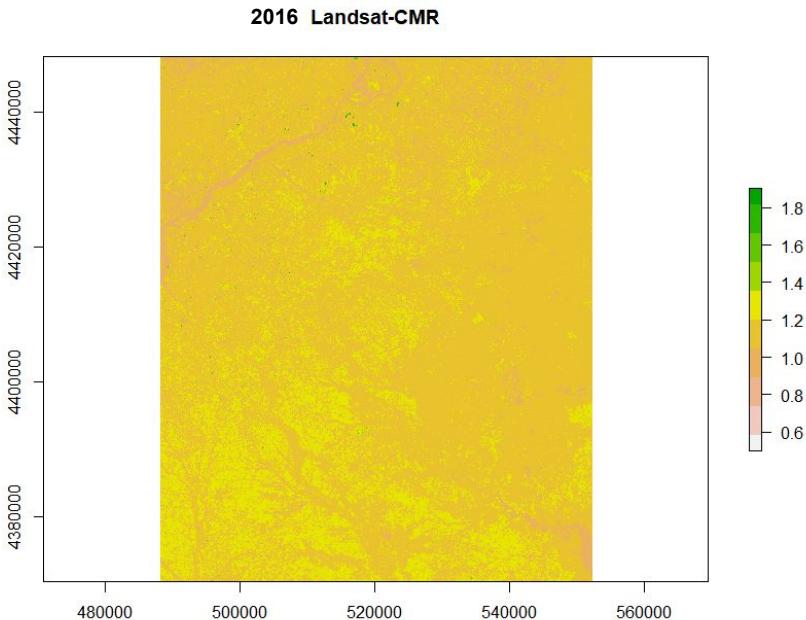
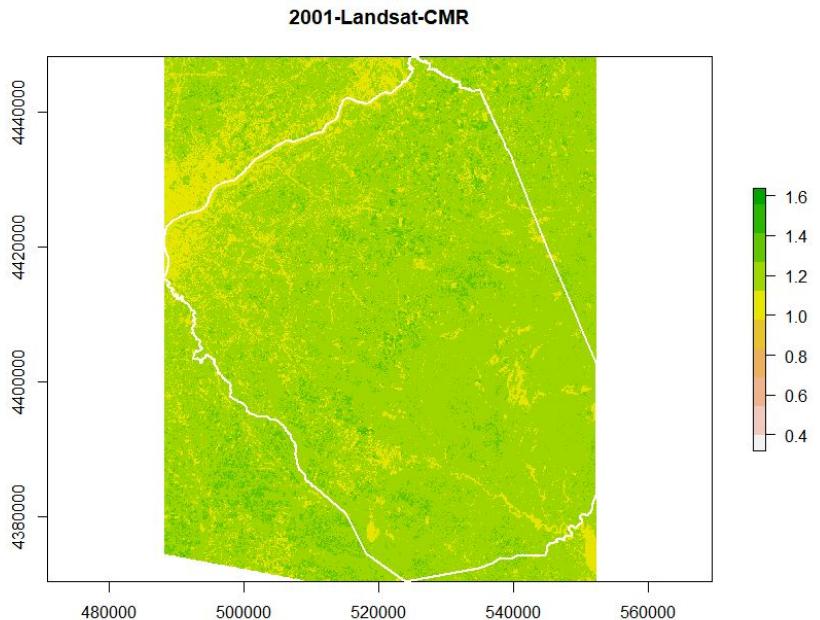
Dataset:

Satellite image: Spectral Indices



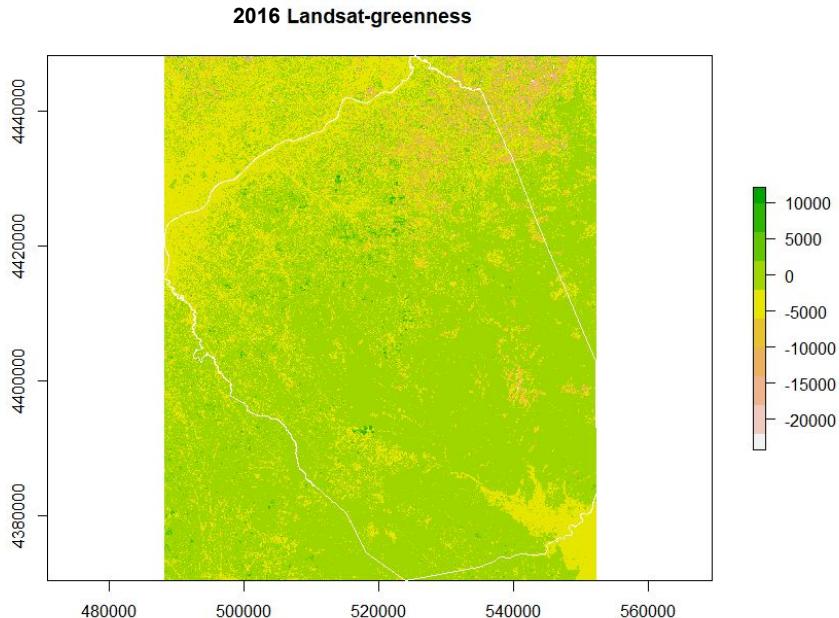
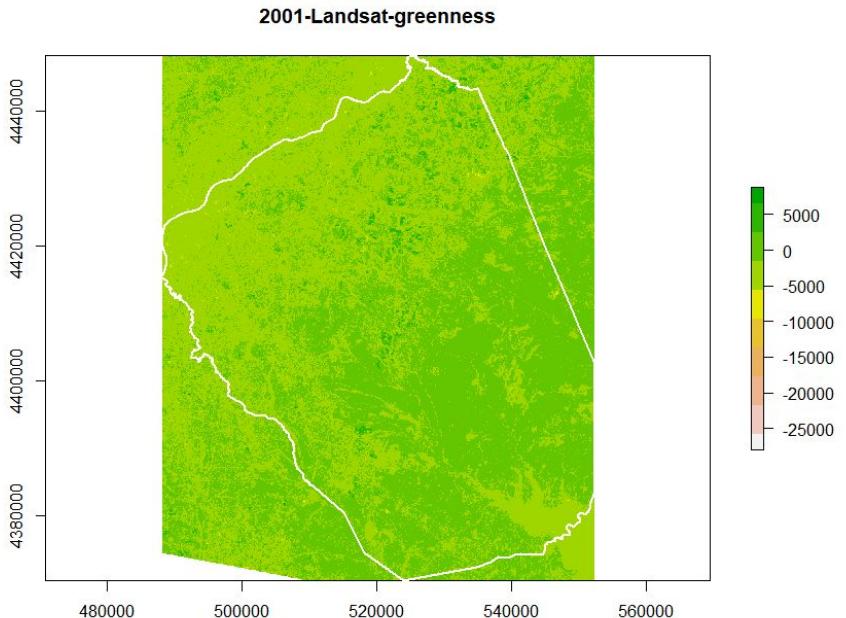
Dataset:

Satellite image: Spectral Indices



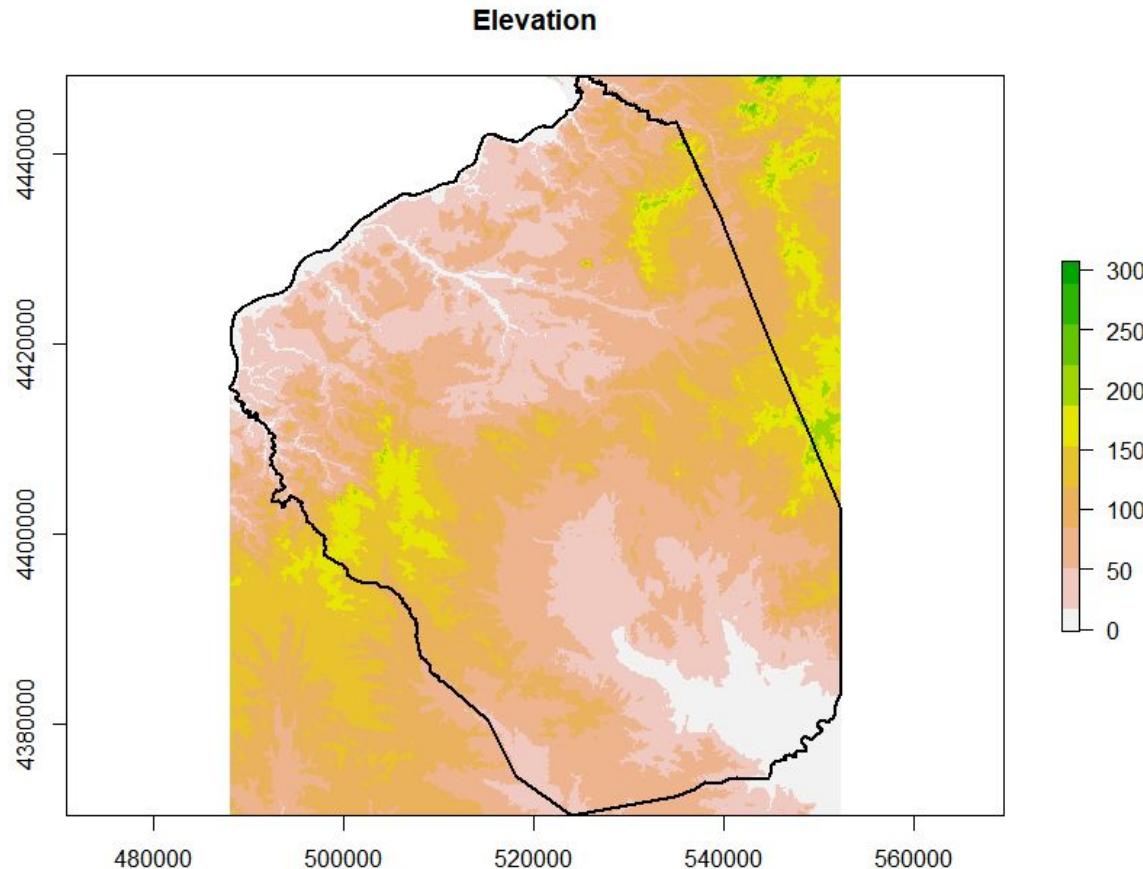
Dataset:

Satellite image: Spectral Indices



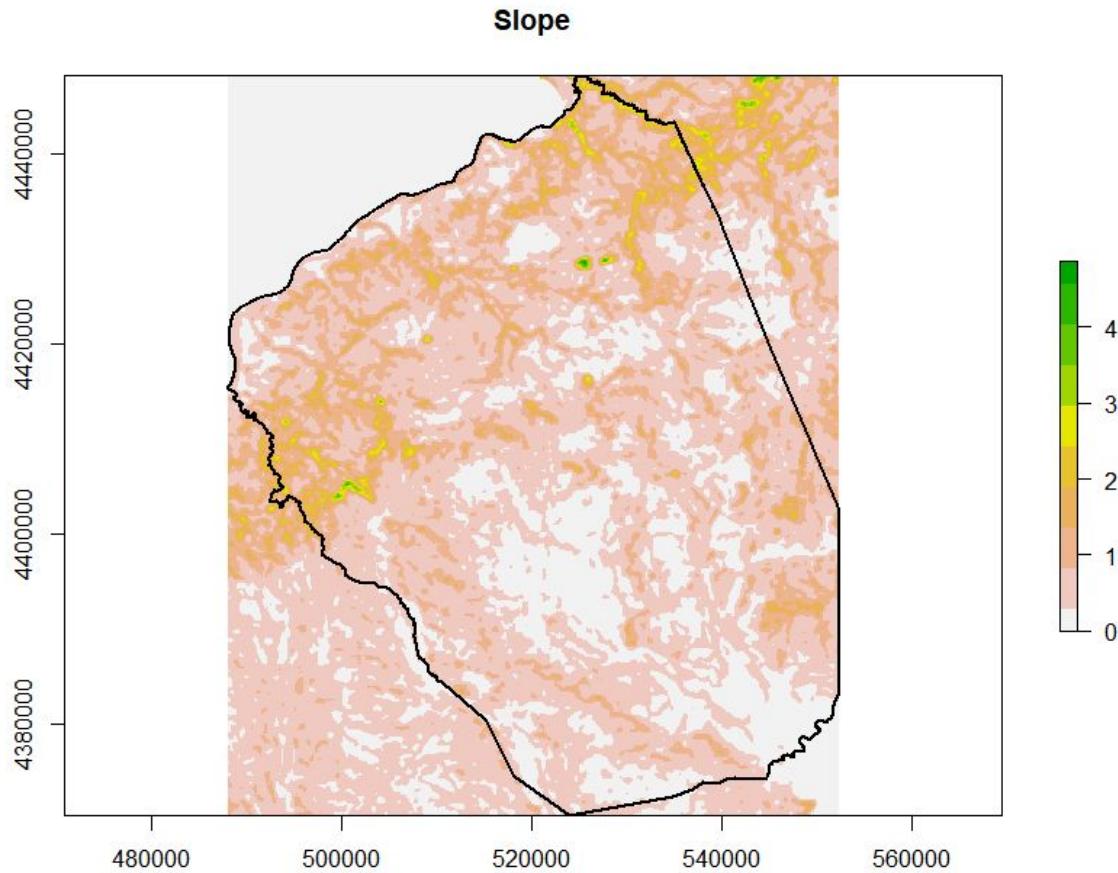
Dataset:

Elevation



Dataset:

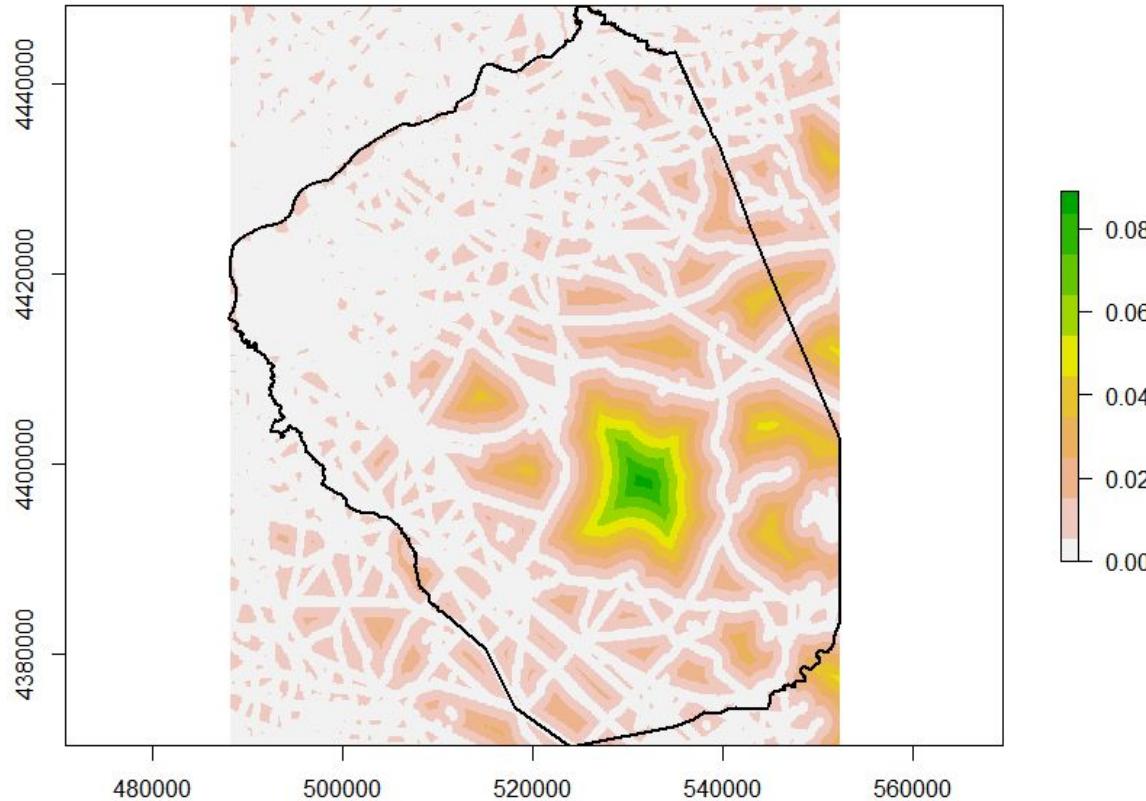
Slope



Dataset:

Disturbances

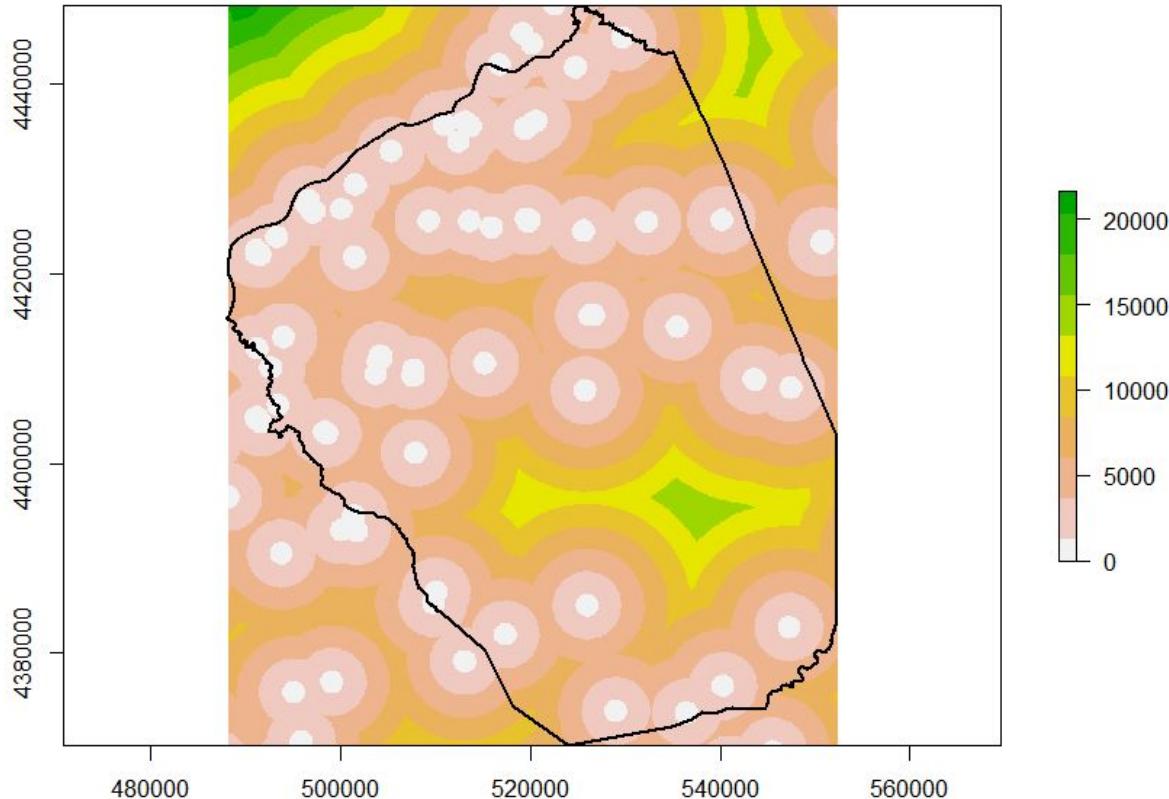
Distance to Roads



Dataset:

Disturbances

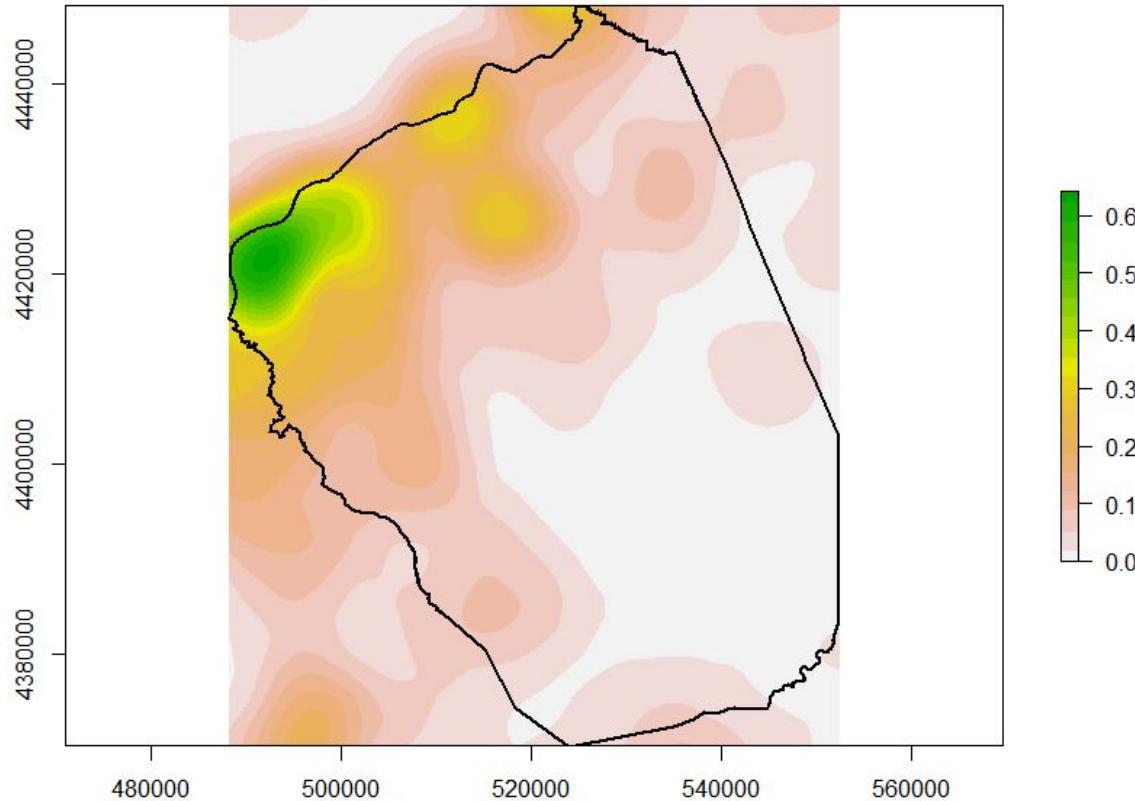
Distance to Waste management brownfields



Dataset:

Disturbances

Brownfields Density



Data Extracted

Sample_ID	Aluminum	Antimony	Arsenic	b2	b3	b4	b5	b6	b7	ndvi	CMR	brightness	greenness	MNDWI	EVI	dem	slope	disroad	dismili	diswaste	brown
BURL01	4600	0.67	35.6	8595	9137	9438	13638	12514	10512	0.18200728	1.190449	25577.82	-738.091	-0.15597431	1.8089414	31.82376	0.5350125	0.0030927309	7005.8833	3005.3953	0.41475967
BURL02	2970	0.00	2.1	9290	10555	10758	12625	11028	10010	0.07984433	1.101698	25392.34	-2764.583	-0.02191540	0.6224163	51.18048	1.4462776	0.0009420578	8293.3047	1281.6006	0.05859118
BURL03	6020	0.64	5.3	9046	9992	10414	13895	12478	10604	0.14319800	1.176726	26557.50	-1437.714	-0.11063641	1.0196251	40.02576	0.1505896	0.0033820907	3874.1838	2991.1370	0.23596206
BURL04	8210	0.75	9.2	8810	9295	9679	14510	13221	10873	0.19971888	1.215948	26716.02	-343.425	-0.17436489	1.8552227	22.63752	0.1811633	0.0000000000	417.8517	1073.3126	0.31157461
BURL05	7490	1.00	15.8	16091	17235	17682	19324	10476	9392	0.04437118	1.115417	35947.28	-5202.947	0.24391036	0.8670398	52.16472	1.6121854	0.0022204514	14940.1201	1406.1650	0.16274378
BURL06	4140	0.39	3.8	7859	8706	9145	12734	13194	11105	0.16403857	1.188113	25046.06	-963.473	-0.20493151	1.0357864	16.73208	1.7285587	0.0015700962	11540.2949	1652.4528	0.28151438
BURL07	4500	0.42	5.3	14297	12562	13129	16179	11421	9980	0.10406715	1.144389	30775.17	-3389.379	0.04757537	-0.6212572	39.04152	0.7779803	0.0006280385	6208.1880	2133.5886	0.20090488
BURL08	2830	0.74	4.1	12680	12721	13723	15601	11549	10581	0.06404310	1.091485	30464.28	-3783.023	0.04829007	1.6531690	20.66904	0.1694955	0.0025317040	8779.7549	2047.2665	0.19359086
BURL10	7250	1.20	15.6	8801	9190	9527	13164	12996	10984	0.16028382	1.183176	25764.80	-1246.614	-0.17154963	2.1049891	10.45815	0.4945253	0.0006280385	5196.5854	2093.3467	0.35535058
BURL12	7400	0.41	5.8	8874	9226	9429	12611	12494	10567	0.14437386	1.182360	25107.93	-1594.797	-0.15046041	3.0235653	20.01288	0.2745165	0.0028951145	5577.9028	1972.9420	0.17606582
CAMD01	8670	1.70	8.4	9693	10036	10576	13478	12500	10465	0.12064522	1.194458	26594.11	-2006.063	-0.10933618	1.7120944	11.81088	0.7786027	0.0000000000	8868.8955	4027.1577	0.49622056
CAMD02	3360	0.00	1.6	8652	9020	9547	13232	14335	11690	0.16177180	1.226262	26532.00	-1141.431	-0.22757439	1.6377778	149.27640	0.6220903	0.0009420578	26354.1426	4625.2568	0.14727801
CAMD03	3520	0.89	8.4	8628	9062	9650	14177	13794	11490	0.18999454	1.200522	26802.13	-518.658	-0.20703535	1.5360342	22.96560	0.8972076	0.0006280385	11406.5117	5299.3960	0.26212120
CAMD06	1970	0.39	7.4	8793	9362	10049	14277	14302	11670	0.17380580	1.225536	27472.28	-776.438	-0.20875591	1.2255783	47.57160	1.7558347	0.0012947345	16406.4902	2584.3567	0.24268474
CAMD07	8140	1.20	13.6	17155	18389	19047	20911	9783	8888	0.04664898	1.100698	37679.71	-5348.738	0.30548063	0.7134655	12.46704	1.4638700	0.0011322125	10722.4297	3811.8894	0.39783421
CAMD08	6720	0.00	5.8	9498	9526	10096	12839	11504	10020	0.11959887	1.148104	25218.67	-2028.906	-0.09405611	3.1441998	80.70769	0.3901238	0.0009930161	19525.0215	3060.5881	0.19948815
CAMD11	3800	1.60	6.3	9676	10576	11069	16601	15465	12723	0.19992772	1.215515	30650.15	-278.970	-0.18774241	1.3239517	58.07016	0.9830751	0.0003140192	6994.1191	1842.0098	0.62953264

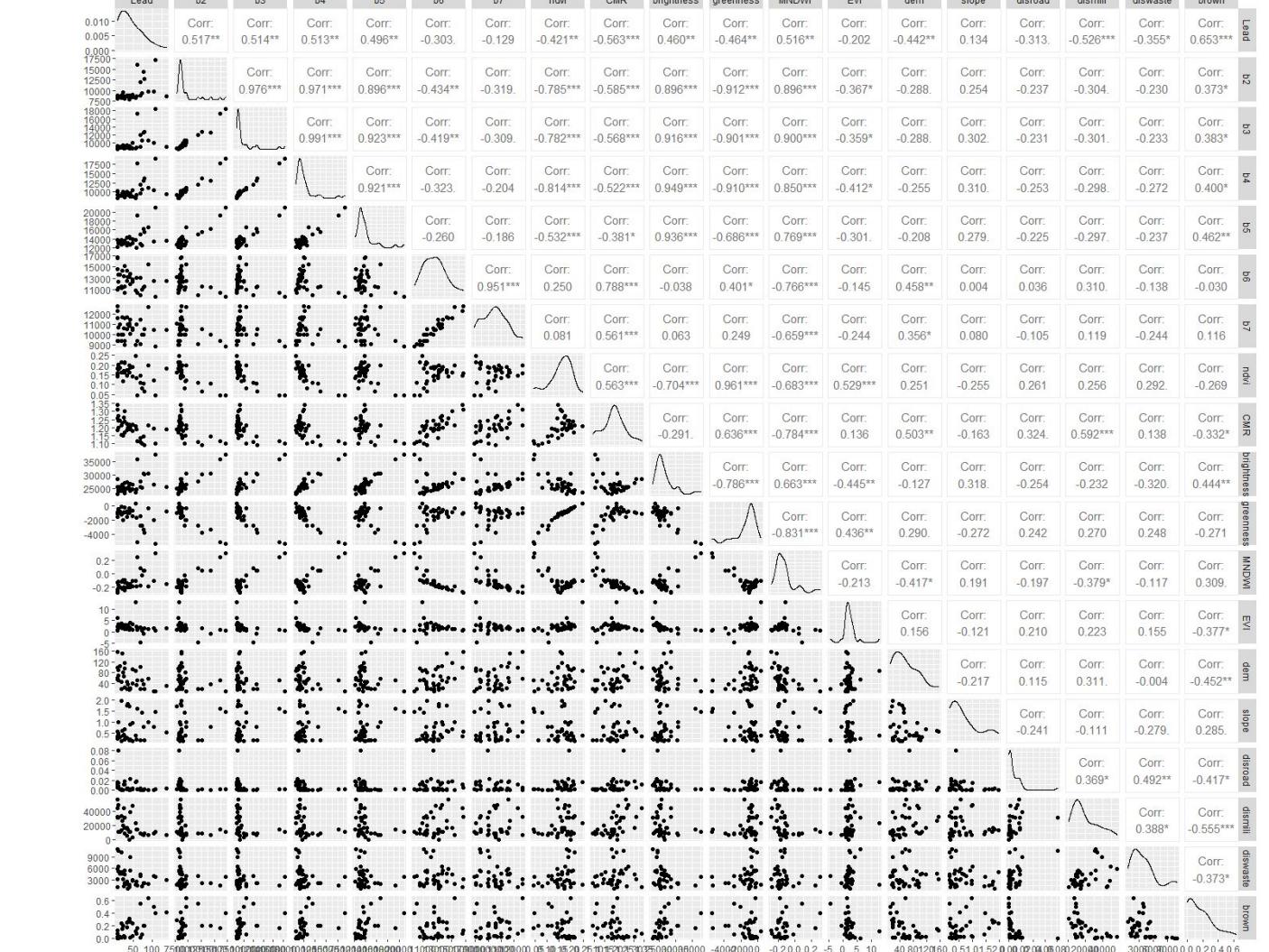
Model

Linear Regression

Lead

	b2	b3	b4	b5	b6	b7	ndvi	CMR	brightness	greenness	MNDWI	EVI	dem	slope	disroad	dismili	diswaste	brown		
Lead		Corr: 0.517**	Corr: 0.514**	Corr: 0.513**	Corr: 0.496**	Corr: -0.303	Corr: -0.129	Corr: -0.421**	Corr: -0.563***	Corr: 0.460***	Corr: -0.464**	Corr: 0.516**	Corr: -0.202	Corr: -0.442**	Corr: 0.134	Corr: -0.313	Corr: -0.526***	Corr: -0.355*	Corr: 0.653***	
Brownfield Density	***		Corr: 0.976***	Corr: 0.971***	Corr: 0.896***	Corr: -0.434**	Corr: -0.319	Corr: -0.785***	Corr: -0.585***	Corr: 0.896***	Corr: -0.912***	Corr: 0.896***	Corr: -0.367*	Corr: -0.288	Corr: 0.254	Corr: -0.237	Corr: -0.304	Corr: -0.230	Corr: 0.373*	
CMR ***		***		Corr: 0.991***	Corr: 0.923***	Corr: -0.419**	Corr: -0.309	Corr: -0.782***	Corr: -0.568***	Corr: 0.916***	Corr: -0.901***	Corr: 0.900***	Corr: -0.359*	Corr: -0.288	Corr: 0.302	Corr: -0.231	Corr: -0.301	Corr: -0.233	Corr: 0.383*	
Band 2 **			***		Corr: 0.921***	Corr: -0.323	Corr: -0.204	Corr: -0.814***	Corr: -0.522**	Corr: 0.949***	Corr: -0.910***	Corr: 0.850***	Corr: -0.412*	Corr: -0.255	Corr: 0.310	Corr: -0.253	Corr: -0.298	Corr: -0.272	Corr: 0.400*	
MNDWI **				***		Corr: -0.260	Corr: -0.186	Corr: -0.532***	Corr: -0.381*	Corr: 0.936***	Corr: -0.688***	Corr: 0.769***	Corr: -0.301	Corr: -0.208	Corr: 0.279	Corr: -0.225	Corr: -0.297	Corr: -0.237	Corr: 0.462**	
Band 3 **					***		Corr: 0.951***	Corr: 0.250	Corr: 0.789***	Corr: -0.038	Corr: 0.401*	Corr: -0.768***	Corr: -0.145	Corr: 0.458**	Corr: 0.004	Corr: 0.036	Corr: 0.310	Corr: -0.138	Corr: -0.030	
Band 4 **						***		Corr: 0.081	Corr: 0.561***	Corr: 0.063	Corr: 0.249	Corr: -0.659***	Corr: -0.244	Corr: 0.356*	Corr: 0.080	Corr: -0.105	Corr: 0.119	Corr: -0.244	Corr: 0.116	
Band 5 **							***	Corr: 0.563***	Corr: -0.704***	Corr: 0.961***	Corr: -0.683***	Corr: 0.529***	Corr: 0.251	Corr: 0.261	Corr: 0.256	Corr: 0.292	Corr: 0.269	Corr: 0.292	Corr: -0.329*	
Greenness **								***	Corr: -0.291	Corr: 0.636***	Corr: -0.784***	Corr: 0.136	Corr: 0.503**	Corr: -0.163	Corr: 0.324	Corr: 0.592**	Corr: 0.138	Corr: -0.332*	Corr: 0.444**	Corr: -0.271

Highly Correlated Variables



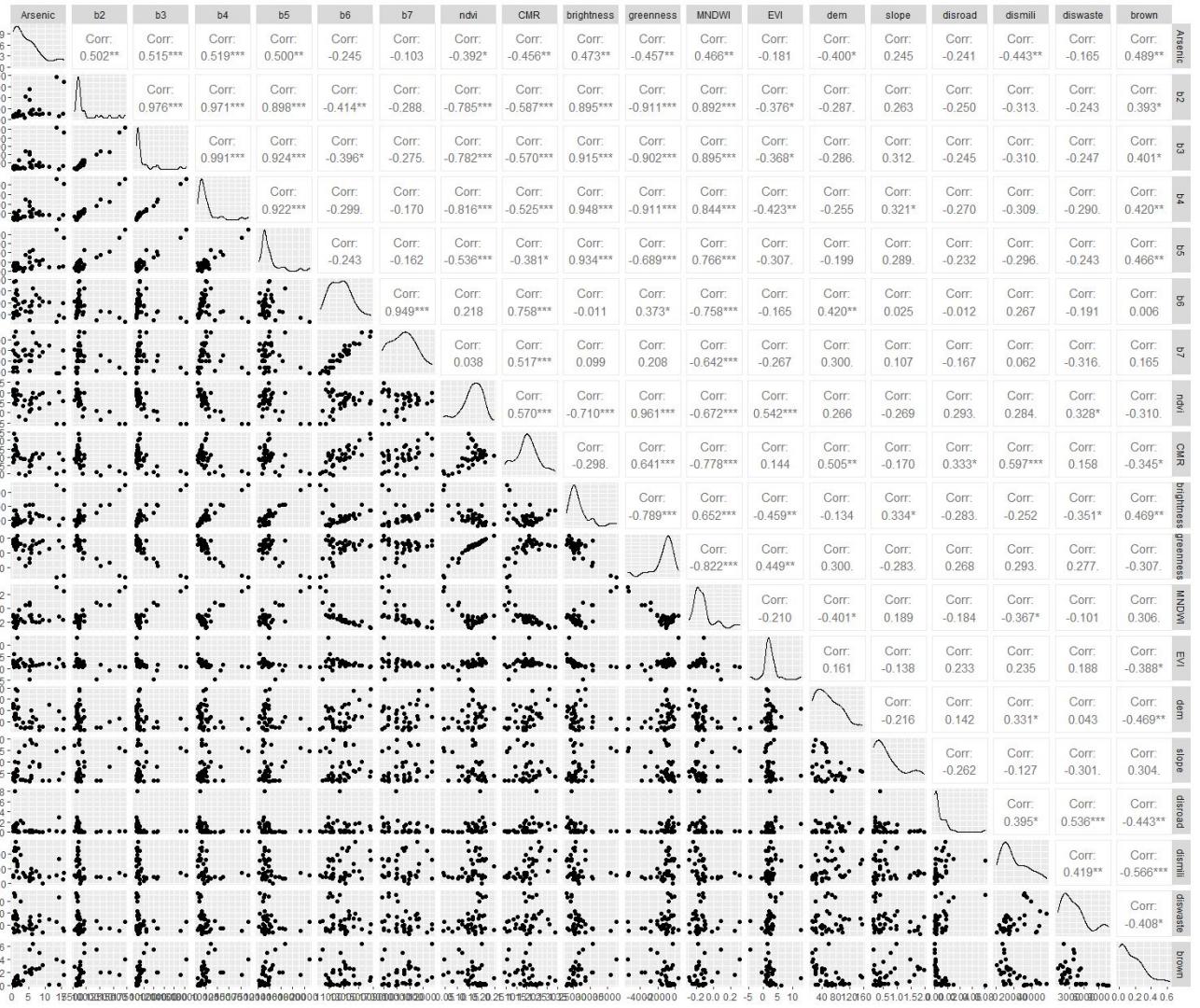
Model

Linear Regression

Arsenic

	b2	b3	b4	b5	b6	b7	ndvi	CMR	brightness	greenness	MNDWI	EVI	dem	slope	disroad	dismili	diswaste	brown	Arsenic	
Band 4 ***			Corr: 0.502**	Corr: 0.515***	Corr: 0.519***	Corr: 0.500**	Corr: -0.245	Corr: -0.103	Corr: -0.392*	Corr: -0.456**	Corr: 0.473***	Corr: -0.457**	Corr: 0.466**	Corr: -0.400*	Corr: 0.245	Corr: -0.241	Corr: -0.443**	Corr: -0.165	Corr: 0.489**	
Band 3 ***				Corr: 0.976***	Corr: 0.971***	Corr: 0.898***	Corr: -0.414**	Corr: -0.288	Corr: -0.785***	Corr: -0.587***	Corr: 0.895***	Corr: -0.911***	Corr: 0.892***	Corr: -0.376*	Corr: -0.287	Corr: 0.263	Corr: -0.250	Corr: -0.313	Corr: -0.243	
Band 2 **					Corr: 0.991***	Corr: 0.924***	Corr: -0.396*	Corr: -0.275	Corr: -0.782***	Corr: -0.570***	Corr: 0.915***	Corr: -0.902***	Corr: 0.895***	Corr: -0.368*	Corr: -0.286	Corr: 0.312	Corr: -0.245	Corr: -0.310	Corr: -0.247	
Band 5 **						Corr: 0.922***	Corr: -0.299	Corr: -0.170	Corr: -0.816***	Corr: -0.525***	Corr: 0.948***	Corr: -0.911***	Corr: 0.844***	Corr: -0.423**	Corr: -0.255	Corr: 0.321*	Corr: -0.270	Corr: -0.309	Corr: -0.290	
Brownfield Density **							Corr: -0.243	Corr: -0.162	Corr: -0.536***	Corr: -0.381*	Corr: 0.934***	Corr: -0.689***	Corr: 0.768***	Corr: -0.307	Corr: -0.199	Corr: 0.289	Corr: -0.232	Corr: -0.296	Corr: -0.243	
Brightness **								Corr: 0.949***	Corr: 0.218	Corr: 0.758***	Corr: -0.011	Corr: 0.373*	Corr: -0.758***	Corr: -0.165	Corr: 0.420**	Corr: 0.025	Corr: -0.012	Corr: 0.267	Corr: -0.191	
MNDWI **									Corr: 0.038	Corr: 0.517***	Corr: 0.099	Corr: 0.208	Corr: -0.642***	Corr: -0.267	Corr: 0.300	Corr: 0.107	Corr: -0.167	Corr: 0.062	Corr: -0.316	
Greenness **										Corr: 0.570***	Corr: -0.710***	Corr: 0.961***	Corr: -0.672***	Corr: 0.542***	Corr: 0.266	Corr: -0.269	Corr: 0.293	Corr: 0.284	Corr: 0.328*	Corr: -0.310
											Corr: -0.298	Corr: 0.641***	Corr: -0.778***	Corr: 0.144	Corr: 0.505**	Corr: -0.170	Corr: 0.333*	Corr: 0.597***	Corr: 0.158	Corr: -0.345*
												Corr: -0.789***	Corr: 0.652***	Corr: -0.459**	Corr: -0.134	Corr: 0.334*	Corr: -0.283	Corr: 0.252	Corr: -0.351*	Corr: 0.469**
													Corr: -0.822***	Corr: 0.449**	Corr: 0.300	Corr: -0.283	Corr: 0.268	Corr: 0.293	Corr: 0.277	Corr: -0.307
														Corr: -0.210	Corr: -0.401*	Corr: 0.189	Corr: -0.184	Corr: -0.367*	Corr: -0.101	Corr: 0.306
															Corr: 0.161	Corr: -0.138	Corr: 0.233	Corr: 0.235	Corr: 0.188	
																Corr: 0.142	Corr: 0.331*	Corr: 0.043	Corr: -0.469**	
																	Corr: -0.262	Corr: -0.127	Corr: -0.301	
																		Corr: 0.304		

Highly Correlated Variables



Model Accuracy Analysis

Arsenic ~ b4 + b3 + b2 + b5 + brownfield + brightness +
MNDWI + greenness

```
lm(formula = Arsenic ~ b4 + b3 + b2 + b5 + brown + brightness +  
    MNDWI + greenness, data = train)
```

Residuals:

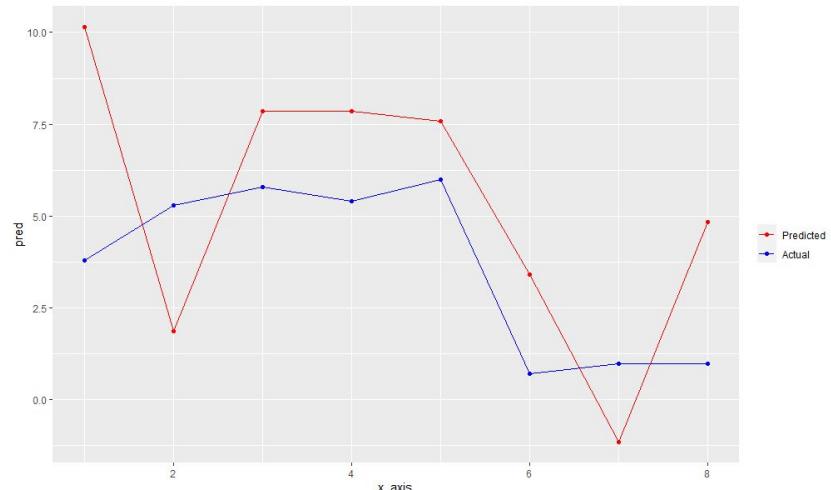
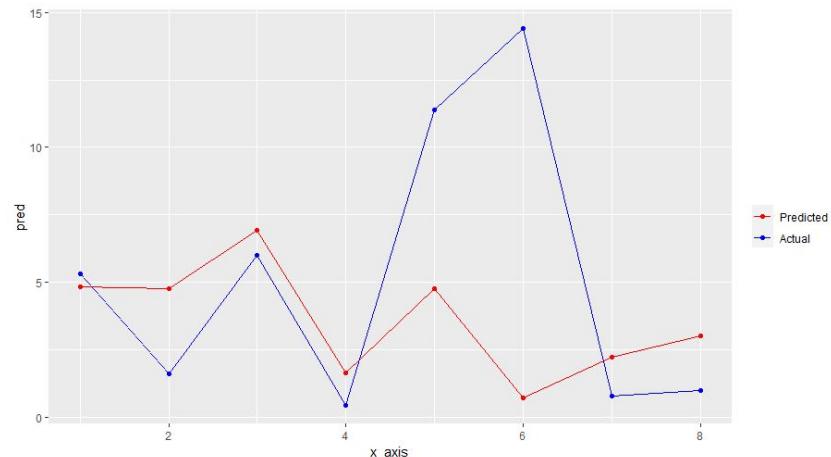
Min	1Q	Median	3Q	Max
-3.8640	-1.3859	-0.0267	0.6565	7.5451

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.804e+01	1.270e+01	-1.420	0.17017
b4	-7.060e-03	8.898e-03	-0.793	0.43643
b3	3.349e-03	4.530e-03	0.739	0.46791
b2	-3.375e-03	4.360e-03	-0.774	0.44761
b5	2.471e-02	1.230e-02	2.009	0.05751
brown	1.453e+01	4.173e+00	3.482	0.00222 **
brightness	-1.165e-02	3.382e-03	-3.446	0.00242 **
MNDWI	-1.767e+02	5.181e+01	-3.411	0.00263 **
greenness	-2.551e-02	1.528e-02	-1.669	0.10986

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.669 on 21 degrees of freedom
Multiple R-squared: 0.7402, Adjusted R-squared: 0.6413
F-statistic: 7.48 on 8 and 21 DF, p-value: 0.0001026

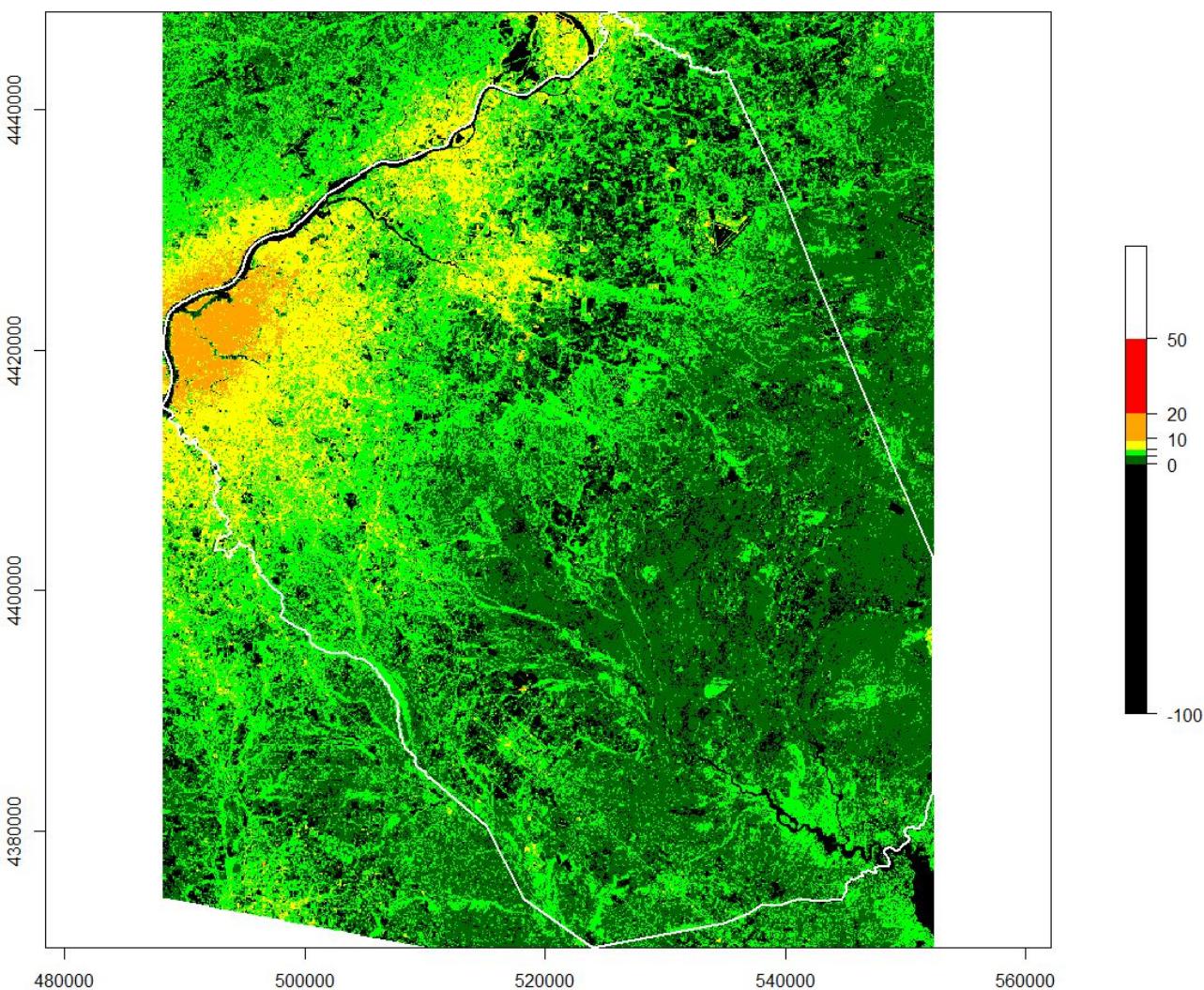


Arsenic Distribution

In 2002

Arsenic Source:

- Industrial buildings such as glass factories
- Smelters
- Farm fields or orchards where arsenic pesticides were used

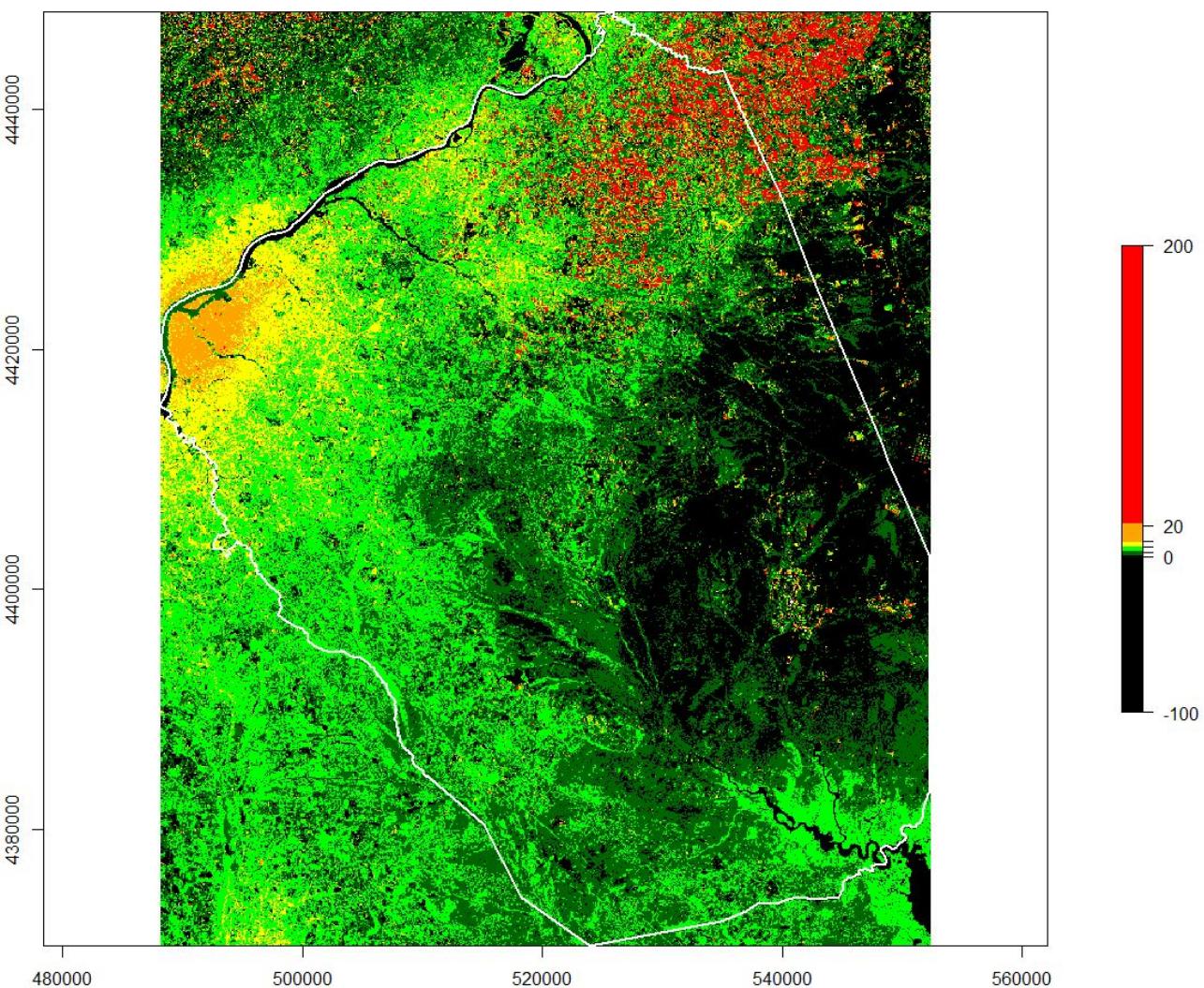


Arsenic Distribution

In 2016

Arsenic Source:

- Industrial buildings such as glass factories
- Smelters
- Farm fields or orchards where arsenic pesticides were used



Latency

- How to increase model accuracy?
- Try other models
(M5 model tree, GA-BP model)
- Visualize metal spatial distribution in 2022
- Use pervious surface raster as a mask