GIS and Heavy Metal Soil Contamination
Steven Kohm

GIS has been used extensively in mapping heavy metal contaminations over large areas or urban and suburban land. Heavy metals are typically defined in environmental engineering applications as elements with atomic numbers greater than Iron and with densities greater than 5 g/cm³ (LaGrega et al. 2001). These metals have been known to affect the central nervous system and even have been linked to specific types of cancer (Cournish et al. 2007). Metals in the environment tend to bioaccumulate due to their low solubility. Bioaccumulation is how metals slowly work their way into the food chain at low concentrations in tiny organisms and increase in lipid tissue of larger organisms as they consume the smaller ones (LaGrega et al. 2001). Under the correct circumstances metals can even bioaccumulate into humans. Luckily, scientists are using GIS to help map these contaminants and mitigate their affects through remediation techniques.

Most studies seem to follow a similar procedure when mapping heavy metals. First soil samples are taken and located through GPS technology. Next the samples are tested for concentrations of different heavy metals typically with spectrography. After all tests are complete statistical analyses are performed in order to normalize and group the data. This is an important step especially if the study is aimed at several different types of metals. Finally, the sample point information is overlaid on a map of the study area and interpolation is used to find the areas of highest concentration. These areas of hot spots are then compared with existing data sets to indicate possible anthropogenic or natural sources.

Collecting soil samples is the first step of the mapping operation. The researcher must choose a finite area where samples are to be collected in order to support his/her specific study. These areas could be political boundaries for example Zheng et al. looked at metal contamination in Galway, Ireland. Study areas can also be more specific, such as Davis’ et al. study of metal contamination in areas of increases mental retardation for South Carolinian children. After the study area is clearly defined the researcher will look at what density of samples can be used to represent their area the most accurately while staying within their budget. Most describe this density in terms of samples per km². Each sample’s exact location is then recorded with GPS technology. Zhou’s et al. study of metal concentrations in Hong Kong Marine sediment specifically mentions differential GPS being used in order to find precise sampling locations while out on the water. The samples concentrations are then tested for metal concentrations using some sort of spectrography method usually, Inductively Coupled Plasma (ICP) spectrography. Depending on what the researcher is specifically looking for, testing may range from one metal to as many as twelve. Common metals of concern include Lead, Chromium, Arsenic, Copper, and Zinc. At this point in the study exact location of each individual sample along with concentration of heavy metals are known.
As you can imagine a dataset like Romic’s et al. with 784 samples focusing on the concentrations of six different metals can get quite complex. For this reason multivariate statistics must be used in order to normalize and group data before entering it in GIS software. This particular branch of statistics looks at interrelationships among many variables. The first method used is Principal Component Analysis, or PCA. PCA helps to condense the number of correlated values in a dataset to a less significant number of uncorrelated values (Jolliffe 2002). The main benefit of this method is to help reduce the variance of the sample data. For example, PCA helped to define four principal components reducing 75% of the variance in Li’s et al. study of metal contamination in urban Hong Kong. The most statistically significant component was identified as Chromium, Cadmium, and Cobalt’s relationship to Iron and Aluminum signifying a natural input from parent rocks. The second multivariate method is known as Cluster Analysis. This method divides data into subsets based on similar statistical qualities (Jolliffe 2002). Metals placed in a subset, or cluster, might be from a similar source. The results of a cluster analysis are visually displayed in a dendrogram allowing the reader to see the relationships between clusters. A statistical analysis including these two methods is imperative when dealing with such large sets of sample data.

Once a statistical analysis is complete the soil sample point data must be overlaid on a map of the study area. The map of the study area might be created through the clip function in GIS software. In order to create a map of metal concentrations for a point on polygon dataset spatial interpolation is used. Inverse distance weighting (IDW) interpolation is one method frequently used. Zheng et al. uses IDW in his study of Galway, Ireland because of its relative simplicity. Zheng explains that there is error in any method of spatial interpolation and his purpose is to find an overall pollution pattern for which IDW is sufficient. Other studies show many different variations of kriging methods. Liu uses methods of disjunctive kriging as well as cokriging in his study of metals in Chinese paddy fields. Lognormal kriging, a disjunctive method, interpolates positive data through logarithmic functions. Cokriging allows for interpolation by means of two variables. Liu asserts the main benefit of cokriging is reducing the number of samples that need to be collected. Through comparison of kriging error values with the cokriging equivalents using progressively fewer sampling locations sampling density can be optimized for specific metals. Error is measured as both root mean square error and $R^2$ values. Semivariograms are a common figure for the studies that used kriging as an interpolation method. However, if samples were obtained with a uniform density the error map created should show similar statistical strength in all points.

After a map of metal concentrations is successfully created and interpolated the next step is to link the areas of the highest metal concentrations with a source. This step is where GIS technology becomes extremely useful. Sources are determined by layering pre-existing datasets of land use, road density, and soil characteristics over the metal concentrations maps then finding similarities. Some basic trends are apparent after looking through several comparable studies. The highest levels of metal contaminations
with respect to anthropogenic sources occur in areas with close proximity to roads and industrial sites. This is most likely the case because gasoline contained lead until only recently. Also in the past industries were not particularly careful with their environmental discharges. Natural sources of metal in the environment typically stem for parent rock material.

Previous studies of metal contamination follow the same this basic trend; sampling, mapping, interpolating, overlaying, analyzing. However, a study currently in press by Rania Bou Kheir et al. uses a novel technique that has not been previously explored. His study focused on Zinc contamination in northern Lebanon. Bou Kheir bypasses the need for soil sampling by using a decision tree model to predict areas of likely metal contamination. The decision tree uses a series of boolean type equations to predict each raster squares metal concentration. Factors for the equations are contained in previously created data sets and include; pH, proximity to roads, proximity to waste sites, etc. The equations are weighted by the strength of correlation between a specific condition and likelihood that the area will contain metals. Bou Kheir compared his decision tree results to soils sample data and found an accuracy of 78%. Depending on the accuracy required for a particular study this decision tree method is a way of saving time and money for researchers.

GIS is an extremely valuable tool when mapping metal contamination over large areas. Information from contaminant maps created with GIS could help to show environmental engineers the areas in the most need of remediation. Furthermore, GIS might be used to map other persistent pollutants in order to expedite their remediation before bioaccumulation becomes a problem. Quickly creating reasonably accurate maps through decision tree modeling is the future for this application of GIS. This idea is so appealing because interpolating between point samples is by no means an exact science and there is a significant chance for error. These contaminant maps should merely serve as a guide where contamination levels are likely to be elevated and special consideration needs to be applied before further developing that land. The decision tree model can save time and resources while still serving as an acceptable guide for land developers and environmental engineers.

**Annotated Bibliography**


This study focuses on Zinc contamination in areas of northern Lebanon. Samples were taken at a density of 1 sample per km² and located with GPS. This is such a novel study because it uses a decision tree model to help predict Zinc concentration. The decision tree factors in order of most importance include; pH, proximity of waste areas, proximity to roads, proximity to cities, distance to drainage line, lithology, land
cover/use, slope and gradient, conductivity, soil type, organic matter, and soil depth. These factors are all obtained from existing GIS databases and are used as clues to which areas might be of the highest concentration of Zinc. GIS software then creates a map of Zinc hot spots according to decision tree factors. These findings were compared to the GIS map of samples and an overall accuracy of 78% percent was found. The beauty of the decision tree model is it can be performed quickly, inexpensively, and simply. This study uses GIS datasets to get around the more time consuming methods of collecting samples and using a spectrograph. Accuracy may suffer slightly, however, depending on your application this might be a cost effective method.


Davis et al. uses GIS as a means of studying metal contaminated soils and their sources in areas of increased mental retardation or developmental delay in South Carolinian children. Around 600 soil samples were taken from 5 different rural or urban areas, 910 km² in total, and analyzed for concentrations of 9 specific metals know to be neurotoxins. The 5 areas were areas in which elevated levels of mental problems in children were observed. A point on polygon overlay was then used to incorporate the sample data with existing data sets including soil type (NRCS), ecological regions (USGS), land use (EPA), and climate information (NCDC) in order to determine information about the area. Once the point samples were incorporated Kriging analysis was used to find metal concentration hot spots and then determine the source. Results found that Ba, Be, and Mn were consistently from natural sources. Pb and Hg were results of anthropogenic sources. Furthermore, As, Cr, Cu, and Ni were both sources. This study might help to point out environmentally irresponsible companies and allow the EPA to enforce a proper remediation of the contaminated sites.


The Piedmonte region of Italy was studied for soil contamination of 6 heavy metals. This study aims to determine metal concentrations, learn about their variability, determine if sources are human or geologically based, and learn of any non-point factors that caused the pollution. 50 sample locations were used to gather data. At each of the 50 locations two samples were taken; the first of the plough horizon and the second from horizon below. Multivariate statistics, including cluster analysis and principal component analysis were used to normalize and group data. Several different GIS data bases were used to determine land use as well as geographic and geologic information about the sampling locations. Cr, Co, and Ni were associated with parent rock material through geologic data bases. Cu and Zn are associated with grape growing operations in vineyard regions of northern Italy. Finally, lead was associated with road density and locations of industrial activities. This study differs from others because it
looks at samples in soil horizons deeper than just the top soil. The deeper soil layers seem to show differences from the top soil layer with the metals Cu and Zn. This was attributed to the different migration rates of the metals. Northern Italy is a region known for superior wine, by determining the locations of high concentrations of metals agriculture can be protected from bioaccumulation of metals in the future.


Soil samples, 5 per km², were taken in the extremely urban Kowloon district of Hong Kong in order to test for concentrations of Cd, Cr, Cu, Ni, Pb and Zn. The statistical analyses of principal component analysis and cluster analysis were used to normalize and group the data, respectively. Finally the point data was overlaid onto a GIS map and compared with several different datasets. These datasets included road data, land use, and geologic information. The highest concentrations of metals tended to be at large road intersections or near industrial sites. Pb was one of the contaminants of the most concern in the study and its origins were determined to be anthropogenic, again, vehicular emissions or poorly managed industrial processes. This study shows us that GIS is an effective way to map contaminants and determine their sources through overlay techniques.


450 different soils samples were taken and absorption spectrometry was used to determine concentrations of metals for HJH plain paddy fields in the Zhejiang province, China. This is such an important region to study the metal concentrations because they could bioaccumulate in the rice and humans could be exposed. Several different methods of spatial interpolation were utilized including kriging, cokriging, and lognormal kriging. The cokriging correlates the concentrations of Cu, Zn and Cr to cation exchange capacity and is able to find a positive relationship. This method of cokriging can help to reduce the number of samples researchers need to take in order to save time and money. This study shows semivariograms and thoroughly takes the reader through the decision making process for each method of spatial interpolation. R² values, model type (linear or exponential), and root mean square error is described for each metal. The results show metal hot spots in the study area and might serve as map for prioritizing remediation techniques.


This study combines the results of 784 soil samples in central Croatia with a soil pollution ranking system called continuous limitation scores. The continuous limitation
score ranges between 0, clean soil, and 4, heavily contaminated soils. The samples points were interpolated with a block-regression kriging method, which sets a threshold for metal concentration. Error analysis of the kriging method included root mean square, mean prediction error, and semivariograms. Metals in question for this study are Cd, Cr, Cu, Ni, Pb and Zn. Several pre-existing GIS databases were used to find relationships between metal concentrations and factors such as ground water depth, land cover, industrialization parameters, and geology. This study uses figures to compare the locations of high metal concentration with Boolean algebra to find suitable agricultural areas in central Croatia. Also a slope map of distances to urban areas helps to show the possible anthropogenic causes of the soil contamination. These results are extremely valuable because they clearly define the areas that are the most fertile for agriculture.


Most people associate Ireland with sprawling green landscapes and majestic limestone cliffs. To assess the recent population boom’s affect on the environment in Galway, Ireland Zheng tested 166 soil samples with a sample density of one sample per 0.25 km². An Inductively Coupled Plasma Atomic Emission Spectrograph test (ICP-AES) allowed for the determination of metal concentration in the soil samples. The multivariate procedures of cluster analysis and principal component analysis helped to group and smooth all the concentration data so definitive trends could be recognized. Increased values of As, Cu, Pb, and Zn were found compared to typical northern European soils. Inverse distance weighting (IDW) was used to interpolate between point concentration values of these 4 elevated metals. When comparing these metal IDW hot spots to a data set of high road density it becomes clear that the metal contamination is caused from traffic pollution. Arsenic concentrations were not found to follow this road density trend and Zheng concludes that their presence is due to coal and peat combustion for heating purposes. This study uses GIS to correlate cause and effects of metal contamination through use of point on poly and line on poly overlays as well as IDW interpolation.


Feng, et al, collected marine sediment samples from 59 sites off the coast of Hong Kong twice annually from 1998-2004. The sediment samples were tested for concentrations of 12 metals in order to determine their spatial distribution and to identify anthropogenic sources environmental impact. Differential global position systems were used in order to monitor exact locations of sampling sites. The 59 sampling sites were separated into 3 areas based on cluster pollution level analysis. The inverse distance
weighting interpolation method and global polynomial interpolation was used to find levels of highest metal concentration, over regulatory upper chemical exceedance levels, and then determine probable sources in these areas. Results identified several anthropogenic sources including wastewater treatment plants, ship painting facilities, and electronic/chemical industrial sites. This article is interesting because they use more than one method of interpolation as well as several advanced statistical analyses in order to find the most accurate results.

**Additional Sources**
