International Journal of Agriculture, Environment & Biotechnology
Citation: IJAEB: 7(4): 787-795 December 2014
DOI Number: 10.5958/2230-732X.2014.01388.6

SOIL SCIENCE

Soil risk assessment of heavy metal contamination near Oil Refinery area, Northeastern India

S.K. Reza*, Utpal Baruah[1] and S.K. Singh[2]

National Bureau of Soil Survey and Land Use Planning, Sector-II, DK-Block, Salt Lake, Kolkata, West Bengal, India
[1]National Bureau of Soil Survey and Land Use Planning, Jamuguri Road, Jorhat, Assam, India
[2]National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra, India

*Corresponding author: reza_ssac@yahoo.co.in

Paper No. 269 Received: 15 September, 2014 Accepted: 18 October, 2014 Published: 20 December, 2014

Abstract

The present paper aims to maps Cr, Cd, Ni and Pb concentration and assess the hazard in the soils of surrounding agricultural fields affected by oil refinery drainage of Digboi refinery of Tinsukia district, Assam using statistics, geostatistics and GIS techniques. The amounts of Cr, Cd, Ni and Pb were determined from 97 samples collected within the contaminated area. Among the heavy metals studied, the mean concentration of Pb was high. The greatest and the smallest standard deviation were observed in the Ni (44.1) and pH (0.47), respectively. Analysis of the isotropic variogram indicated that the Cr and Cd semivariograms were well described with the circular model, with the distance of spatial dependence being 1240 and 1022 m, respectively, while the Pb and Ni were well describe with Gaussain model, with the distance of spatial dependence being 1930 and 2321 m, respectively. The ordinary kriging maps of Cr, Cd, Ni and Pb showed that high concentrations of heavy metals were located in the low lying area. Indicator kriged probability maps of soil Cr, Cd, Ni and Pb were prepared based on the concentrations to exceed the respective Food and Agriculture Organization maximum permissible limit (MPL) value of 100, 3, 30 and 50 mg kg^{-1}, respectively. It was seen that whole studied area had a higher than 0.99% probability to exceed the MPL value of Pb. About 10% area of the study site was having higher concentration than MPL value of Cd and Ni concentrated at the centre and north-west corner of the study area, respectively.

Highlights

- Apart from transport and municipal services, industrial plants constitute the main source of heavy metals released to environment.
- A good variogram structure of heavy metals was observed, showing that there are clear spatial patterns of heavy metals on the distribution map and also that the current sampling density is sufficient to indicate such spatial patterns.
- The kriging interpolated map showed areas with high values of heavy metal concentrations. The probability map produced based on indicator kriging provided useful information for hazard assessment.

Keywords: Heavy metals, geostatistics, spatial variability, accuracy assessment, risk assessment

Apart from transport and municipal services, industrial plants constitute the main source of heavy metals released to environment (Hjortenkrans *et al.,* 2006). A higher metal content in soils occurs most frequently within urbanized areas (Singh and Kumar, 2006) and around industrial facilities (Li *et al.,* 2006; Shukurov *et al.,* 2006). The petrochemical and refinery sector is counted among significant sources of environmental pollution, with an impact on the content of pollutants in the plants cultivated on these soils (Gogoi *et al.,* 2003; Hassan *et al.,* 2005; Jamrah *et al.,* 2007). During the last three decades, it has become more apparent that the total concentrations of heavy metals in soils and plants, their chemical forms, mobility and availability to the food chain provide the basis for understanding a range of problems in crop, animal and human health. Soils are the ultimate sink for trace elements in the terrestrial environment and have a great capacity for receiving, purifying and decomposing wastes and pollutants of different kinds (Boon and Soltanpour, 1992). The effects of industrial effluents on heavy metals contents soil and plant are well documented (Karaczun *et al.,* 2007; Owamah, 2013).

Risk assessment involves calculation of risk in affected areas and provides valuable information regarding feasible rehabilitation options. The methodology is mainly based on the principle “source – pathway – target”. Risk is better assessed if quantitative techniques are used to account for spatial and temporal variations. A probabilistic assessment takes into account variability of parameters and uncertainty in measurement (Korre *et al.,* 2002). Geostatistics is extensively used to assess the level of soil contamination and calculate the risk in contaminated sites, by preserving the spatial distribution and uncertainty of the estimates. In addition, geostatistics and GIS provide useful tools for the study of spatial uncertainty and hazard assessment (McGrath *et al.,* 2004; Komnitsas and Modis, 2006; Reza *et al.,* 2013).

In the northeastern India, Digboi refinery is the Asia’s oldest refinery was set up at Digboi in 1901 in upper Assam district of Tinsukia. Digboi is also known as the birth place of Indian oil industry. The refinery had an installed capacity 0.50 MMTPA (million metric tonnes per annum). Our study investigate the extent of contamination of heavy metals (Cr, Cd, Ni and Pb) in soil by Digboi refinery using statistics, geostatistics and GIS techniques in order to reveal the spatial distribution patterns and provide a basis for hazard assessment.

Materials and Methods

Study area

The study was carried out near the Digboi refinery area of Tinsukia district, Assam, north-eastern India, extended between 27°21′048″ to 27°24′50″ N latitude and 95°31′27″ to 95°36′42″ E longitude covering an area of 2378 ha (Figure 1). The climate is humid subtropical. The average annual rainfall ranges between 2100 and 2900 mm with maximum rainfall during July–September. The climate is moderately warm during summer but cold in winter. Mean monthly minimum and maximum temperatures were 7 °C and 36 °C, respectively.

Soil sampling and analysis

A total of 97 surface soil samples were collected from a depth of 0-25 cm (plough layer) using a square 500×500 m grid (Figure 1) covering not only the waste disposal site, but also the surrounding cultivated areas with the help of a hand-held global positioning system. Soil samples were air-dried and ground to pass through a 2 mm sieve. A combined glass calomel electrode was used to determine the pH of aqueous suspension (1:2.5 soil:solution ratio). Organic carbon was determined by the Walkley and Black (1934) method. Digestion of 0.50 g samples was performed with concentrated HNO_3, HF and $HClO_4$ in a microwave digester (model Start D, Milestone). Subsequently, the total concentration of heavy metals was determined by a Shimadzu AA6300 atomic absorption spectrophotometer.

Statistical analysis

The main statistical parameters, including mean, standard deviation, variance, coefficient of variance, and extreme maximum and minimum values, which are generally accepted as indicators of the central tendency and of the data spread, were analyzed. The Pearson correlation coefficients were estimated for all possible paired combinations of the response variables to generate a correlation coefficient matrix. These statistical parameters were calculated with EXCEL® 2007 and SPSS 15.0® (SPSS Inc., Chicago, III., USA).

Geostatistical analysis based on GIS

Spatial interpolation and GIS mapping techniques were employed to produce spatial distribution and risk assessment maps for the four observed heavy metals, and the software used for this purpose was ArcGIS v.9.3 (ESRI Co, Redlands, USA). The first step was taking the log-transformation of all non-normally distributed target variables (heavy metal contents) to ensure (in most cases) the normality of residuals. In ArcGIS, kriging can express the spatial variation and allow a variety of map outputs, and at the same time minimize the errors of predicted values. Moreover, it is very flexible and allows users to investigate graphs of spatial autocorrelation. In kriging, a semivariogram model was used to define the weights of the function (Webster and Oliver, 2001), and the semivariance is an autocorrelation statistic defined as follows (Mabit and Bernard, 2007):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where $z(x_i)$ is the value of the variable z at location of x_i, h the lag and $N(h)$ the number of pairs of sample points separated by h. For irregular sampling, it is rare for the distance between the sample pairs to be exactly equal to h. That is, h is often represented by a distance band.

Best-fit model with minimum root mean square error (RMSE) was selected for each heavy metal. Using the model semivariogram, basic spatial parameters such as nugget(C_0), sill $(C + C_0)$ and range (A) was calculated which provide information about the structure as well as the input parameters for the kriging interpolation (Dagar and Esfahan, 2013). Nugget is the variance at zero distance, sill is the lag distance between measurements at which one value for a variable does not influence neighboring values and range is the distance at which values of one variable become spatially independent of another (Lopez-Granados *et al.,* 2002; Reza *et al.,* 2010).

Indicator kriging

The probability maps of soil Cr, Cd, Ni and Pb concentration to exceed the respective FAO (2000) maximum permissible limit value (MPL) of 100, 3, 30 and 50 mg kg^{-1} were prepared using indicator kriging. Indicator kriging is a nonlinear geostatistics where the conventional linear kriging estimators are applied to the data after a nonlinear transformation. Here the nonlinear transform is to a discrete (binary) indicator variable. These techniques have been widely applied by soil scientists (Van Meirvenne and Goovaerts, 2001; Reza *et al.,* 2012; 2013).

Let us assume that a soil property z at location x take value $z(x)$. In geostatistics, we treat this value as a realization of the random function $Z(x)$. An indicator transformation of $z(x)$ can be defined by

$\omega_c(x) = 1$ if $z(x) \leq z_c$, 0 otherwise,

Where z_c is a threshold value of the property. In indicator geostatistics, $\omega_c(x)$ is regarded as a realization of the random $\Omega_c(x)$,

$\Omega_c(x) = 1$ if $z(x) \leq z_c$, else 0.

It can be seen that

$\text{Prob}[Z(x) \leq z_c] = E[\Omega c(x)] = G[Z(x); z_c]$,

Where Prob[ ], E[ ] denote, respectively, the probability and the expectation of the terms within the square brackets, and $G[Z(x); z_c]$ is the cumulative distribution function of $Z(x)$ at value z_c. The principal of IK is to estimate the conditional probability that $z(x)$ is smaller than or equal to a threshold value z_c, conditional on a set of observations of z at neighbouring sites, by kriging $\Omega_c(x)$ from a set of indicator-transformed data.

A set of data on z is transformed to the indicator variable $\omega_c(x)$. The variogram of the underlying random function $\Omega_c(x)$ is then estimated by

$$\gamma_{\Omega c}(h) = \frac{1}{2M_h} \sum_{i=1}^{M_h} [\omega c(x_i) - \omega c(x_i + h)]^2$$

Where M_h pairs of observations that are separated by the lag interval h. A set of estimates of this indicator variogram at different lags may then be modeled by one of the authorized continuous functions used to describe variograms (Webster and Oliver, 2001).

An estimate of the indicator random function may then be obtained for a location *x* by kriging from the neighbouring indicator-transformed data. IK is equivalent to simple kriging of the indicator variables $\omega_c(x)$ using the mean within the kriging neighbourhood as the expectation.

Accuracy assessment

Accuracy of the soil maps was evaluated through cross-validation approach (Davis, 1987; Reza *et al.*, 2010). Among three evaluation indices used in this study, mean absolute error (MAE), and mean squared error (MSE) measure the accuracy of prediction, whereas goodness of prediction (*G*) measures the effectiveness of prediction (Reza *et al.*, 2010). MAE is a measure of the sum of the residuals (*e.g.* predicted minus observed) (Voltz and Webster, 1990).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}[\![z(x_i) - \hat{z}(x_i)]\!]$$

Where $\hat{z}(x_i)$ is the predicted value at location i. Small MAE values indicate less error. The MAE measure, however, does not reveal the magnitude of errorthat might occur at any point and hence MSE will be calculated,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}[z(x_i) - \hat{z}(x_i)]^2$$

Squaring the difference at any point gives an indication of the magnitude, *e.g.* small MSE values indicate more accurate estimation, point-by-point. The *G* measure gives an indication of how effective a prediction might be relative to that which could have been derived from using the sample mean alone (Schloeder *et al.*, 2001).

$$G = \left[1 - \frac{\sum_{i=1}^{N}[z(x_i) - \hat{z}(x_i)]^2}{\sum_{i=1}^{N}[z(x_i) - \bar{z}]^2}\right] \times 100$$

Where *z* is the sample mean. If *G* = 100, it indicates perfect prediction, while negative values indicate that the predictions are less reliable than using sample mean as the predictors. The comparison of performance between interpolations was achieved by using mean absolute error (MAE).

Results and Discussion

Descriptive statistics of heavy metals and other soil properties

The statistical characteristics of soil Cr, Cd, Ni and Pb are listed in Table 1. The median of each heavy metal was lower than the mean, which indicates that the effects of abnormal data on sampling value were not great. In the present investigation, among the heavy metals studied (Cr, Cd, Ni and Pb), the mean

Table 1. Summary statistics of heavy metal concentrations and selected soil properties

	pH	Organic carbon (%)	Cr	Cd	Pb	Ni
			mg kg^{-1}			
Mean	4.7	3.44	74.10	1.68	87.84	45.20
Median	4.5	2.67	63.21	1.52	81.56	32.78
SD	0.47	1.82	13.88	2.00	22.96	44.10
CV (%)	10.0	52.9	18.7	119.0	26.1	97.6
Minimum	3.7	0.54	47.28	0.08	22.24	0.08
Maximum	5.9	7.01	106.48	8.12	127.44	293.92
Skewness	–0.02	0.14	0.11	2.12	–0.63	2.94
Kurtosis	–0.54	–1.35	–0.38	3.44	–0.11	13.53
Distribution pattern			Normal	Lognormal	Normal	Lognormal

SD, Standard deviation; CV, Co-efficient of variation

Table 2. Semivariogram model and parameters of heavy metals

Heavy metals	Fitted model	Nugget (C_0)	Sill ($C+C_0$)	Range (A) (m)	Nugget/Sill
Cr	Circular	0.411	1.981	1240	0.207
Cd	Circular	0.950	1.156	1022	0.821
Pb	Gaussian	0.175	0.737	1930	0.237
Ni	Gaussian	0.276	0.597	2321	0.462

concentration of Pb was high. A higher concentration of lead has also been found in soil impacted with petroleum exploration and production activities (Asia *et al.,* 2007; Owamah, 2013). The greatest and the smallest standard deviation were observed in the Ni (44.1) and pH (0.47), respectively. Organic carbon, Cd and Ni exhibit a high variation (>50%) according to guidelines provided by Warrick (1998). Skewness is the most common form of departure from normality. If a variable has positive skewness, the confidence limits on the variogram are wider than they would otherwise be and consequently, the variances are less reliable. A logarithmic transformation is considered where the coefficient of skewness is greater than one (Webster and Oliver, 2001). Therefore, a logarithmic transformation was performed for Cd and Ni because their skewness was greater than one.

Semivariogram analysis of heavy metals

Semivariogram analysis was used to characterize and quantify spatial variability and RMSE was used for different theoretical semivariogram models to fit the experimental semivariogram values for each micronutrient. Analysis of the isotropic variogram indicated that the Cr and Cd semivariograms were well described with the circular model, with the distance of spatial dependence being 1240 and 1022 m, respectively, while the Pb and Ni semivariogram was well described with the Gaussain model, with the distance of spatial dependence being 1930 and 2321 m, respectively (Table 2).

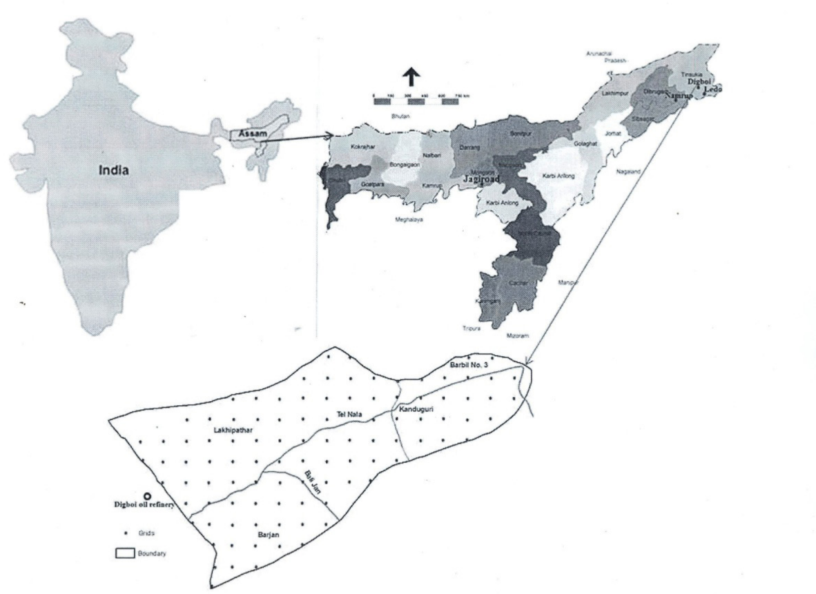

Figure 1. Location and grid map of the study area

In the semivariogarm analysis, the nugget values represent the variability of measured heavy metals level at zero distance, which are positive in this study for all the heavy metals. This spatial random variance is caused by the artificial nature of heavy metal pollution in soil; meaning that anthropogenic inputs are a significant source of heavy metals in the study area. The sill, sum of partial sill and nugget, is the maximum variance between data pairs and reflects the variations of regionalized variables in the study area. The ratio of nugget and sill is commonly used to express the spatial autocorrelation of regional variables, which also indicates the predominant factors among all natural and anthropogenic factors (Robertson *et al.,* 1997). The ratios of nugget and sill between 0.25 and 0.75 represented moderate spatial dependence; those below 0.25 represented strong spatial dependence; and all others represented weak dependence. Cr and Pb were strongly spatially dependent suggesting that they are affected by anthropogenic factors only while Ni was moderately spatially dependent suggesting that they are affected by either anthropogenic or natural factors or both.

Table 3. Evaluation performance of ordinary kriged map of heavy metals through cross-validation

Heavy metals	Mean absolute error (MAE)	Mean square error (MSE)	Goodness of prediction (G)
Cr	0.004	160.7	15.7
Cd	0.048	3.8	28.0
Pb	0.107	333.2	36.1
Ni	9.33	661.7	2.3

Spatial distribution and risk assessment of heavy metals pollution

Using the available measurements for Cr, Cd, Pb and Ni concentration as well as the aforementioned structural models, spatial maps of these pollutants were produced using the ordinary kriging procedure (Journel and Huijbregts, 1978). The spatial distribution maps of Cr, Cd, Pb and Ni (Figure 2a–d, respectively) showed that high concentration of heavy metals was located in the low lying paddy

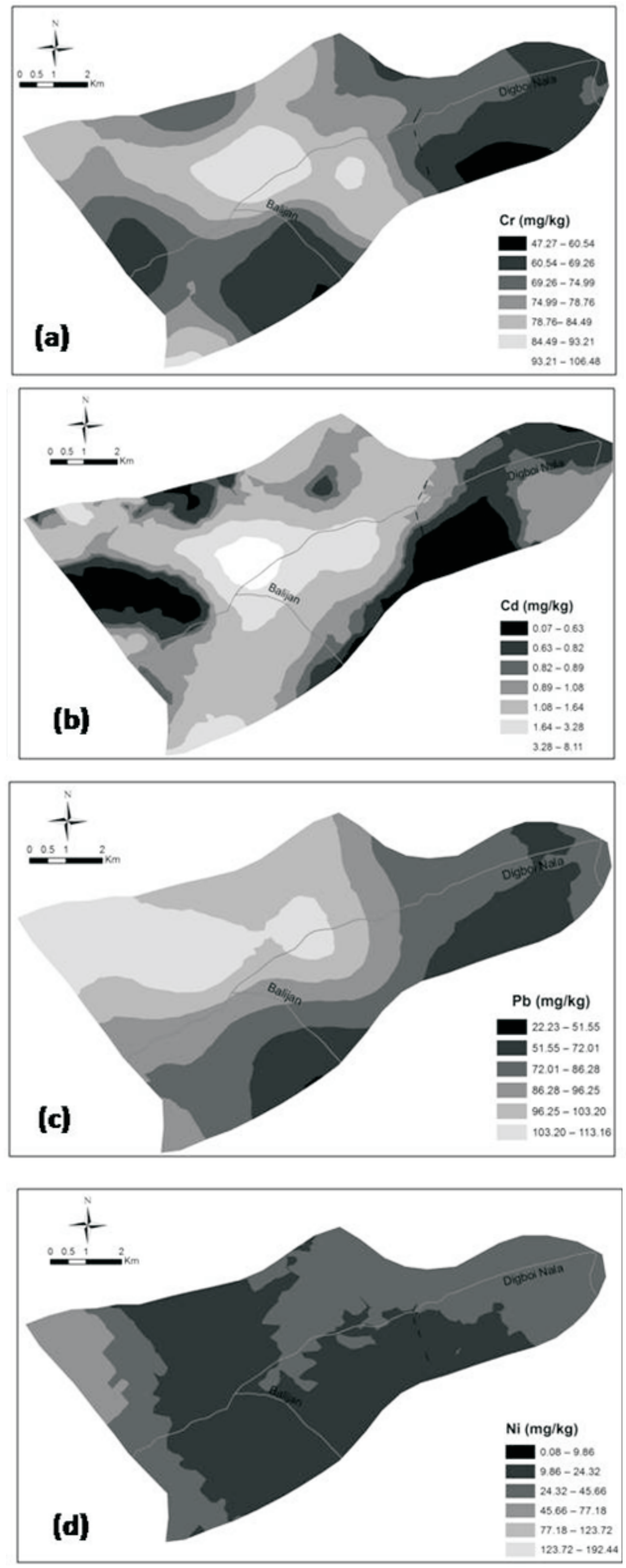

Figure 2 Spatial distribution maps of (a) Chromium; (b)

Cadmium; (c) Lead; (d) Nickel

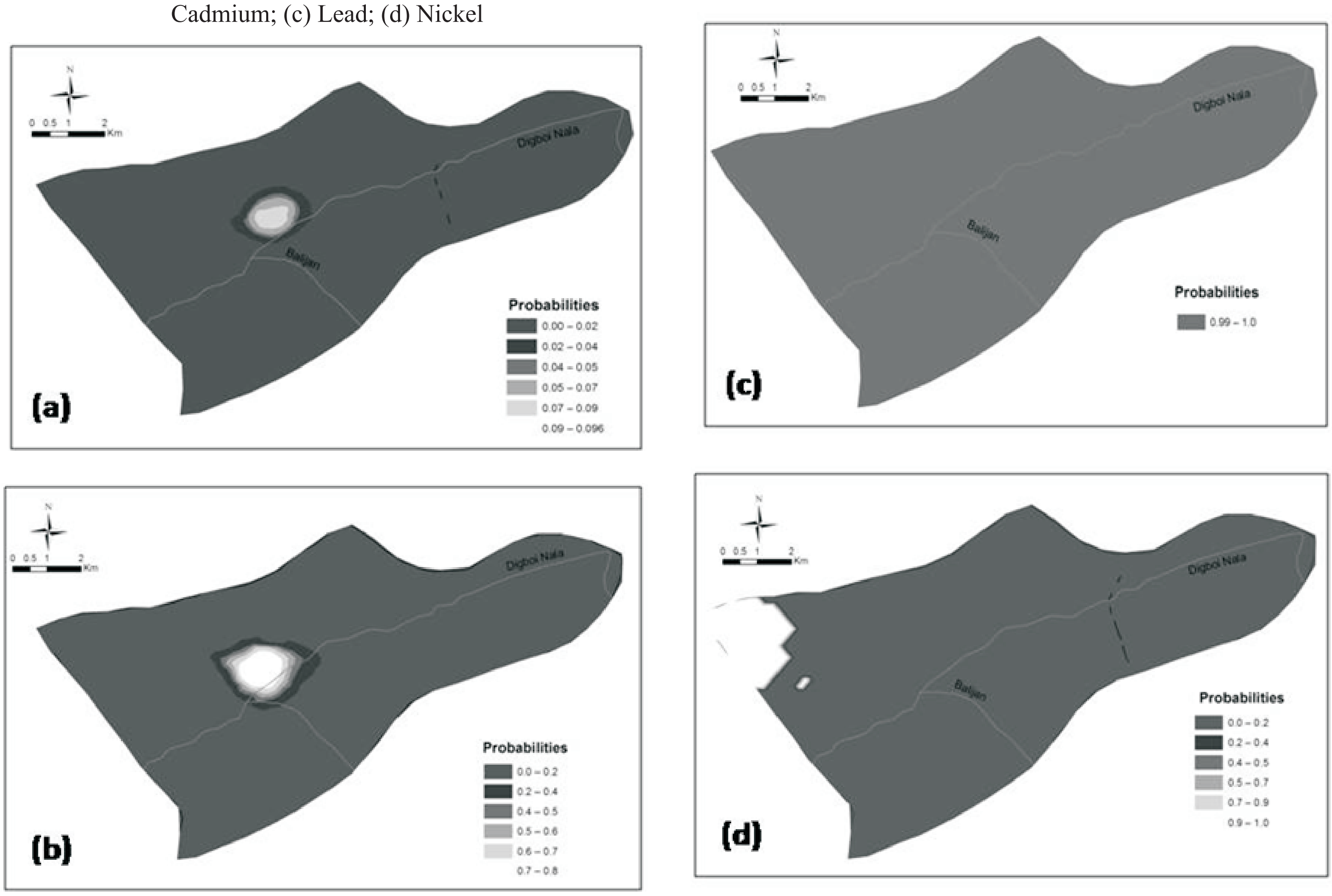

Figure 3. Risk assessment maps of (a) Chromium; (b) Cadmium; (c) Lead; (d) Nickel

field situated in the middle of the centre of the study area. Evaluation indices resulting from cross-validation of spatial maps of soil properties (Table 3) for all the soil heavy metals the prediction of goodness (G) value was greater than zero, which indicates that spatial prediction using semivariogram parameters is better than assuming mean of observed value as the values for any unsampled location. This also shows that semivariogram parameters obtained from fitting of experimental semivariogram values were fairly reasonable to describe the spatial variation.

In order to obtain data that may be used in the future for the assessment of the health risk due to elevated soil heavy metals concentration in cultivated areas, spatial maps of the probability that these pollutants exceed the corresponding maximum permissible limits (MPL) are produced. Figure (3a–d) shows the indicator kriged probability maps of soil Cr, Cd, Pb and Ni based on the concentrations to exceed the respective FAO (2000) MPL value of 100, 3, 50 and 30 mg kg^{-1}, respectively. It was seen that whole study area has higher than 0.99% probability to exceed this MPL value of Pb. About 10% area of the study site was having higher concentration than MPL value of Cd and Ni concentrated at the centre and north-west corner of the study area, respectively.

Conclusion

Geostatistics and statistics have been employed for assessment and mapping of soil pollution in the agricultural soils around the Digboi refinery area in the Tinsukia district of northeastern, India. A good variogram structure of heavy metals was observed, showing that there are clear spatial patterns of heavy metals on the distribution map and also that the current sampling density is sufficient enough to indicate such spatial patterns. The ordinary kriging interpolated map showed areas with high values of heavy metal concentrations. The probability map produced based on indicator kriging interpolation provided useful information for hazard assessment.

References

Asia, I.O., Jegede, S.I., Jegede, D.A., Ize-Iyamu, O.K. and Akpasubi, E.B. 2007. The effects of petroleum exploration and production operations on the heavy metals contents of soil and groundwater in the Niger Delta. *International Journal of Physical Sciences* **2:** 271–275.

Boon, D.Y. and Soltanpour, P.N. 1992. Lead, cadmium and zinc contamination of aspen garden soils and vegetation. *Journal of Environmental Quality* **21:** 82–86.

Dagar, M. and Esfahan, E.Z. 2013. Development of pedotransfer function for prediction of soil nitrogen and its spatial distribution in conservation of Damavand's rangelands. *International Journal of Bioscience* **3:** 74–81.

Davis, B.M. 1987. Uses and abuses of cross-validation in geostatistics. *Mathematical Geology* **19:** 241–248.

FAO.2000. Land and water digital media series. Soil and terrain database, land degradation status and soil vulnerability assessment for Central and Eastern Europe. Food and Agriculture Organization (FAO). Available: http://www. FAO.org/docrep/TO551E/TO551eob.htm.

Gogoi, B.K., Dutta, N.N., Goswami, P. and Mohan, T.R. 2003. A case study of bioremediation of petroleum-hydrocarbon contaminated soil at a crude oil spill site. A*dvances in Environmental Research* **7:** 767–782.

Hassan, H., Taha, R., Al-Rawas, A., Al-Shandoudi, B., Al-Gheithi, K. and Al-Barami, A. 2005. Potential uses of petroleum-contaminated soil in highway construction. *Construction and Building Materials* **19:** 646–652.

Hjortenkrans, D., Bergback, B. and Haggerud, A. 2006. New metal emission patterns in road traffic environments *Environmental Monitoring and Assessment* **117:** 85–98.

Jamrah, A., Al-Futaisi, A., Hassan, H. and Al-Oraimi, S. 2007. Petroleum contaminated soil in Oman: Evaluation of bioremediation treatment and potential for reuse in hot asphalt mix concrete. *Environmental Monitoring and Assessment* **124:** 331–341.

Journel, A.G. and Huijbregts, C.J. 1978. Mining Geostatistics. Academic Press, London.

Karaczun, Z.M., Indeka, L. and Obidoska, G. 2007. Heavy metal content in soil and consumption yield in the vicinity of petrochemical plant. *Land Reclamation* **38**: 19–24.

Komnitsas, K. and Modis, K. 2006. Soil risk assessment of As and Zn contamination in a coal mining region using geostatisretics. *Science of the Total Environment* **371:** 190–196.

Korre, A., Durucan, S. and Koutroumani, A. 2002. Quantitative-spatial assessment of the risks associated with high Pb loads in soils around Lavrio, Greece. *Applied Geochemistry* **17:**1029–1045.

Li J., Xie, Z.M., Xu J.M. and Sun Y.F. 2006. Risk assessment for safety of soils and vegetables around a lead/zinc mine. *Environmental Geochemistry and Health* **28:** 37–44.

Lopez-Granados, F., Jurado-Exposito, M., Atenciano, S., Gracia-Ferrer, A., De La Orden, M.S. and Gracia-Toreres, L. 2002. Spatial variability of agricultural soil parameters in Southern Spain. *Plant and Soil* **246:** 97–105.

Mabit, L. and Bernard, C. 2007. Assessment of spatial distribution of fallout radionuclides through geostatistics concept. *Journal of Environmental Radioactivity* **97:** 206–219.

McGrath, D., Zhang, C. and Carton, O.T. 2004. Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. *Environmental Pollution* **127:** 239–248.

Owamah, H.I. 2013. Heavy metals determination and assessment in a petroleum impacted river in the Niger delta region of Nigeria. *Petroleum and Environmental Biotechnology* **4:** 135.

Reza, S.K., Baruah, U. and Sarkar, D. 2012. Mapping risk of soil phosphorus deficiency using geostatistical approach: A case study of Brahmaputra plains, Assam, India. *Indian Journal of Soil Conservation* **40:** 65–69.

Reza, S.K., Baruah, U. and Sarkar, D. 2013. Hazard assessment of heavy metal contamination by the paper industry, north-eastern India. *International Journal of the Environmental Studies* **70:** 23–32.

Reza, S.K., Sarkar, D., Baruah, U. and Das, T.H. 2010. Evaluation and comparison of ordinary kriging and inverse distance weighting methods for prediction of spatial variability of some chemical parameters of Dhalai district, Tripura. *Agropedology* **20:** 38–48.

Robertson, G.P., Klingensmith, K.M., Klug, M.J., Paul, E.A., Crum, J.C. and Ellis, B.G. 1997. Soil resource, microbial activity, and primary production across an agricultural ecosystem. *Ecological Applications* **7:** 158–170.

Schloeder, C.A., Zimmermen, N.E. and Jacobs, M.J. 2001. Comparison of methods for interpolating soil properties using limited data. *Soil Science Society of America Journal* **65:** 470–479.

Shukurov, N., Pen-Mouratov, S. and Steinberger, Y. 2006. The impact of the Almalyk Industrial Complex on soil chemical and biological properties. *Environmental Pollution* **136**: 331–340.

Singh, S. and Kumar, M. 2006. Heavy metal load of soil, water and vegetables in periurban Delhi. *Environmental Monitoring and Assessment* **120:** 79–91.

Van Meirvenne, M. and Goovaerts, P. 2001. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma* **102**: 75–100.

Voltz, M. and Webster, R. 1990. A comparison of kriging, cubic splines and classification for redicting soil properties from sample information. *Journal of Soil Science* **41:**473–490.

Walkley, A. and Black, I.A. 1934. An examination of the digtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science* **37:** 29–38.

Warrick, A.W. 1998. Spatial variability. p. 655–675**.** In: D. Hillel (eds.) *Environmental soil physics*. Academic Press, New York.

Webster, R. and Oliver, M.A. 2001. *Geostatistics for Environmental Scientists*. John Wiley and Sons Ltd., Chichester.