

23. Stull, R. B., *An Introduction to Boundary Layer Meteorology*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989, p. 620.
24. Naseema Beegum, S. *et al.*, Spatial distribution of aerosol black carbon over India during pre-monsoon season. *Atmos. Environ.*, 2009, **43**, 1071–1078.
25. Babu, S. S., Sathees, S. K. and Moorthy, K. K., Aerosol radiative forcing due to enhanced black carbon at an urban site in India. *Geophys. Res. Lett.*, 2002, **29**, 1880, doi: 10.1029/2002GL015826.
26. Kunhikrishnan, P. K., Gupta, K. S., Ramachandran, R., Prakash, W. J. and Nair K. N., Study on thermal internal boundary layer structure over Thumba, India. *Ann. Geophys.*, 1993, **11**, 52–60.
27. Kumar, K. R. *et al.*, Characterization of aerosol black carbon over a tropical semi-arid region of Anantapur, India. *Atmos. Res.*, 2011, **100**, 12–27.
28. Kohler, I., Dameris, M., Ackermann, I. and Hass, H., Contribution of road traffic emissions to the atmospheric black carbon burden in the mid-1990s. *J. Geophys. Res.*, 2001, **106**, 17997–18014.

ACKNOWLEDGEMENTS. This work has been supported by grants from Indian Space Research Organization under the project ‘Studies on aerosol environment at Kolkata located near the land–ocean boundary as a part of ARFI network under ISRO-GBP’ being carried out at S.K. Mitra Centre for Research in Space Environment, University of Calcutta, Kolkata.

Received 19 July 2013; revised accepted 17 April 2014

Forecasting the distribution of heavy metals in soil and groundwater near municipal solid waste dumpsites using linear regression

Srikanth Srigriri^{1,*}, Hanmandlu Madasu², Balaram Vysetti³ and Vandana Parth³

¹Department of Ecology and Environmental Sciences, Pondicherry Central University, Pondicherry 605 014, India

²Department of Electrical Engineering, Indian Institute of Technology New Delhi, New Delhi 110 016, India

³CSIR-National Geophysical Research Institute, Hyderabad 500 007, India

The levels of heavy metals are measured at different dumpsites with different distances and directions under the jurisdiction of Greater Municipal Corporation of Hyderabad for ascertaining the soil and groundwater quality and forecasting as a part of integrated municipal solid waste (MSW) management study. The datasets indicate a steady decrease in the concentrations of ions and heavy metals in groundwater with distance from the MSW dumpsites. Similar trends are observed for the levels of heavy metals in soil at dump sites around the MSW dumpsite. In this study, we

have used both linear and quadratic regression to predict water and soil constituents. As the datasets on the components of solid waste and groundwater are limited, the desired level of forecasting accuracy could not be achieved. However, for some components the results are promising. This study suggests that improvement can be achieved by removing the outliers from the dataset. If the errors are large for a component, it would mean that we need a better way of separation of this component from the waste.

Keywords: Groundwater, heavy metals, municipal solid waste, soil.

POLLUTION is one of the major public health concerns in many large cities worldwide. However, in many cases only little attention has been given to this issue, particularly in developing countries. Example is the case of Hyderabad, where municipal solid waste (MSW) dumpsites are not scientifically maintained. One of the main activities leading to this problem includes unorganized dumping and burning of MSW which contains high levels of heavy metals. Such activities tend to increase the elemental background levels in the surrounding soil and groundwater, driving to adverse temporal variations of heavy metal levels in soils. Anthropogenically derived chemicals are an important source of environmental pollution^{1,2}. They contribute to the load of pollutants in urban run-off/leachate. In some areas close to MSW dumpsites, concentration of pollutants has reached levels which are toxic to humans and other living organisms^{3–5}. Therefore, the measurements of the fluxes of pollutants from the atmosphere in urban environments can aid in the assessment of soil and groundwater quality and can be used to determine temporal and seasonal variability of pollution sources.

Soil constitutes part of vital environmental, ecological and agricultural resources that have to be protected from further degradation as an adequate supply of healthy food needed for the world’s increasing population. Heavy metals can affect both the yield of crops and their composition⁶. Thus the elemental status of a cultivated land has to be determined to identify yield-limiting deficiencies of essential micronutrients of plants grown on polluted soils.

Some heavy metals are essential in trace amounts, namely Zn, Cu and Mn for plants and in addition, Co and Ni for animals⁷. Not much information is available on the toxicity of several metals, including Cd on either plants or animals. On the other hand, high concentrations of metals become toxic to plants and possibly are dangerous to human health. The three metals, Pb, Hg and Cd and the metalloid arsenic have all caused major human health problems in various parts of the world⁸. A number of cases of health problems related to environmental Cd poisoning have been reported. Some of the metals are phytotoxic and some are toxic to both plants and animals through their entry into the food chain^{8,9}. Recent studies

*For correspondence. (e-mail: srikanr1@gmail.com)

indicate that the platinum group of elements (Pt, Pd and Rh) is currently posing new environmental pollution problems. At present there is insufficient information about the toxicity of these elements, although some studies have revealed that platinum, in particular, would affect human health if platinum-contained dust is inhaled or contacted directly or indirectly through the food chain^{10,11}.

Baseline data for the occurrence of heavy metals as contaminants are needed as one of the criteria for assessment of critical heavy metal levels in agricultural soils¹². Over the last two decades, the study of the sources, fluxes and pathways of heavy metals has attracted attention of both national and international research communities. Environmental pollution data tend to vary extensively and are subjected to various types of uncertainties due to several factors such as distance from pollution sources and pathways, natural background variation and pollution build-up or degradation over time. Several heavy metals such as Ni, Cr and Mn are contained as trace elements in some rock types of volcanic and metamorphic origin¹³. During weathering processes, the primary crystalline structures of these rock-forming minerals are completely broken down and relevant chemical elements are thus either adsorbed in the topsoil or transported towards surface water or groundwater targets. Thus, environmental variability affects the exact variant in pollution levels between population units.

The urban and suburban population in and around Hyderabad city, India, greatly depends on groundwater for various purposes other than drinking¹⁴. Presence of any component in excess concentration compared to values prescribed by the World Health Organization¹⁵ will result in water unsuitable for irrigation, domestic or industrial uses¹⁶. The dissolved physico-chemical parameters in groundwater play a significant role in classifying and assessing water quality. Residual sodium carbonate (RSC) can be used as a decisive factor for finding the suitability of irrigation water¹⁷⁻²⁰. It was observed that the criteria used in the classification of water for a particular purpose may not satisfy the suitability standards for other purposes, but better results can be obtained only by considering the combined chemistry of all the ions rather than individual or paired ionic characters. Chemical categorization also throws light on the concentration of various predominant cations, anions and their interrelationships. Physico-chemical properties of soils depend on natural and anthropogenic factors, together acting on different spatial and temporal scales. Natural rock weathering and organic matter decomposition are related to parent material, geomorphology of the area, presence of vegetation, wind, the climatic conditions and other interactions with the environment. The effects of these processes are strictly time-dependent and exposed in a complex structure of soils. In contrast, soil management practices significantly affect pedological properties

by changing soil structure mechanically due to agricultural and urban activities, and by changing chemical composition through pollution load. The presence of any element in a fatal concentration in the soil could be due to both natural and anthropogenic factors²¹. Therefore, it is often difficult to discriminate among the different causes. The parent material largely influences heavy metal content in many soil types, with concentration sometimes exceeding the critical values. The present study was taken up to establish the levels of potentially toxic elements in soil environment around waste disposal site and to establish the levels of dissolved major ions and metal content to classify the groundwater and examine the water quality for drinking and irrigation purposes. In this case, various methods have been used to study critically the geochemical characteristics of groundwater in Hyderabad city. The major objectives of this study are: (i) to assess the levels of heavy metals in soil and groundwater distributed in the surrounding environment of MSW dumpsites of Greater Hyderabad Municipal Corporation (GHMC); (ii) to compare these levels at different MSW dumpsites, and (iii) to forecast the distribution of heavy metals in soil and groundwater near MSW dumpsites using linear regression.

The dumpsites are located in the north (Jawahar Nagar), southeast (Autonagar) and northwestern (Dundigal) parts of Hyderabad city (Figure 1). The extension of three dumpsites varies from over 200 to 400 acres of area and receives on an average 200–300 tonnes of municipal and industrial solid waste per day (~50,000 tonnes/annum according to GHMC records). The soil cover is a well-developed residue of weathered granite and consists of clay loam, red loam and sandy loam. The area is in the semi-arid zone with subtropical climatic conditions. The temperature varies between 25°C and 45°C. It receives more than 80% precipitation from SW monsoon with an average rainfall of 812 mm. Groundwater occurs in the weathered and fractured zones under water-table semi-confined conditions. The depth of weathering and fractured zones dominantly control the occurrence and movement of groundwater in these rocks. The rocks possess negligible primary porosity, but secondary porosity and permeability have occurred as a result of deep fracturing and weathering giving rise to potential aquifers. The general pattern of groundwater flow in the area is from southwest to northeast. The transmissivity of granite aquifer ranges from 30 to 200 m²/day (ref. 22). Major part of the study area is covered with pedi-plain having shallow weathering.

Ninety-four soil samples were collected during a systematic soil sampling programme. To avoid influence from various arbitrary surface conditions like waste and humus and to get assured natural in-place soil, the selected depth of sampling was from 10 to 25 cm depth. Normally, anthropogenic pollutants contaminate the upper layer of the soil. In case of natural pollutants, the entire soil at all

depths shows high level of metal enrichment. The samples were taken from geographically distributed (north, south, east and west) sites at grid intervals of 300–500 m. Only plastic equipment was used during sampling instead of metal tools to avoid any cross-contamination. The sampling period was two months between 2009 and 2011. The samples were collected in self-locking polythene bags and sealed to avoid leakage. The soil samples were air-dried and kept in an oven for 48 h at 60°C. The dried samples were then disaggregated with mortar and pestle and then finely powdered to –250 mesh size (US standard) using a swing grinding mill to make them homogeneous. The pH of soil suspension was measured using 1 : 1 soil to water mixture as recommended in the *Soil Survey Manual*²³. Weighing of sample was accomplished using analytical balance with precision of 0.0001 g.

Groundwater samples were collected from 60 bore wells in use (depth 150'–250'; diameter 8"–10") located around dumpsites at the target interval of 200–400 m in a network formation. The samples were collected during April and May (summer) when the water levels are low and the mineral contents in water are likely to reach a maximum. Samples were collected in pre-cleaned high-density polyethylene bottles from representative bore wells distributed throughout the area. The collected samples were filtered using Whatman No. 42 filter paper, and acidified with nitric acid (AR-grade) to pH < 2 (0.2% v/v). The number of samples varies from one site to the other depending upon the availability of bore wells or pumps within a particular watershed. On-site observations like location, source and depth of the bore wells were recorded. Water pH, total dissolved solids and temperature were measured instantly with corresponding pH/EC/TDS/°C portable meter. Total alkalinity was determined in non-acidified samples by titration against 0.1 M hydrochloric acid using methyl orange and phenolphthalein as indicators¹⁷. Anions (nitrate, fluoride, chloride) were analysed by double junction electrode at 25°C. Sulphate ion was determined by turbidimetric method.

Elemental composition in soil samples was determined using an X-ray fluorescence spectrometer (XRF; type Philips MagiX PRO model PW 2440 XRF) with a rhodium (Rh) anode 4 kW tube. Its high-level performance enables a sensitive and accurate determination of trace and major elements (As, Cr, Cu, Ni, Pb and Zn). With the PW 2440 XRF, it is possible to scan the elements of interest from boron to uranium. The MagiX PRO is a sequential instrument with a single goniometer-based measuring channel covering the complete elemental array. Suitable software 'super Q' was used to take care of dead time correction and inter-element matrix effects. International soil reference materials (SO-1, SO-2, SO-3, SO-4) obtained from CCRMP, CANMET Mining and Mineral Sciences Laboratories, Ontario, Canada were used to prepare the calibration curves for major and trace elements and to check the accuracy of the analytical

data²⁴. Pressed pellets for XRF analysis were prepared using collapsible aluminium cups, with a backing of boric acid. They were then pressed into pellets at 25 tonnes/inch under a hydraulic press. A PerkinElmer® Model ELAN® DRCTM II ICP mass spectrometer (PerkinElmer, Inc, Shelton, CT, USA) was used for trace element analysis utilizing the methodologies described in the literature^{25,26}. The sample introduction consisted of a standard Meinhard® nebulizer with a cyclonic spray chamber. All quantitative measurements were performed using the instrument software (ELAN v. 3.1). This software uses knowledge-driven routines in combination with numerical calculations (quantitative analysis) to perform an automated interpretation of the spectrum of interest. Several well-known isobaric interferences are programmed and the corrections are automatically applied.

The type of water that predominated in the study area was assessed based on hydrochemical facies, whereas the suitability of groundwater for irrigation was evaluated based on sodium adsorption ratio, percentage of sodium, residual sodium carbonate and the US salinity diagram. High concentrations of major ions (Ca^{++} , Mg^{++} and F^-) observed in bore wells were attributed to differential weathering of minerals such as pyroxenes, plagioclase, feldspars and apatite together with dissolution/precipitation reactions along fractures and joints in the granites. The high NO_3^- level (>50 mg/l) in groundwater is ascribed to the oxidation of ammonia and similar sources from leachates emanating from municipal waste. Although water is not suitable for domestic purposes, it is however, found to be suitable for irrigation purposes with little risk in the development of detrimental levels of exchangeable sodium.

The extent of municipal/hazardous/industrial waste dumped as landfill in the city outskirts, abundantly contaminates soil resources. The heavy metals (As, Cr, Cu, Ni, Pb, Zn) in soil samples were quantified and natural background values were used to delineate their derivation as geogenic or anthropogenic. The average concentrations of As, Cr, Pb were found to exceed the threshold and natural background values, whereas the maximum concentration of Cu, Ni and Zn exceeded the prescribed threshold limit. Soil pH varies from 5.7 to 8.9 and is acidic to near neutral and alkaline in nature. Soil pH significantly affects the solubility and mobility of these metals, as most of them are soluble in acid soils than in neutral or slightly basic soils. The methodology used has proved to be a useful tool to separate geological and anthropogenic causes of variation in soil heavy metal content and to identify common pollution sources.

Various forecasting and prediction techniques are available in the literature. Selection of the technique(s) normally depends on the availability of data, the quality of available models and some assumptions. Each method is different in terms of accuracy, scope, time and cost. To facilitate an adequate level of forecasting accuracy, the

developer has to be responsive to the characteristics of different methods, and determine if a particular method is appropriate for the situation on hand before embarking on its usage in real applications. As a result, the choice of a forecasting model is one of the important factors that will influence the forecasting accuracy.

Scientific data were collected from various dumpsites around Greater Hyderabad. Here we will use the data for past periods and forecast the values.

In statistics, linear regression is an approach adopted for modelling the relationship between a scalar variable y and one or more explanatory variables denoted as X (ref. 27). The case of one explanatory variable is called simple regression. More than one explanatory variable is multiple regressions²⁸.

In linear regression, scientifically collected data are modelled using linear functions, and unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, linear regression refers to a model in which the conditional mean of y given the value of X is an affine function of X . Less commonly, linear regression could refer to a model in which the median, or some other quantile of the conditional distribution of y given X is expressed as a linear function of X . Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X , rather than on the joint probability distribution of y and X , which is the domain of multivariate analysis.

Linear regression is the first type of regression analysis used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than those which are nonlinearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

Linear regression has many practical uses. Most applications of linear regression fall into one of the following two broad categories: If the goal is prediction, or forecasting, linear regression can be used to fit a predictive model to an observed dataset of y and X values. After developing such a model, if an additional value of X is then given without its accompanying value of y , the fitted model can be used to make a prediction of the value of y .

Given a variable y and a number of variables X_1, \dots, X_p that may be related to y , linear regression analysis can be applied to quantify the strength of the relationship between y and the X_j , to assess which X_j may have no relationship with y at all, and also which subsets of X_j contain redundant information about y .

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the 'lack of fit' in some other norm (as with least absolute deviation regression), or by minimizing a penalized version of the least squares loss function as in ridge regression. Conversely, the least squares approach can be used to fit models that are not

linear models. Thus, while the terms 'least squares' and 'linear model' are closely linked, they are not synonymous²⁹.

A brief description of the linear and quadratic regression models and the estimation of their coefficients using the least squares estimation techniques are given in Appendix A.

Following the economic benefits of forecasting³⁰, we embark on predicting waste materials from dump sites. In this study, three dump sites: Autonagar, Dundigal and Jawahar Nagar have been selected for the forecasting of components whether of solid waste or polluted groundwater. Three analyses are done on the solid and liquid wastages at each site. These comprise: ICP-MS analysis of metals in groundwater, major ion concentration in the groundwater and XRF analysis of the soil. The data are divided into two sets: training set and test set which is used to check the forecasting accuracy. The sample size varies from site to site and one analysis to another. ICP-MS analysis (Autonagar – training: 16, test: 5; Dundigal and Jawahar Nagar – training: 17, test: 5), major ion concentration (Autonagar – training: 16, test: 5; Dundigal – training: 17, test: 5; Jawahar Nagar – training: 17, test: 5) and XRF analysis (Autonagar – training: 17, test: 10; Dundigal – training: 35, test: 10; Jawahar Nagar – training: 17, test: 5). We have applied two types of regression: linear regression and quadratic regression on each component of the solid and groundwater wastes to build the models based on the training dataset. Then the learned models have been used for the forecast; the error of forecasts is determined using the test dataset.

We present the coefficients of linear and quadratic regression using equations (A1) and (A2) respectively (see Appendix A). As an example, we consider Cr and Mn for modelling using the regression.

Linear regression model (A1) for Cr:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

The coefficients are $\beta_0 = 47.4362$; $\beta_1 = 0.4380$.

Quadratic regression model (A2) for Mn:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, \dots, n.$$

The coefficients are:

$$\begin{aligned} \beta_0 &= 1695.81465738939; \\ \beta_1 &= -1.63921843163745; \\ \beta_2 &= 0.000857889295506935. \end{aligned}$$

Tables 1–3 (Autonagar), 4–6 (Dundigal) and 7–9 (Jawahar Nagar) display two types of error measures: per unit error and sum of squared error. Maximum errors are witnessed at the Dundigal site and minimum at the Jawahar Nagar site. For certain components like Cd and Sb the

RESEARCH COMMUNICATIONS

Table 1. ICP-MS analysis of metals in groundwater (Autonagar)

Method	Error	Li	Be	B	Al	Si	V	Cr	Mn	Se	Rb	Pb	
Linear regression	Per unit error	4.7586	1.7665	2.1182	2.418	1.4007	3.319	0.7457	1.828	6.673	1.4497	4.9667	
	Sum of squared error	6.7531	0.8803	1.1809	1.6305	0.7164	3.2268	0.2006	1.0608	11.978	0.4506	6.5448	
Quadratic regression	Per unit error	4.7586	3.2758	2.8426	2.4174	1.3255	2.4096	0.7475	1.0314	7.7437	1.3614	5.7128	
	Sum of squared error	6.7531	3.403	2.2728	1.6314	0.6464	1.9673	0.2014	0.6199	18.546	0.4937	11.288	
		Fe	Ni	Co	Cu	Zn	As	Sr	Mo	Ag	Cd	Sb	Ba
Linear regression	Per unit error	5.4073	7.24	5.5469	6.4223	1.8781	9.0636	12.279	5.2428	0.7146	9.0492	8.7093	2.7583
	Sum of squared error	7.3285	14.595	10.84	9.8567	1.1489	22.961	37.533	5.5037	0.1724	18.994	20.392	3.397
Quadratic regression	Per unit error	5.048	7.1549	5.2884	6.8771	2.6581	8.3625	11.291	4.1491	0.7363	8.963	7.3622	2.4669
	Sum of squared error	6.615	14.29	9.9573	11.017	2.245	19.861	32.169	3.5085	0.158	18.601	14.254	2.9935

Table 2. Major ion concentration in groundwater (Autonagar)

Method	Error	HCO ₃ ⁻	CO ₃ ⁻	F ⁻	Cl ⁻	NO ₃ ⁻	SO ₄ ²⁻	Na ⁺	Mg ⁺⁺	K ⁺	Ca ⁺⁺
Linear regression	Per unit error	4.0266	1.5978	1.3499	5.809	7.9874	5.9307	1.7928	8.1651	4.8252	4.1863
	Sum of squared error	6.3846	0.8119	0.4422	10.7803	24.3111	11.4156	1.1998	20.4011	4.7884	6.7174
Quadratic regression	Per unit error	3.8815	1.4465	1.2157	5.4374	8.0838	6.4253	1.7841	9.0359	1.2848	4.432
	Sum of squared error	6.2628	0.8072	0.3955	8.196	22.0077	12.6279	1.2142	24.5548	0.4857	7.4466

Table 3. XRF analysis of soil (Autonagar)

Method	Error	Ba	Co	Cr	Cu	Mo	Ni	Pb	V	Zn
Linear regression	Per unit error	1.5779	3.4592	3.9538	2.5135	3.6186	4.0466	2.5436	2.2346	2.3266
	Sum of squared error	0.3949	1.3879	2.831	0.873	2.2089	3.0664	1.2604	0.7132	0.6782
Quadratic regression	Per unit error	1.3658	6.0739	4.0426	2.5989	4.0603	3.8321	2.7268	3.09	2.1737
	Sum of squared error	0.2778	4.7082	3.0099	0.9452	2.7718	2.7811	1.3878	2.1636	0.6158

Table 4. ICP-MS analysis of metal in groundwater (Dundigal)

Method	Error	Li	Be	B	Al	Si	V	Cr	Mn	Se	Rb	Pb	
Linear regression	Per unit error	1.527	4.308	4.591	8.579	0.564	8.251	10.61	4.228	17.88	13.9	1.589	
	Sum of squared error	0.569	5.278	5.762	25.35	0.11	18.37	35.63	13.03	76.8	60.18	0.734	
Quadratic regression	Per unit error	1.85	6.595	3.166	13.31	0.609	9.147	9.193	4.156	7.29	12.2	1.211	
	Sum of squared error	0.9	14.67	3.163	69.92	0.116	22.05	27.03	13.63	15.96	41.29	0.382	
		Fe	Ni	Co	Cu	Zn	As	Sr	Mo	Ag	Cd	Sb	Ba
Linear regression	Per unit error	1.388	6.201	2.893	6.989	3.973	23.61	8.425	2.781	16.27	25.08	25.72	4.596
	Sum of squared error	0.829	11.58	3.357	21.57	4.074	149.7	17.7	2.038	66.64	205.3	217.3	8.723
Quadratic regression	Per unit error	1.152	15.62	2.889	8.708	4.042	16.34	7.072	3.247	15.48	24.05	19.4	4.517
	Sum of squared error	0.454	73.84	3.349	29.84	4.137	74.65	11.68	2.831	60.02	188.9	117.8	8.504

Table 5. Major ion concentration in groundwater (Dundigal)

Method	Error	HCO ₃ ⁻	CO ₃ ⁻	F ⁻	Cl ⁻	NO ₃ ⁻	SO ₄ ²⁻	Na ⁺	Mg ⁺⁺	K ⁺	Ca ⁺⁺
Linear regression	Per unit error	1.721	4.3345	1.5006	3.408	6.6924	1.0163	1.4748	6.2773	1.8324	4.0632
	Sum of squared error	1.0864	9.0901	0.7243	3.5957	10.1255	0.3327	0.7051	19.1765	0.9745	5.5598
Quadratic regression	Per unit error	1.5095	4.9834	1.3428	3.1763	3.5689	1.0944	1.4333	5.1024	2.1974	5.0072
	Sum of squared error	1.0638	11.5651	0.4474	3.3502	3.6558	0.361	0.6949	15.5166	1.2114	8.8

errors are very large indicating that there might be outliers due to unscrupulous data collection. But we have refrained from correcting the outliers, as it requires large

datasets. Figures 2 (Autonagar), 3 (Dundigal) and 4 (Jawahar Nagar) depict the errors of forecasting of three components in each dump site. These figures correspond

Table 6. XRF analysis of soil (Dundigal)

Method	Error	As	Ba	Co	Cr	Cu	Mo	Ni	Pb	Rb	V	Y	Zn	Zr
Linear regression	Per unit error	10.03	3.625	16.95	7.934	4.122	2.919	5.257	4.525	4.87	2.516	2.822	5.342	6.445
	Sum of squared error	17.81	1.59	85.46	9.942	2.441	1.386	3.837	3.543	5.05	0.913	1.087	3.326	7.507
Quadratic regression	Per unit error	9.936	3.705	16.04	8.666	4.511	2.915	5.717	7.099	5.023	2.549	2.873	9.352	6.794
	Sum of squared error	17.55	1.987	61.73	12.84	2.96	1.383	5.39	9.768	5.327	0.955	1.128	10.39	8.398

Table 7. ICP-MS analysis of metal in groundwater (Jawahar Nagar)

Method	Error	B	Al	Si	V	Cr	Mn	Fe	Ni
Linear regression	Per unit error	0.9181	2.0738	1.9873	0.9857	1.8014	4.1276	2.4021	4.0344
	Sum of squared error	0.266	1.1826	1.0215	0.2629	0.6607	5.6841	1.1792	4.183
Quadratic regression	Per unit error	1.7285	2.1201	3.8005	0.9857	1.5621	3.7966	2.7786	4.0381
	Sum of squared error	0.8654	1.1774	4.8696	0.2907	0.5017	4.4671	1.5616	4.3604

		Co	Cu	Zn	As	Se	Mo	Ag	Cd	Ba	Pb
Linear regression	Per unit error	1.9643	1.3105	7.5815	2.1226	9.1435	1.8259	0.4628	0.1631	1.8869	1.1979
	Sum of squared error	0.9388	0.5163	14.8336	1.7514	34.9149	1.0847	0.0528	0.0097	1.4218	0.3658
Quadratic regression	Per unit error	4.586	3.5929	5.2114	2.883	9.0632	3.2792	0.3423	0.1628	1.875	1.0679
	Sum of squared error	10.194	3.4327	12.4131	3.4798	29.7724	2.966	0.0339	0.0104	1.3932	0.3025

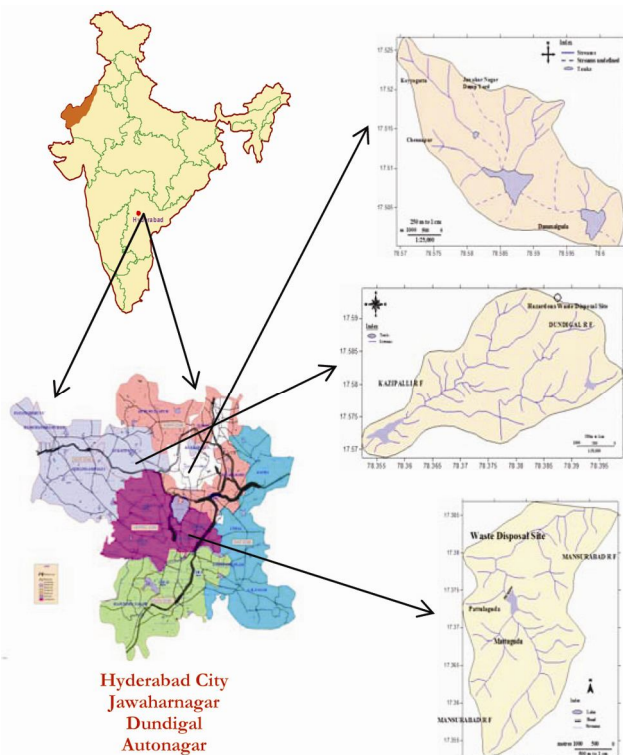


Figure 1. Watershed map of the study area.

to linear regression. Three forecasting results of HSO₃, Sb and Mn using quadratic regression are shown in Figure 5. Inclusion of all figures of quadratic regression would require additional spacing but with not much usefulness as the errors are almost similar for most of the components.

The results of forecasts of the materials (solid waste and liquid waste) at three dumpsites are presented using both linear regression and quadratic regression. Owing to the limited data, the forecasting errors are found to be large at dump sites like Dundigal. The main problem is that the components (constituents) from solid and groundwater are separated manually and automation needs to be introduced.

An analysis of the results shows that in some cases linear regression is better than quadratic regression and vice versa in other cases. We cannot choose one as superior to the other. If we have additional information about the exogenous variables like population density, literacy and waste management facilities, we can improve the results of forecasting. However, this preliminary study has opened new vistas for improvement in the waste management.

Future work will be aimed at finding the outliers in the datasets and trying out fuzzy models.

Appendix A

Linear regression

In linear regression, the model specification is that the dependent variable, y_i is a linear combination of the parameters (but need not be linear in the independent variables). For example, in simple linear regression for modelling n data points, there is one independent variable x_i , and two parameters β_0 and β_1

$$\text{Linear: } y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n. \quad (A1)$$

RESEARCH COMMUNICATIONS

Table 8. Major ion concentration in groundwater (Jawahar Nagar)

Method	Error	F ⁻	Cl ⁻	NO ₃ ⁻	SO ₄ ²⁻	HCO ₃ ⁻	CO ₃ ⁻	Na ⁺	Mg ⁺⁺	K ⁺	Ca ⁺⁺
Linear regression	Per unit error	1.6096	3.1866	2.6413	2.7136	3.1872	1.7132	2.0708	8.1398	1.4243	2.0428
	Sum of squared error	1.1995	3.0855	2.9664	1.9076	4.2184	0.612	1.7821	16.3024	0.502	1.4178
Quadratic regression	Per unit error	1.6714	2.7533	3.2144	2.4059	3.1372	1.7002	1.7665	11.2655	2.17	2.0513
	Sum of squared error	1.1856	2.2825	4.9804	1.7956	4.0503	0.6344	0.8526	28.2237	1.0676	1.2696

Table 9. XRF analysis of soil (Jawahar Nagar)

Method	Error	As	Ba	Co	Cr	Cu	Ni	Pb	Rb	V	Zn	Zr
Linear regression	Per unit error	3.4102	1.3918	4.74	6.7019	1.9789	1.5272	1.0348	1.3186	0.8177	1.3044	3.8571
	Sum of squared error	3.9728	0.4676	7.3139	11.7734	2.0305	0.4803	0.4229	0.5881	0.1682	0.4098	5.8088
Quadratic regression	Per unit error	3.3806	1.2202	4.6574	6.6091	1.8393	1.4875	1.0209	0.8268	1.4322	1.4994	4.7934
	Sum of squared error	4.248	0.3842	7.2385	17.7509	1.707	0.4587	0.4261	0.2338	0.6078	0.589	9.7972

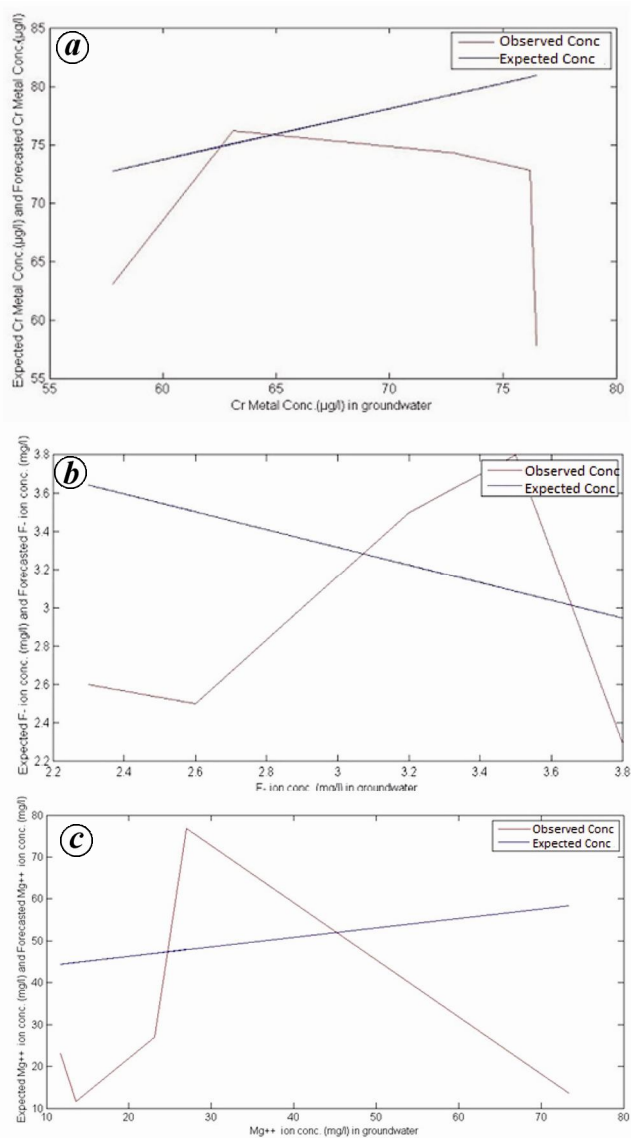


Figure 2. a, ICP-MS analysis of Cr metal; b, F⁻ ion concentration; c, Mg⁺⁺ ion concentration in groundwater (Autonagar).

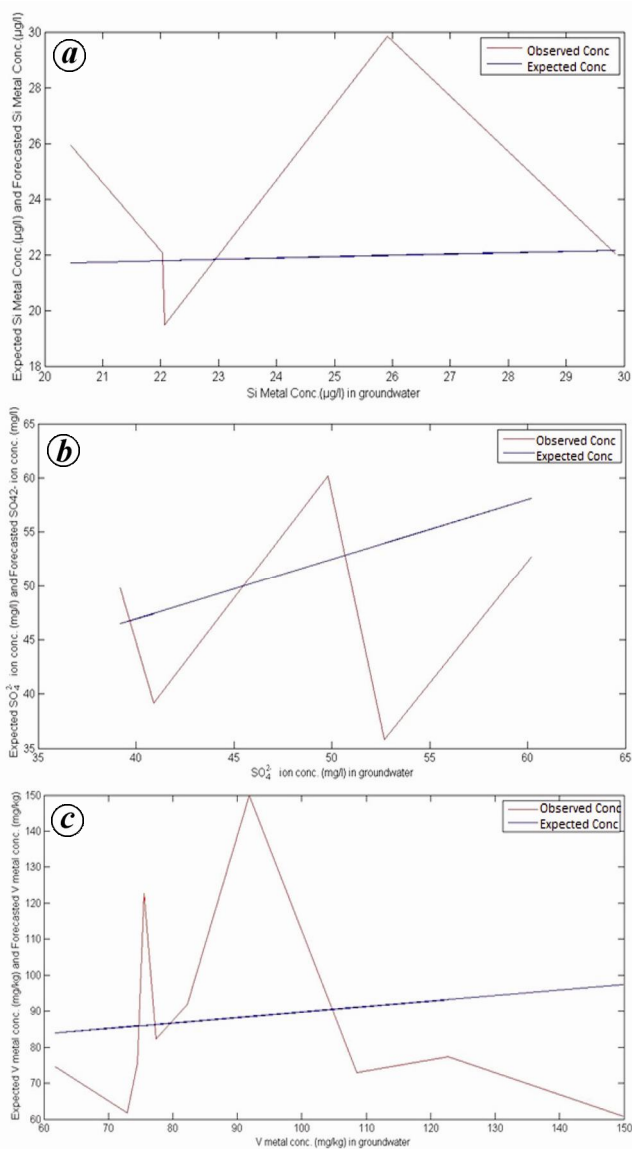


Figure 3. a, ICP-MS analysis of Si metal; b, SO₄²⁻ ion concentration; c, XRF analysis of V metal in groundwater (Dundigal).

(In multiple linear regressions, there are several independent variables or functions of independent variables.) Adding a term in x_i^2 to the preceding regression gives

$$\text{Quadratic: } y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, \dots, n. \quad (\text{A2})$$

This is still linear regression; although the expression on the right hand side is quadratic in the independent variable x_i , it is linear in the parameters β_0 , β_1 and β_2 .

In both cases, ε_i is an error term and the subscript i indexes a particular observation. Given a random sample from the population, we estimate the population parameters and obtain the sample linear regression model

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \quad (\text{A3})$$

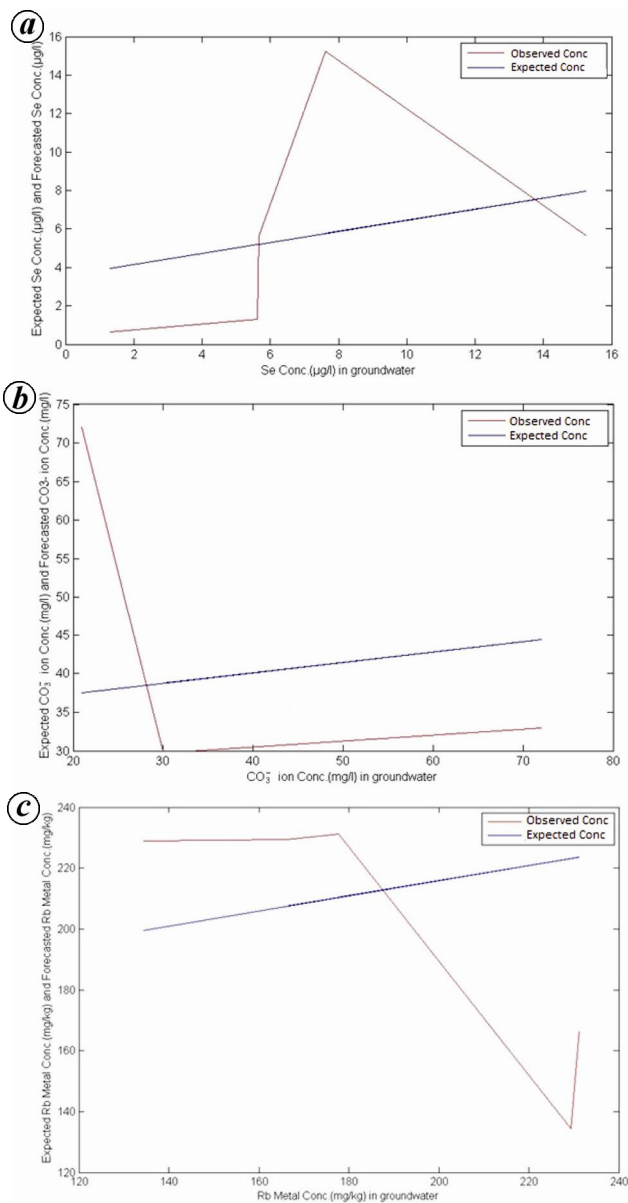


Figure 4. a, ICP-MS analysis of Se metal; b, CO_3^- ion concentration; c, XRF analysis of Rb metal in groundwater (Jawahar Nagar).

The residual $e_1 = y_i - \hat{y}_i$, is the difference between the value of the dependent variable predicted by the model, \hat{y}_i and the true value of the dependent variable y_i . One method of estimation is ordinary least squares. This method obtains parameter estimates that minimize the

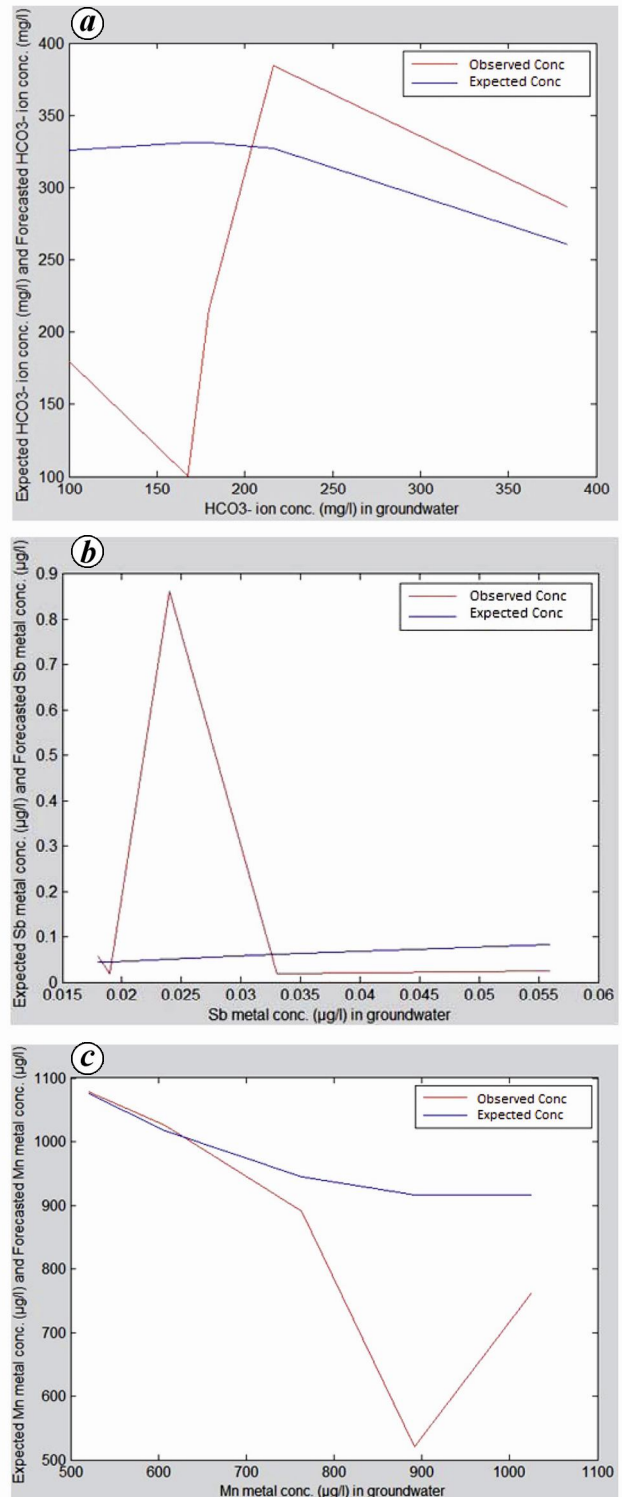


Figure 5. Results of forecasting using quadratic regression on (a) HCO_3^- ; (b) Sb; (c) Mn.

sum of squared residuals, SSE, also sometimes denoted as RSS

$$SSE = \frac{1}{N} \sum_1^N e_i^2. \tag{A4}$$

Minimization of this function results in a set of normal equations, a set of simultaneous linear equations in the parameters, which are solved to yield the parameter estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$.

In the case of simple regression, the formulas for the least squares estimates are

$$\hat{\beta}_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \tag{A5}$$

where \bar{x} is the mean (average) of the x values and \bar{y} is the mean of the y values.

Given a dataset $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y_i and the p -vector of regressors x_i is linear. This relationship is modelled through a disturbance term or error variable ε_i – an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = x_i' \beta + \varepsilon_i, \quad i = 1, \dots, n, \tag{A6}$$

where (A6) takes the form $x_i' \beta$ is the inner product between vectors x_i and β .

Often these n equations are stacked together and written in vector form as

$$y = X\beta + \varepsilon, \tag{A7}$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ x_{12} & \dots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Some remarks on terminology and general use y_i are called the regress and exogenous variable, response variable, measured variable or dependent variable. The deci-

sion as to which variable in a dataset is modelled as the dependent variable and which are modelled as the independent variables may be based on a presumption that the value of one of the variables is caused by, or directly influenced by the other variables. Alternatively, there may be an operational reason to model one of the variables in terms of the others, in which case there need be no presumption of causality.

x_i are called regressors, endogenous variables, explanatory variables, covariates, input variables, predictor variables or independent variables. The matrix X is sometimes called the design matrix.

(a) Usually a constant is included as one of the regressors. For example, we can take $x_{i1} = 1$ for $i = 1, \dots, n$. The corresponding element of β is called the intercept. Many statistical inference procedures for linear models require an intercept to be present. So it is often included even if theoretical considerations suggest that its value should be zero.

(b) Sometimes one of the regressors can be a nonlinear function of another regressor or of the data, as in polynomial regression and segmented regression. The model remains linear as long as it is linear in the parameter vector β .

(c) The regressors x_{ij} may be viewed either as random variables, which we simply observe, or they can be considered as predetermined fixed values which we can choose. Both interpretations may be appropriate in different cases, and they generally lead to the same estimation procedures; however, different approaches to asymptotic analysis are used in these two situations.

β is a p -dimensional parameter vector. Its elements are also called effects, or regression coefficients. Statistical estimation and inference in linear regression focuses on β .

ε_i is called the error term, disturbance term or noise. This variable captures all other factors which influence the dependent variable y_i other than the regressors x_i .

The relationship between the error term and the regressors, for example whether they are correlated, is a crucial step in formulating a linear regression model, as it will determine the method to be used for estimation.

A fitted linear regression model can be used to identify the relationship between a single predictor variable x_j and the response variable y when all the other predictor variables in the model are ‘held fixed’. Specifically, the interpretation of β_j is the expected change in y for a one-unit change in x_j when the other covariates are held fixed. This is sometimes called the unique effect of x_j on y . In contrast, the marginal effect of x_j on y can be assessed using a correlation coefficient or simple linear regression model relating x_j to y .

Care must be taken when interpreting regression results, as some of the regressors may not allow for marginal changes (such as dummy variables, or the intercept term), while others cannot be held fixed (recall the example

from the introduction: it would be impossible to ‘hold t_i fixed’ and at the same time change the value of t_i^2 .

It is possible that the unique effect can be nearly zero even when the marginal effect is large. This may imply that some other covariate captures all the information in x_j , so that once that variable is in the model, there is no contribution of x_j to the variation in y . Conversely, the unique effect of x_j can be large while its marginal effect is nearly zero.

The meaning of the expression ‘held fixed’ may depend on how the values of the predictor variables arise. If the experimenter directly sets the values of the predictor variables according to a study design, the comparisons of interest may literally correspond to comparisons among units whose predictor variables have been ‘held fixed’ by the experimenter. Alternatively, the expression ‘held fixed’ can refer to a selection that takes place in the context of data analysis. In this case, we ‘hold a variable fixed’ by restricting our attention to the subsets of the data that happen to have a common value for the given predictor variable. This is the only interpretation of ‘held fixed’ that can be used in an observational study.

The notion of a ‘unique effect’ is appealing when studying a complex system where multiple interrelated components influence the response variable. In some cases, it can be literally interpreted as the causal effect of an intervention that is linked to the value of a predictor variable. However, it has been argued that in many cases multiple regression analysis fails to clarify the relationships between the predictor variables and the response variable when the predictors are correlated with each other and are not assigned following a study design.

Least squares estimation techniques

Errors-in-variables models (or ‘measurement error models’) extend the traditional linear regression model to allow the predictor variables X to be observed with error. This error causes standard estimators of β to become biased. Generally, the form of bias is attenuation, meaning that the effects are biased toward zero.

Ordinary least squares

Ordinary least squares (OLS) is the simplest and thus most common estimator. It is conceptually simple and computationally straightforward. OLS estimates are commonly used to analyse both experimental and observational data. The OLS method minimizes the sum of squared residuals and leads to a closed-form expression for the estimated value of the unknown parameter β

$$\hat{\beta} = (X'X)^{-1} X'y = \left(\frac{1}{n} \sum x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum x_i y_i \right). \quad (\text{A8})$$

The estimator is unbiased and consistent if the errors have finite variance and are uncorrelated with the regressors

$$E[x_i \varepsilon_i] = 0. \quad (\text{A9})$$

It is also efficient under the assumption that the errors have finite variance and are homoscedastic, meaning that $E[\varepsilon_i^2|x_i]$ does not depend on i . The condition that the errors are uncorrelated with the regressors will be generally satisfied in an experiment, but in the case of observational data, it is difficult to exclude the possibility of an omitted covariate z that is related to both the observed covariates and the response variable. The existence of such a covariate will generally lead to a correlation between the regressors and the response variable, and hence to an inconsistent estimator of β . The condition of homoscedasticity can fail with either experimental or observational data. If the goal is either inference or predictive modelling, the performance of OLS estimates can be poor if multicollinearity is present, unless the sample size is large.

In simple linear regression, where there is only one regressor (with a constant), the OLS coefficient estimates have a simple form that is closely related to the correlation coefficient between the covariate and the response.

Generalized least squares

Generalized least squares (GLS) is an extension of the OLS method that allows efficient estimation of β when either heteroscedasticity or correlations, or both are present among the error terms of the model, as long as the form of heteroscedasticity and correlation is known independently of the data.

To handle heteroscedasticity when the error terms are uncorrelated with each other, GLS minimizes a weighted analogue to the sum of squared residuals from OLS regression, where the weight for the i th case is inversely proportional to $\text{var}(\varepsilon_i)$. This special case of GLS is called ‘weighted least squares’. The GLS solution to estimation problem is

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y, \quad (\text{A10})$$

where Ω is the covariance matrix of the errors. GLS can be viewed as applying a linear transformation to the data so that the assumptions of OLS are met for the transformed data. For GLS to be applied, the covariance structure of the errors must be known up to a multiplicative constant.

1. Boulding, J. R., *Practical Handbook of Soil, Vadose Zone, and Groundwater Contamination: Assessment, Prevention, and Remediation*, Lewis Publishers, Boca Raton, FL, 1995.

2. Chandrasekhar, K., Chary, N. S., Kamala, C. T., Venkateswara-Rao, J., Balam, V. and Anjaneyulu, Y., Risk assessment and pathway study of arsenic in industrially contaminated sites of Hyderabad: a case study. *Environ. Int.*, 2003, **29**, 601–611.
3. Marina, I., Margherita, L., Angela, N. and Alfredo, P., Toxicity identification evaluation of leachates from municipal solid waste landfills: a multispecies approach. *Chemosphere*, 2003, **52**, 85–94.
4. Chandrasekhar, K., Kamala, C. T., Chary, N. S., Balam, V. and Garcio, G., Potential of *Hemidesmns indicus* for phytoextraction of lead from industrially contaminated soils. *Chemosphere*, 2005, **58**, 507–514.
5. Gaur and Adholeya, A., Prospects of arbuscular mycorrhizal fungi in phytoremediation of heavy metal contaminated soils. *Curr. Sci.*, 2004, **86**(4), 528–534.
6. Khurana, M. P. W. S. and Singh, P., Waste water use in crop production: a review. *Resour. Environ.*, 2012, **2**(4), 116–131.
7. Kirmani, M. Z., Mohiuddin, S., Farah, N., Naqvi, I. M. and Zahir, E., Determination of some toxic and essential trace metals in some medicinal and edible plants of Karachi city. *J. Basic Appl. Sci.*, 2011, **7**(2), 89–95.
8. Hutton, M., Human health concerns of lead, mercury, cadmium and arsenic. In *Lead, Mercury, Cadmium and Arsenic in the Environment* (eds Hutchinson, T. C. and Meema, K. M.), John Wiley, 1987.
9. Shankar, A. K., Cervantes, C., Loza-Tavera, H. and Avudainayagam, S., Chromium toxicity in plants. *Environ. Int.*, 2005, **31**, 739–753.
10. Ramavati, M., Balam, V., Satyanarayanan, M., Sawant, S. S. and Ramesh, S. L., Anthropogenic platinum, palladium and rhodium concentrations in road dusts from Hyderabad city, India. *Environ. Earth Sci.*, 2011, **62**, 1085–1098.
11. Balam, V., Kamala, C. T., Ramavathi Mathur, Srinivasa Rao, A. and Satyanarayanan, M., PGE in road dust/top soil from major Indian cities: implications on human health and environment. In *On a Sustainable Future of the Earth's Natural Resources* (ed. Ramkumar, Mu.), Springer Earth System Sciences, 2013, pp. 85–108.
12. Rashad, M. and Shalaby, E. A., Dispersion and deposition of heavy metals around two municipal solid waste (MSW) dumpsites, Alexandria, Egypt. *Am. Eurasian J. Agric. Environ. Sci.*, 2007, **2**(3), 204–212.
13. Alloway, B. J., *Heavy Metals in Soils*, Blackie Academic and Professional, London, 1995, p. 368.
14. Reddy, M. V., Sagar Babu, K., Balam, V. and Satyanarayanan, M., Assessment of the effects of municipal sewage, immersed idols and boating on the heavy metal and other elemental pollution of surface water of the eutrophic Hussainsagar Lake (Hyderabad, India). *Environ. Monit. Assess.*, 2012, **184**(4), 1991–2000.
15. WHO, *Guidelines for Drinking Water Quality*, World Health Organization, 2004, 3rd edn, ISBN 9241546387.
16. Chapman, D., *Water Quality Assessments – A Guide to Use of Biota, Sediments and Water in Environmental Monitoring*, UNESCO/WHO/UNEP, University Press, Cambridge, 1996, p. 609.
17. APHA, *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association, Washington, DC, 1998, 20th edn.
18. BIS, Drinking water. Bureau of Indian Standard, IS 10500, 1991.
19. Balam, V., Recent trends in inorganic analysis of environmental materials. *Res. J. Chem. Environ.*, 2002, **6**(2), 69–79.
20. Vandana, P., Murthy, N. N. and Saxena, P. R., Hydrogeochemistry of subsurface water around waste disposal sites in Hyderabad city, India: geogenic and anthropogenic consequence. *J. Toxicol. Environ. Health Sci.*, 2011, **3**(9), 254–263.
21. Vandana, P., Murthy, N. N. and Saxena, P. R., Assessment of heavy metal contamination in soil around hazardous waste disposal sites in Hyderabad city (India): natural and anthropogenic implications. *J. Environ. Res. Manage.*, 2011, **2**(2), 27–34.
22. Ahmed, S., Saxena, V. K., Subrahmanyam, K. and Kumar, D., Spatial variability and correlation of hydrochemical parameters in Meheshwaram watershed, Andhra Pradesh. In Proceedings of the International Conference on Sustainable Development and Management of Groundwater Resources in Semi-Arid Region with Special Reference to Hard Rocks (eds Thangarajan, M., Rai, S. N. and Singh, V. S.), Oxford & IBH, New Delhi, 2002, pp. 268–278.
23. *Soil Survey Manual*, Soil Conservation Service, US Department of Agriculture Handbook No.18, 1993.
24. Govil, P. K., X-ray fluorescence analysis of major, minor and selected trace elements in new IWG reference rock samples. *J. Geol. Soc. India*, 1985, **26**, 38–42.
25. Balam, V., Characterization of trace elements in environmental samples by IC-MS. *At. Spectrosc.*, 1993, **14**(6), 174–179.
26. Balam, V. et al., Use of xenon as an internal standard for the determination of trace elements in water samples by ICP-MS. *At. Spectrosc.*, 2012, **33**(2), 41–47.
27. Draper, N. R. and Smith, H., *Applied Regression Analysis*, Wiley Series in Probability and Statistics, 1998.
28. Cohen, J., Cohen, P., West, S. G. and Aiken, L. S., *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, NJ, 2003, 2nd edn.
29. Kaw, A. and Kalu, E. E., *Numerical Methods with Applications*, 2008, 1st edn; <http://www.autakaw.com>
30. Pindyck, R. S. and Rubinfeld, D. L., *Econometric Models and Economic Forecasts*, McGraw-Hill, Boston, Irwin, 1998, 4th edn.

ACKNOWLEDGEMENT. We thank Mr Mohit Meena for providing computational assistance during this work.

Received 3 April 2013; revised accepted 25 March 2014

Hydrogen generation by gamma radiolysis of aqueous suspension of nano zirconia

A. G. Kumbhar^{1,*}, Y. K. Bhardwaj² and D. B. Naik³

¹Water and Steam Chemistry Division, ²Radiation Technology Development Division, ³Radiation and Photochemistry Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400 085, India

Aqueous suspension of nano zirconia with methanol was irradiated with gamma (γ) rays. Hydrogen (H_2) generated in this process was studied as a function of pH and γ -dose. In the presence of 1 M methanol and at pH 3.0, low gamma dose irradiation showed optimum H_2 generation. This is explained on the basis zeta potential and surface charge on zirconia particles. Positive surface charge at low pH could be the reason for enhanced H_2 generation. Maximum H_2 yield $G(H_2)$ of 3.7 was observed. This is 400 times more compared to nano pure water. The method can be utilized for building a medium-scale H_2 generation plant.

*For correspondence. (e-mail: agk@barc.gov.in)