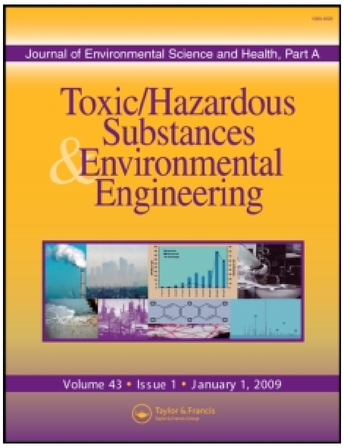

Journal of Environmental Science and Health, Part A
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597268

Geostattstical analysis of soil arsenic content in taiwan

Tsun-Kuo Chang [a]; Guey-Shin Shyu [b]; Yu-Pin Lin [c]; Nan-Chang Chang [a]
[a] Graduate Institute of Agricultural Engineering, National Taiwan University, Taipei, Taiwan,Republic of China [b] Department of Environmental Engineering, Tung-Nan Junior College of Technology, Taipei, Taiwan,Republic of China [c] Department of Landscape Architecture, Chinese Culture University, Taipei, Taiwan,Republic of China

To cite this Article Chang, Tsun-Kuo, Shyu, Guey-Shin, Lin, Yu-Pin and Chang, Nan-Chang'Geostattstical analysis of soil arsenic content in taiwan', Journal of Environmental Science and Health, Part A, 34: 7, 1485 — 1501
To link to this Article: DOI: 10.1080/10934529909376907
URL: http://dx.doi.org/10.1080/10934529909376907

J. ENVIRON. SCI. HEALTH, A34(7), 1485–1501 (1999)

GEOSTATISTICAL ANALYSIS OF SOIL ARSENIC CONTENT IN TAIWAN

Key Words: Arsenic, geostatistics, kriging, Geographic Information Systems

Tsun-Kuo Chang[1], Guey-Shin Shyu[2], Yu-Pin Lin[3] and Nan-Chang Chang[1]

[1]Graduate Institute of Agricultural Engineering, National Taiwan University, Taipei, Taiwan 10617 Republic of China
[2]Department of Environmental Engineering, Tung-Nan Junior College of Technology, Taipei, Taiwan 22202 Republic of China
[3]Department of Landscape Architecture, Chinese Culture University, Taipei, Taiwan 11114 Republic of China

ABSTRACT

A combined geostatistical and computergraphic approach was developed for illustrating the soil Arsenic map of Taiwan. Data were collected from the Environmental Protection Administration's study targeting agricultural soils in Taiwan. The range and arithmetic mean of the As contents in the surface soils (0 to 15 cm) of the study samples are as follows: 0.01 to 16.16, 5.65 (mg/kg dry soil). Variograms (Semivariograms) indicated spatial correlation at distances up to 195.0 km. The data exhibited some anisotropy, but this had little effect on kriging. An exponential variogram model was fitted using least squares and used to krige a grid covering Taiwan. Soils southwest of Taiwan tended to contain higher levels of As than average. The map will be useful in future research to determine the geographic distribution of regional patterns of plants and groundwater As content, the relationship between As and parent soil material, and correlation with of occurrence of blackfoot disease.

INTRODUCTION

Arsenic is an element toxic to humans and is suspected to cause the blackfoot disease in Taiwan (Hung and Laio, 1996). Maps illustrating the geographic distribution of soil As in Taiwan would contribute to better understanding of the correlation between soil As content and the occurrence of blackfoot disease. In addition, these maps would be useful in considering the effects of As content in groundwater and crops.

Rational soil management requires an understanding of how properties such as pH, organic matter, texture, and nutrient status vary across the land. We can use soil sampling and analysis of relevant soil properties to develop maps that delineate areas that would benefit from different management, or to identify areas of particular concern. Inherent in this process is the assumption that a property measured at a given point represents the surrounding unsampled neighborhood. The validity of this assumption depends on the spatial variability of the soil property. Geostatistics provides the means to characterize and quantify spatial variability, to use this information in rational interpolation, and to estimate the variance of the interpolated values. Variance estimation provides valuable information on the sampling density and configuration necessary for estimating a property to a specified precision. Geostatistics has been used to characterize spatial variability and to map a variety of soil properties at scales ranging from centimeters to kilometers (Warrick et al., 1986), and it may prove useful across even greater distances (White et al., 1997). Examples of other works include those of Burgess and Webster (1980), Bierkens and Burrough (1993), Bourennane et al. (1996), Davies (1997), and Smith et al. (1993) among others.

Studies by the Environmental Protection Administration (EPA) in a collaborative research program initiated in 1983 aimed to determine the As, Cd, Cu, Cr, Hg, Ni, Pb and Zn trace element contents in soils and other soil properties, such as cation-exchange capacity and pH. Soils were sampled from 878 sites representing important agricultural production areas across Taiwan as described in reports on the elemental contents of soils in Taiwan. Total soil As ranged from

0.01 to 16.16 mg kg^{-1}, with an estimated arithmetic mean of 5.65 mg kg^{-1}. In another study, Chen and Lee (1995) detailed the As contents of 101 soils in 21 representative agricultural soil profiles in Taiwan ranging from 3.12 to 14.7 mg As kg^{-1} with an arithmetic mean of 8.88 mg kg^{-1}.

The primary objective of this study was to develop maps illustrating the geographic distribution of As in surficial soil horizons of Taiwan using geostatistics and geographic information systems (GIS). We used variography, linear regression, and sensitivity analysis to characterize the spatial variability of soil As based on data from EPA. We used geostatistical interpolation, i.e., punctual kriging, to estimate and map the total soil As content throughout Taiwan.

MATERIALS AND METHODS

Data

Data were derived from the EPA studies described above. The sampling sites for the study are shown in Figure 1. The sampling was conducted from 1983 through 1986. Samples were taken from geographically well-distributed sites at a target interval of 4 km network. Soils were sampled at a depth of 0-15 cm. Total soil As for 878 samples was determined by means of acid digestion, followed by flame atomic absorption spectrophotometry.

Geostatistical Methodology

Geostatistics (Journel and Hijbregts, 1978; ASCE, 1990a, 1990b) consists of a collection of techniques for the analysis of spatially correlated data. Such geostatistical techniques as kriging incorporate the spatial or temporal characteristics of actual data into statistical estimation processes. These techniques can be linear, such as point kriging, ordinary kriging and block kriging.

Geostatistics provides a model of the spatial correlation of data within a statistical framework, including spatial and temporal covariance functions. Not surprisingly, these models are generally referred to as spatial or temporal structures, and are defined in terms of the correlation between any two points separated by either spatial or temporal distances. A great deal of collected

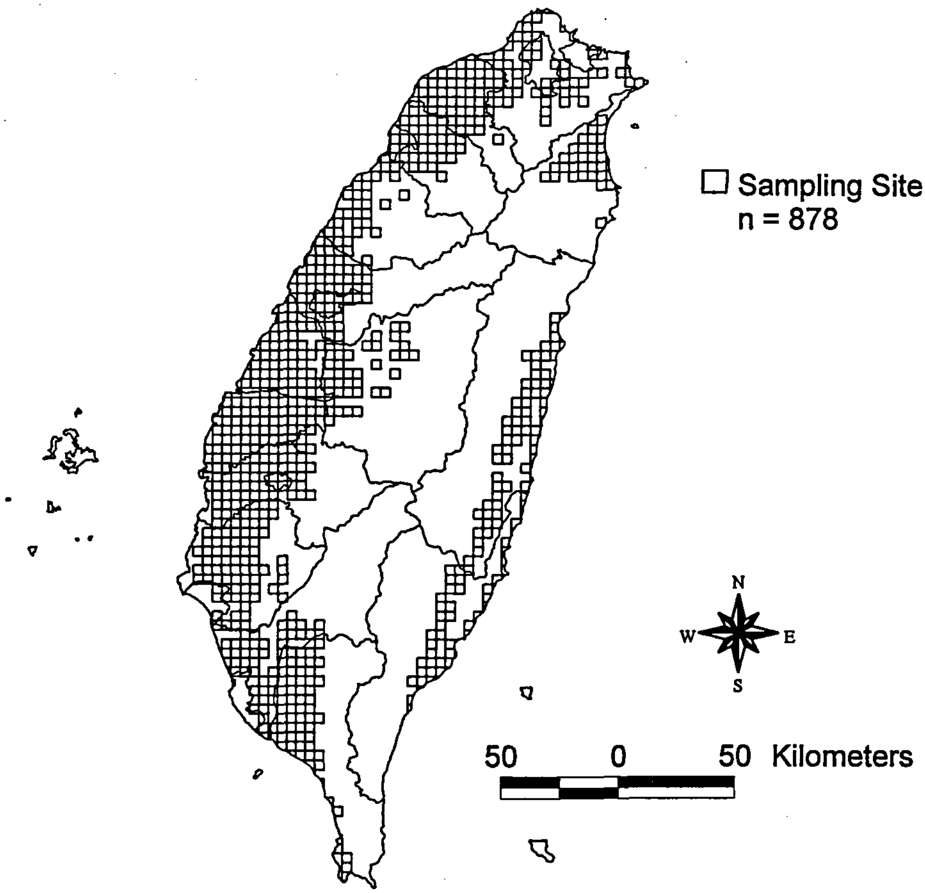

FIGURE 1

General map showing the soil sampling sites in Taiwan.

environmental data indicate that points which are closer in a given direction display higher correlation values than do those that are separated farther.

Geostatistical estimates are calculated as weighted sums of the adjacent sampled concentrations. These weights depend on the exhibited correlation structure. To illustrate, if data appear to be highly continuous in space, those points closer to the estimated points receive higher weights than do those farther away. The criterion for the selection of these weights is minimization of the estimation variance. In this framework, geostatistical estimates may be regarded as most accurate among all linear estimators (i.e., the Best Linear Unbiased Estimator).

<u>Geostatistical Structural Analyses</u>

The first task in any geostatistical investigation is to identify the variogram of the investigated variable in space or time. This task, referred to as variography, is usually performed by determining the estimated variogram of the data collected. Variography is initiated by grouping the available pair-values into a number of lags or distance classes in accordance with their in-between distances. Variograms provide a means of quantifying the commonly observed relationship where samples close together tend to have more similar values than samples farther apart. The variogram $\gamma(h)$ is defined as

$$\gamma(h)\ \frac{1}{2}Var[Z(x)-Z(x+h)] \qquad (1)$$

where (h) is the lag distance separating pairs of point sand Var is the variance of the argument. $Z(x)$ is the value of the regionalized variable of interest at location x , and $Z(x+h)$ is the value at the location $x+h$. An experimental variogram $\gamma^{*}(h)$, is given by

$$\gamma^{*}(h)\ [1/2n(h)]\ \sum_{i=1}^{n(h)}[Z(x_i+h)-Z(x_i)]^2 \qquad (2)$$

where $n(h)$ is the number of pairs separated by the lag distance h.

The main features of a typical variogram are three-fold: the (1)range, (2)sill, and (3)nugget effect. Range is the distance at which the variogram reaches its maximum value. A pair of samples whose in-between distance is greater than the range is uncorrelated. This means that the range is regarded as a measure of the spatial continuity of the investigated variable. Sill, as the upper limit of the variogram which tends to level off at large distances, is a measure of the population variability of the investigated variable; generally, the higher the sill, the greater the variability in the population. The nugget effect is exhibited by the apparent jump in the variogram at the origin, a phenomenon which may be attributed to the small-scale variability of the investigated process and/or to measurement errors. The most accurate estimations can be determined if the investigated variable is well structured. Such a variable will have a variogram with a long range (i.e., high continuity), low sill values (i.e., small population variance) and a small nugget effect (i.e., insignificant small-scale variabilities or measurement errors).

Exploratory data analysis by linear regression were done using STATISTICA, Windows Version (StatSoft, 1994). Geostatistical analysis was done using GS+/386 Version 1.0 (Gamma Design Software, 1995). We calculated variograms at lag intervals ranging from 300 m to 50 km across the full extent of the data (418.9km). Punctual kriging was carried out into a rectangular 106 column by 211 row grid consisting of 22366 2 by 2 km cells. 64 neighboring data points were used to krige each cell. Kriged estimates and estimate standard deviations were transferred to Microsoft Excel 97 (Microsoft, 1997), sorted to conform with the vector display format of the GIS, Arc/Info (ESRI, 1993), and transferred to Arcview 3.0 (ESRI, 1997) for analysis, reclassing, and display.

RESULTS AND DISCUSSION

Descriptive statistics and frequency distribution histograms for the data sets are shown in Table 1 and Figure 2. The kriged estimates of total soil As are shown on the interpolated map (Figure 3).

Variography

A global experimental variogram calculated at a representative lag interval of 10 km across the full extent (418.9 km) of the data (n=878) is shown in Figure 4. At short to moderate lags, i.e., 10 to 300 km, the variogram was relatively stable. The variance increased from a nugget variance of approximately 4.05 mg^2 As kg^{-2} to approach a sill of about 12.3 mg^2 As kg^{-2}, which remained relatively constant up to a lag distance of about 300 km. Beyond 300 km, the variance increased and then became extremely large. Instability in the variogram at long lag distances has been noted by others (Armstrong, 1984), and as a result, variograms are typically calculated only to one-half the maximum distance between points (Englund and Sparks, 1991).

TABLE 1
Statistical summary of soil As data set and kriged estimates for conterminous Taiwan.

	n	min	max	media	mean	Std	KS-c	KS	KS/KS-c	Skewness	kurtosis
		———— mg As kg^{-1} ————									
original	878	0.01	16.16	4.61	5.65	3.21	0.027	0.073	2.704	0.177	-0.679
kriged	8891	1.01	9.34	4.67	4.77	1.94	0.009	0.051	5.667	0.154	-0.886

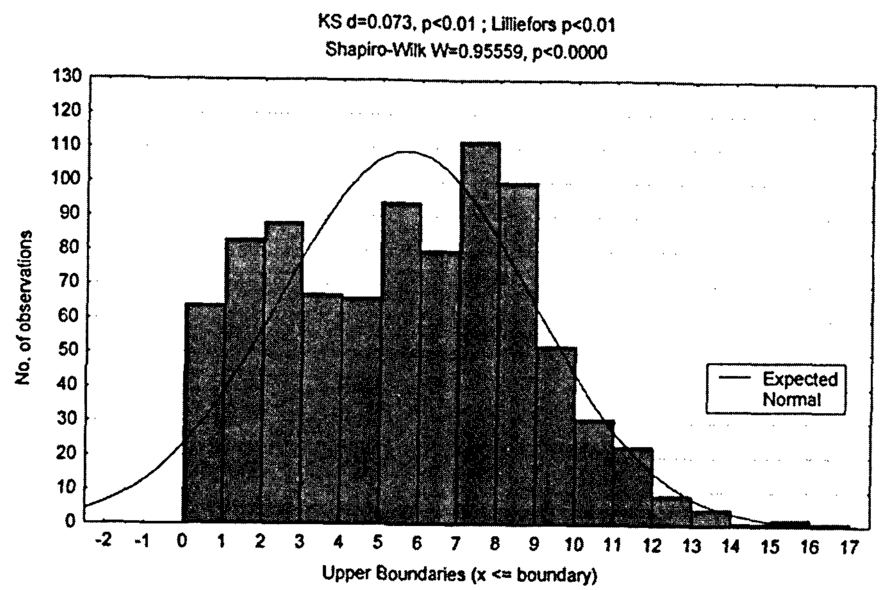

FIGURE 2
Frequency distributions of soil As for conterminous Taiwan.

Experimental variograms were calculated for the data at a variety of lag intervals to a maximum lag of 300 km. Least squares model fitted of these variograms generated a relatively consistent set of best-fit models. Parameters from representative models are shown in Table 2. An exponential model always had the best fit compared with the other models (linear, linear with sill, spherical, or Gaussian) available in the software. The exponential model is of the form:

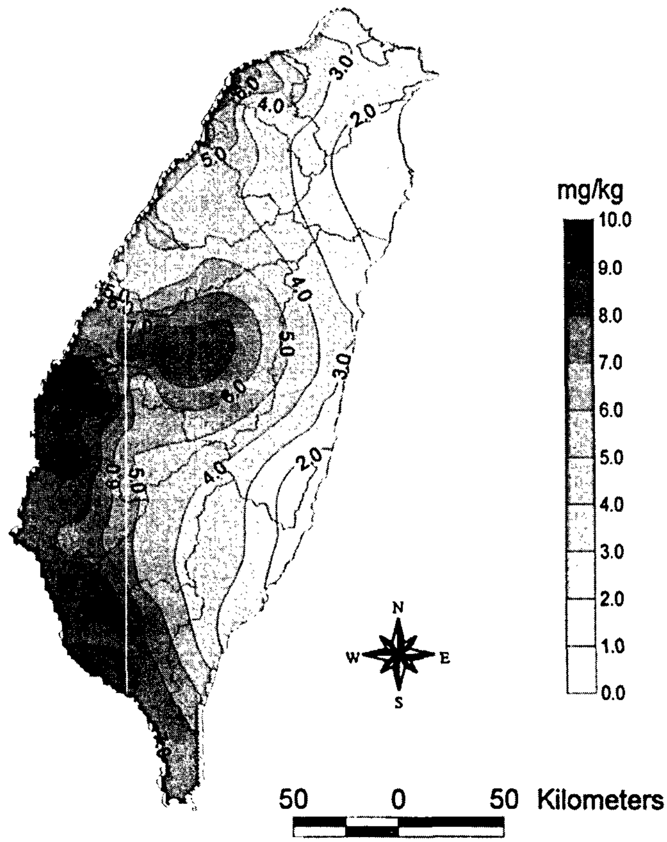

FIGURE 3
Map of the Kriged estimates of the total soil As for conterminous Taiwan.

$$\gamma(h) = C_0 + C[1 - \exp(-h / A_0)] \qquad (3)$$

where C_0 is the nugget variance($C_0$0), C is the structural variance (C C_0), and A_0 is a range parameter, different from the true range in a spherical or linear-sill model. Range in the exponential variogram model is usually assumed to be the point at which the model attains about 95% of the sill ($C + C_0$), which can be estimated as 3 A_0.

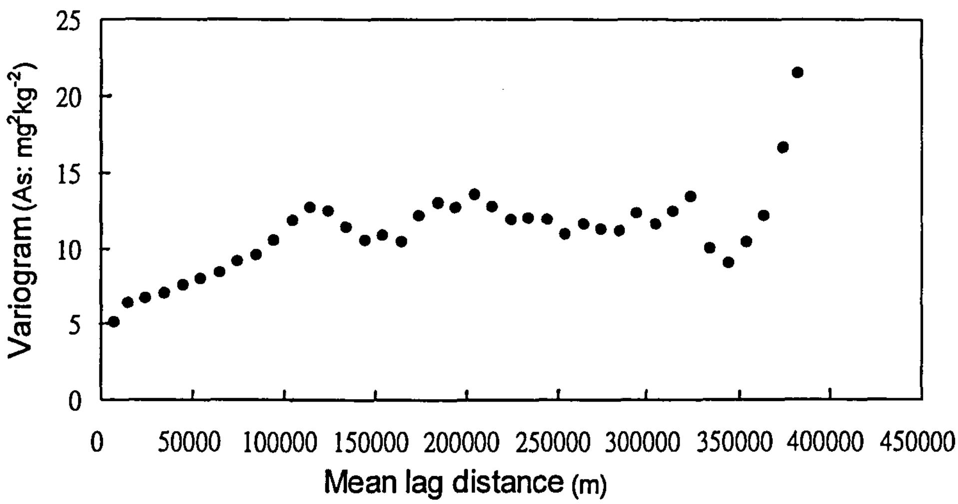

FIGURE 4
Global experimental variogram of total soil As calculated using a lag interval of 10 km across the full extent of the data.

The modeled nugget variance ranged from 2.83 to 4.79 mg^2 As kg^{-2}, corresponding to a standard deviation of from 2.01 to 2.83 mg As kg^{-1}. The experimental error in As determinations was usually less than this, suggesting the presence of spatial variability at lags smaller than those in the data set. The soil element content exhibited different levels of spatial variability at different scales, and the variogram across the extent probably contained localized spatial variation nested within it (Berndtsson et al., 1993). The modeled sill was very consistent, ranging from 12.25 to 12.31 mg^2 As kg^{-2}, slightly less than the sample standard deviation of 3.21 mg As kg^{-1}. The range estimated from the model range parameter A_0 varied from 186 to 196.2 km, beyond which the soil As exhibited no significant spatial correlation. The best-fit model of variogram is shown in Figure 5. Others have analyzed a variety of soil parameters, such as pH, saturated electrical conductivity, sand content, sorbed P, Na, and organic C, and found them to have spatial correlation ranges from centimeters to tens of kilometers (Warrick et al., 1986).

TABLE 2

Least squares best-fit variogram model parameters at various lag intervals. In all cases, the best fit was an exponential model: $\gamma(h) = C_0 + C[1 - \exp(-h/A_0)]$, where $\gamma(h)$ = variance at lag interval h, C_0 = nugget variance, C =structural variance, and A_0 = range parameter

Range (km)	Lag interval (h) (km)	Nugget (C_0) mg^2 As kg^{-2}	Sill (C_0+C) mg^2 As kg^{-2}	Range[a] parameter (A_0) (km)	Estimated[b] range ($3A_0$) (km)	Model reduced sum of squares	Model R^2
300	0.3	4.01	12.25	63.2	189.6	2327.96	0.640
200	0.3	4.11	12.75	73.0	216.0	1174.11	0.736
100	0.3	4.47	11.93	74.0	222.0	245.17	0.774
300	5.0	4.07	12.31	65.4	196.2	45.07	0.848
200	5.0	4.32	13.11	82.9	248.7	27.55	0.877
100	5.0	4.79	14.33	122.1	366.3	1.87	0.962
300	10.0[c]	4.05	12.30	65.0	195.0	21.04	0.855
300	20.0	4.06	12.29	64.8	194.4	8.621	0.874
300	30.0	3.79	12.25	62.0	186.0	4.727	0.890
418.9	40.0	3.69	12.37	63.5	190.5	4.278	0.890
418.9	50.0	2.83	11.99	52.4	157.2	1.846	0.929

[a] A_0 in the exponential model is not the range but a parameter indicative of the range.

[b] Range in the exponential variogram model is usually assumed to be the point at which the model attains about 95% of the sill ($C + C_0$), which can be estimated as $3A_0$.

[c] Parameters of the adjusted least squares best-fit model at a lag interval of 10 km, with C_0 and A_0 adjusted to provide better fit at the smallest lags. This is the model used for kriging.

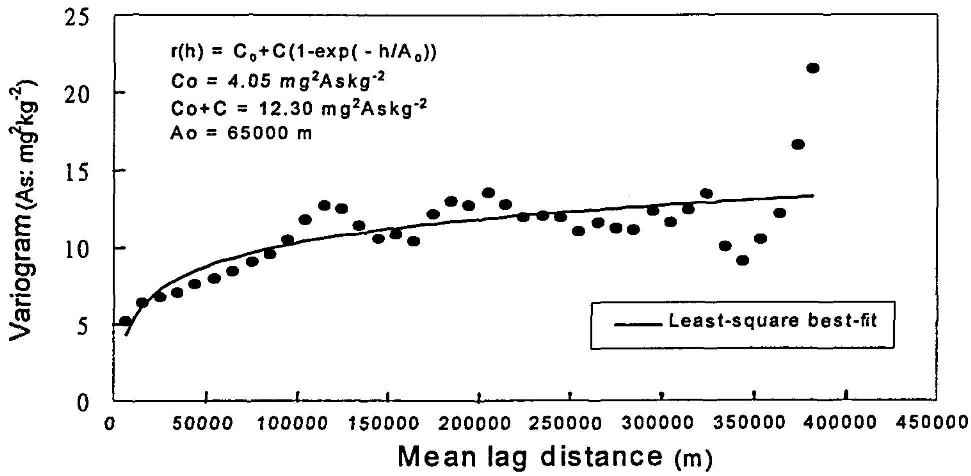

FIGURE 5
Global variogram of total soil As calculated using a lag interval of 10 km and a maximum lag distance of 40 km, and using the least squares best-fit variogram model.

Interpolation via Kriging

Kriging from several of the variogram models shown in Table 2 produced only minor differences in the results, so we only present here results obtained with a representative lag interval of 10 km. The least squares best-fit model for the variogram calculated with a 10 km lag interval is shown superimposed on the variogram in Figure 5; model parameters appear in Table 2. Kriging is particularly sensitive to model parameters at the shortest lags, but a least squares best-fit model may not fit the variogram well at the smallest lag intervals (Journel and Huijbregts, 1978; Armstrong, 1984). To obtain a better fit for the nugget variance and the initial portion of the variogram, the best-fit model nugget variance (C_0) and range parameter (A_0) were adjusted slightly. Weighted least squares methods giving greater emphasis to points at smaller lags are an alternate approach (Cressie, 1985). The adjusted model is shown superimposed on the variogram in Figure 5, and the adjusted model parameters are listed in Table 2.

Kriging from the adjusted least squares best-fit model (Figure 5) and classifying the results by means of deciles produced the map of total soil As shown in Figure 3. We chose decile classification because it provides a geographical representation of the frequency distribution of the kriged estimates. No other statistical differentiation between these classes is meant, nor should it be inferred. Descriptive statistics for the 22366 As estimates included in this map are shown in Table 1, and the estimated standard deviations of the kriged estimates are mapped in Figure 6. The mean total soil As for Taiwan estimated from kriging was similar to that from the original data, 5.65 vs. 4.77 mg As kg^{-1}, respectively. The kriged estimates had a smaller range than the original data, with both the minimum and maximum truncated slightly. Kriging also normalized the data somewhat, reducing skewness and kurtosis compared with the original data.

The kriging estimated standard deviations of As ranged from 2.01 to 2.83 mg As kg^{-1}, with a mean of 2.42 mg As kg^{-1}. These estimated standard deviations reflect both the variability of soil As, as indicated by the standard deviation of the original data (3.21 mg As kg^{-1}), and the uncertainty inherent in interpolating from widely dispersed sites. The nugget variance comprises both of these sources of variability, and represents, in effect, the minimum variance that can be expected for a kriged estimate. Expressed as a standard deviation, the modeled nugget variance was equivalent to 4.05 mg As kg^{-1}. In comparison, the estimated standard deviations for the kriged estimates are quite reasonable. The highest estimated standard deviations (Figure 6) occurred where data were sparsest (Figure 1). This is a natural consequence of the fact that variances estimated from kriging depend only on the variogram model and the sample site configuration; they are not a direct measure of the variance of the neighboring data used to estimate an unsampled point. Acquiring additional data in undersampled areas would reduce estimation variance.

The kriged map of total soil As indicated generally higher than average As in southwest Taiwan. Soil parent material mineralogy may be a predominant factor affecting the total As content of the soils.

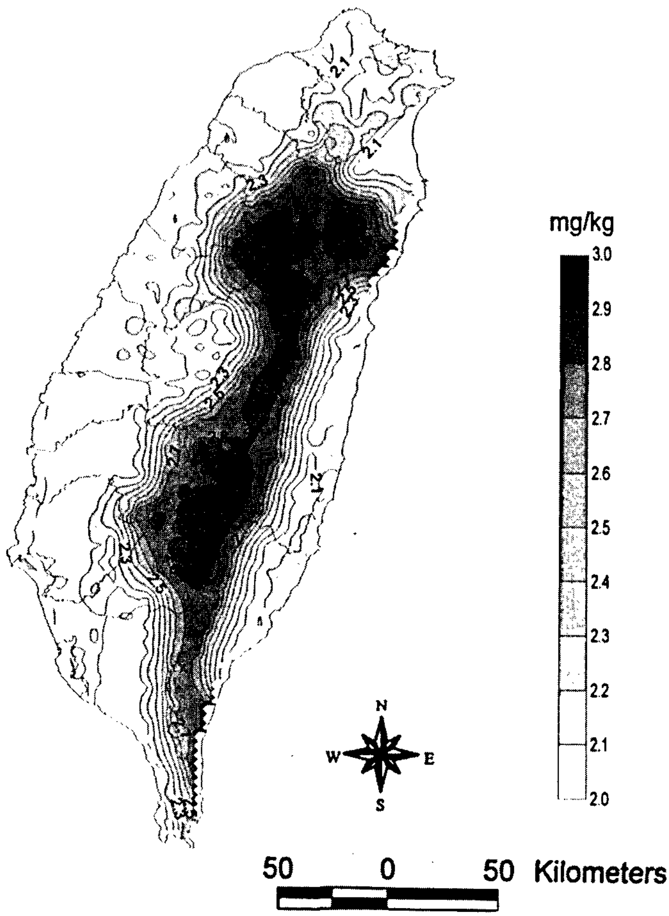

FIGURE 6
Map of the estimated standard deviations of the kriged estimates of total soil As for conterminous Taiwan.

Kriging is a statistically based interpolation method that has proven effective in mining, soil analysis, ecology, and other disciplines. Kriging is considered to be very sensitive to violations of the basic assumptions, yet it has proven very robust in many circumstances. Simple sensitivity analysis was useful here in judging the effects of anisotropy estimation via kriging. Other studies have shown that ordinary kriging following careful exploratory statistical analysis and proper variography can yield reasonable results despite violations of the assumptions. Statisticians are working to develop more robust geostatistical methods to handle difficult data (Henley, 1981; Armstrong, 1984; Cressie, 1984; Rivoirard, 1994).

Geostatistical analysis of total soil As data from 878 sites across Taiwan produced an interpretable exponential variogram (Figure 5). Soil As showed spatial correlation at distances up to 195.0 km, a much greater range than reported for other soil properties. Interpolation by kriging produced digital maps showing the geographic distribution of soil As (Figure 3) and estimates of the standard deviations of the interpolated values (Figure 6). Estimation of the variance associated with interpolation is a major benefit of kriging, and one that other interpolators lack. The estimated standard deviations for soil As were high relative to the estimates themselves, emphasizing the uncertainty inherent in interpolating across large areas from variable and widely scattered data. The standard deviation map will be useful in guiding additional sampling or data acquisition to improve map accuracy and precision. Digital maps facilitate such improvements, allow analysis and manipulation in a GIS, and permit electronic data transfer.

Readers are cautioned to consider the nature of the original data, their extent and spatial variability, the kriging resolution, and the estimated standard deviations when using and interpreting these maps. There are biases inherent in the data set used. The EPA study sampled agricultural soils where As content may have been reduced by crop harvesting or increased by the addition of As-containing pesticides. Some soils exhibit considerable variation in the total As content in different horizons within the profile (Chen and Lee, 1995). Further, we emphasize that the maps delineate gross trends in total soil As on the scale of conterminous Taiwan. Small-scale interpolated maps show broad, average trends, and may misrepresent them as continuous or as having greater precision than the data allow. In reality, local situations may be much more complex, e.g., where parent materials change abruptly within relatively small distances. A global variogram determined across a large and diverse land area may be inadequate to represent regional or localized spatial variability. Regional analyses of spatial variability, with regions based perhaps on soil parent material, may provide more accurate and precise estimates of total soil As. The magnitude of the standard deviations emphasizes the uncertainty inherent in interpolating from variable, widely spaced data. The map of estimated standard deviations indicates clearly where an investment in obtaining more data would reduce estimation variances and improve the map.

CONCLUSIONS

Kriging is a statistically based interpolation method that has proven effective in mining, soil analysis, ecology, and other disciplines. This study has shown that ordinary kriging following careful exploratory statistical analysis and proper variography can yield reasonable results. The kriged map of total soil As indicated generally higher than average As in southwest Taiwan. Soil parent material mineralogy may be a predominant factor influencing the total As content of soils. Geostatistics and GIS are essential tools for analyzing georeferenced information and advancing our understanding of spatial variability at various scales. Geostatistics and GIS will be indispensable in characterizing and summarizing this information to provide quantitative support to decision and policy making for agriculture, health and natural resource management.

ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council, Republic of China, for financially supporting this research under Contract No. NSC-87-2621-P-002-012 .We also thank the EPA-ROC for providing the data.

REFERENCES

Armstrong, M. Improving the estimation and modeling of the Variogram. In G. Verly et al.(ed.).Geostatistics for natural resource characterization. Proc. NATO Advanced Study Institute on Geostatistics for Natural Resources Characterization, South Lake Tahoe, CA.6-17 Sept. 1983Dordrecht, the Netherlands (1984), pp.1-19.

ASCE Task Committee on Geostatistical Techniques in Geohydrology(S. Rouhani, Chairman and Principle Author), 116(5), 612-632 (1990a).

ASCE Task Committee on Geostatistical Techniques in Geohydrology(S. Rouhani, Chairman and Principle Author), 116(5), 633-658 (1990b).

Berndtsson, R., Bahri, A. and Jinno, K., Soil Sci. Soc. Am. J., 57, 1323-1329 (1993).

Bierkens, M.F.P. and Burrough, P.A., J. Soil Sci., 44, 369-381 (1993).

Bourennane, H., King, D., Chery, P. and Bruand, A., Eur. J. Soil Sci., 47, 473-483 (1996).

Burgess, T.M. and Webster, R., J. Soil Sci., 31, 315-331 (1980).

Chen, Z.S. and Lee, D.Y., Journal of the Chinese Institute of Environmental Engineering, Vol. 5, No.3, 205-211 (1995).

Cressie, N. Math. Geol. 17, 563-586 (1985).

Davies, B.E., Water Air Soil Poll., 94, 99-107 (1997).

Englund, E., and Sparks, A.GEO-EAS, geostatistical environmental assessment software, user's guideUSEPA Rep.600/4-88/033.ENMSL, Environmental Protection Agency, Las Vegas, NV.10, 11-13 (1988).

Environmental Systems Research InstitutePC ARC/INFO version 3.4D ESRI, Redlands, CA (1993) .

Environmental Systems Research InstituteArcview 3.0 ESRI, Redlands, CA (1997).

Gamma DesignGS+: Geostatistics for the agronomic and biological sciences, Version 2.3, Gamma Design , Plainwell, MI (1994).

Henly, S.Nonparametric geostatisticsJohn Wiley & Sons, New York (1981).

Hung, T.C. and Liao, S.M., Toxicological and Environmental Chemistry, 56(1-4) , 63-73 (1996).

Journel, A.G. and Huikbregts, C.J.Mining geostatisticsAcademic Press, London (1978).

Microsoft CorporationMicrosoft Office 97 Chinese Professional Edition, Microsoft Excel 97 Microsoft Taiwan Corp, Taipei (1997).

Rivoirard, J.Introduction to disjunctive kriging and non-linear geostatistics Oxford Univ. Press, New York (1994).

Smith, J.F., Halvorson, J.J. and Papendick, R.I., Soil Sci. Soc. Am. J., 57, 743-749 (1993).

StatSoft, Inc.STATISTICA for WindowsStatSoft, Inc., Tulsa OK (1994).

Warrick, A.W., Meyers, D.E. and Nielsen, D.R.. Geostatistical methods applied to soil science In A.Klute (ed.) Methods of soil analysis, Part 1. 2nd ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI (1986), pp.53-82.

White, J.G., Welch, R.M. and Norvell, W.A., Soil Sci. Soc. Am. J. 61:185-194 (1997).

Received: November 20, 1998