農業工程學報　第 52 卷第 2 期　　Journal of Chinese Agricultural Engineering
中華民國 95 年 6 月出版　　Vol. 52, No. 2, June 2006

比較基線法與有限混合分佈模式對台灣農田土壤中銅濃度的分級

Comparing the Classification of Copper Concentrations in Taiwan Farmland Soils by Baseline Method and Finite Mixture Distribution Model

經濟部水利署水利規劃試驗所
工程員

楊 松 岳*

Song-Yue Yang

國立台灣大學生物環境系統工程學系
教授

張 文 亮

Wen-Lian Chang

摘　要

基線法(Baseline Method, BM)通常用來描述土壤中重金屬的背景濃度，但是由於污染物質的移動，使取得未受人爲干擾與影響的土壤樣本變得非常困難，特別是在都會與工業區。本研究的目的是以有限混合分佈模式(Finite Mixture Distribution Model, FMDM)來分級全台 918 個農田土壤樣本中銅的濃度並與傳統的基線法比較。本研究除了求出 FMDM 模式的參數與分級外，並進一步的計算每一種土壤被歸類在每一等級的比例。由這些結果計算出每一等級的擊中率(hit rate)，並以數值的高低來判斷所代表的意義。與 BM 法相較，FMDM 更適合描述重金屬銅在土壤中的分佈，而且模式的適合度是可以評估的，而 BM 則無法評估。此外，FMDM 可以有效的解決 BM 在取樣過程中難以判斷取樣點是否有被污染的困難。特別是從環境風險評估的觀點來看的時候，當受人爲影響與污染土壤在取樣中被包含進來時，BM 會高估污染土壤的界定值導致低估了污染土壤的比例，而 FMDM 能有效的將大部分污染土壤從土壤樣本中分離出來。

關鍵詞：土壤污染，銅，背景值，有限混合分布。

ABSTRACT

The baseline method (BM) is typically adopted to determine the background concentration of heavy metals in soils. The long-range transport of contaminants makes the sampling of undisturbed or unaffected soil difficult, thus making it difficult to determine the background concentration, especially in urban and industrial areas. This

*通訊作者，經濟部水利署水利規劃試驗所，413 台中縣霧峰鄉吉峰村中正路 1340 號，acton@ms14.url.com.tw

study adopted the finite mixture distribution model (FMDM) not only to classify Cu concentrations based on 918 representative Taiwan farmland soil samples, but also to compare the results with those obtained using a traditional BM. The parameters and classes of FMDM were obtained; the proportion of each soil in each class, and the hit rate of each class were calculated. The results showed that FMDM was more effective than BM in describing Cu distribution in soils, and the goodness of fit of FMDM could be evaluated, but that of BM could not. In addition, FMDM effectively overcame the difficulty of defining the background site. In particular, from the perspective of environmental risk assessment, BM underestimates the proportion of contaminated soil included during sampling; however, FMDM can effectively separate most of the contaminant soils from soil samples.

Keywords: Soil pollution, Copper, Background value, Finite mixture distribution.

INTRODUCITON

Copper (Cu) is an essential element in plants in which Cu functions as a member a prosthetic groups of enzyme systems and as a facultative activator of enzyme systems (Alloway, 1995). Cu exhibits a great capacity to interact chemically with mineral and organic components of soil, and to readily precipitate with various anions such as sulfide, carbonate and hydroxide (Kabata-Pendias and Pendias, 2001).

The immobile characteristic of soil profiles is such that Cu has great affinity to accumulate in surface soils. The regularity of the distribution of Cu in soils is governed by two main factors - the parent material and the soil formation processes. While this factors the initial distribution of Cu in soil, bioaccumulation and anthropogenic sources most strongly affect the concentration of Cu in surface soils. The utilization of Cu-containing material and agricultural or industrial emissions contaminate soil with Cu. The corrosion of Cu alloy construction materials causes the local or incidental input of Cu into soils, and metal smelters contribute to the global long-distance pollution of the atmosphere (Kabata-Pendias and Pendias, 2001). Also, the application of sewage sludges, municipal composts, pig and poultry slurries to agricultural land causes the deliberate non-point pollution of agricultural soils. Such practices are a greater threat than point source pollution to sustained food and fiber production, because their extent and impact are greater but less evident until plant growth becomes visually retarded and exhibits symptoms of metal toxicity (Alloway, 1995).

The background values of elements are critical to regulatory agencies in estimating the extent of pollution and assessing risk. In some contexts, the full description of the term “background concentration” could be “natural background concentration”, which refers to the soil that is naturally present at sites that have not been disturbed or affected by human activities (Gough, 1993; Portier, 2001). However, finding such a site in a developed area is difficult, so indirect estimates of concentrations have been made, such as in remote areas, and in ancient or historical materials (Fergusson, 1990). Another type of “background concentration” called “anthropogenic background” is the concentration typically observed in a region, resulting from human activities but not associated with a specific contamination activity. The Cu concentration in general Taiwanese farmlands soil belongs to the latter type.

The anthropogenic soils that form as a result of various human activities transport small quantities of chemicals and deposit them thinly over a broad area. The concentrations of elements in such soils are expected to exceed those in background soils. The contaminant soils that are generated by the point source releases of contaminants are geographically constrained in scope and exhibit intense enhancement (Portier, 2001).

In recent years, some methodologies have been proposed to determine the thresholds for determining the concentrations of heavy metals in soils. The baseline concentration, expressing an expected range of concentrations around a mean, is usually used to describe elemental concentrations in soils (Gough et al., 1988; Dudka, 1993; Gough, 1993). However, certifying a background site as an anthropogenic one is difficult, because the activities that cause anthropogenic enhancement cannot be easily identified. A statistical approach, "finite mixture distributions", has in recent years been developed and discussed (Portier, 2001; Yang and Chang, 2005).

A frequency distribution that comprises more than one component distribution frequency is defined as a finite mixture distribution (Everitt, 1996). The degree of the contamination of an element in soil can be specified as a probability distribution function with three parts – background, anthropogenic and contaminant soils (Portier, 2001). Yang and Chang (2005) applied the finite mixture distribution model to determine the concentration of cadmium in Taiwan farmland soil. This study classified Cu concentrations in Taiwan farmland soils using the baseline method (BM) and the finite mixture distribution model (FMDM) and studied whether these classifications can effectively distinguish anthropogenic and contaminant soils from background soil.

MATERIALS AND METHODS

The database of Cu concentrations herein was established by the Environmental Protection Agency, Executive Yuan (1989). The database contains 918 geographically representative 1600 hectare farmland plots in Taiwan. The soil was extracted using 0.1N HCl solution. The metal concentration in the extracted solution was analyzed using a Perkin-Elmer 305A atomic absorption spectrophotometer.

THEORY AND MODELS

Two methods were adopted to classify the Cu concentrations in this study.

Baseline Method (BM)

The baseline method was applied to the samples as a whole, regardless of background or whether the anthropogenic or contaminant component distribution of the sample is of interest. The central tendency and variation of data were expressed as geometric mean (GM) and geometric standard deviation (GSD). The range of baseline concentrations, which included 95% of the sample population, was between GM/GSD^2 and $GM \times GSD^2$ (Dudka et al., 1995; Chen et al., 1999). The upper baseline concentration, defined as 97.5% of the top of the log-normal distribution, was used as a background value to evaluate soil samples to determine possible contamination by heavy elements (Chen et. al, 2002).

Finite Mixture Distribution Model (FMDM)

According to Yang and Chang (2005), the probability density function (p.d.f.) of Cu concentration in soil has three components and can be expressed as Eq. (1).

$$p(x|\psi) = \pi_1 f_1(x|\theta_1) + \pi_2 f_2(x|\theta_2) + (1-\pi_1-\pi_2) f_3(x|\theta_3) \quad \text{.....(1)}$$

where π_1, π_2 and π_3 are the relative abundances of components as proportions of the total population, and must satisfy the constraints $0 \le \pi_1,\ \pi_2 \le 1$. All of the components, f_1(x), f_2(x) and f_3(x), follow two-parameter log-normal distributions (Ames and Prych, 1995; Ames and Hawkins, 1997); θ_1, θ_2 and θ_3 represent background, anthropogenic and contaminant soils, respectively. Then, two points of intersection, X_1 and X_2, of the three functions, f_1(x), f_2(x) and f_3(x), are given by Eqs. (2) and (3).

$$\pi_1 f_1(X_1|\theta_1) = \pi_2 f_2(X_1|\theta_2) \quad \text{.....(2)}$$

$$\pi_2 f_2(X_2|\theta_2) = \pi_3 f_3(X_2|\theta_3) \quad \text{.....(3)}$$

X_1 is defined as the threshold of Anthropogenic Class, and X_2 is defined as the threshold of Contaminant Class.

The relative proportion F_{ij} of soil type i classified as soil Class j can be calculated using Eq. (4).

$$F_{ij}(x|\theta_i) = \pi_i \left[F_i(X_j|\theta_i) - F_i(X_{j-1}|\theta_i)\right] \quad \text{.....(4)}$$

i = 1, 2, 3; j = 1, 2, 3

where

$$F_i(x|\theta_i) = \frac{1}{2}\left[1 + erf\left(\frac{\ln x - \mu_y}{\sigma_y \sqrt{2}}\right)\right]$$

$x = \exp(y)$

In this study, $X_0 = 0$ and $X_3 = \infty$ are set. The hit rate of Class j is given by Eq. (5) (Chow, 2002).

$$h_j = \frac{F_{jj}(x)}{\sum_{i=1}^{3} F_{ij}(x)} \quad \text{.....(5)}$$

j = 1, 2, 3

where F_{jj} is the proportion of soil type j classified in the same soil Class j. The likelihood-ratio χ^2 test is given by Eq. (6), to evaluate the goodness of fit:

$$\chi^2 = \sum_{j=1}^{m} \frac{(O_j - E_j)^2}{E_j} \quad \text{.....(6)}$$

χ^2 has (m-k-1) degrees of freedom, where m is the number of classes of the original population and k is the number of estimated parameters (Krebs, 1999; Liu et al., 2002). In this study, a software package called "Rmix" was used to implement the Maximum Likelihood approach to estimate the parameters and perform the χ^2 test of the model (Du, 2002).

RESULTS AND DISCUSSION

For BM, GM and GSD yielded Cu concentrations in soil of 4.05 and 0.33 mg/kg, and the threshold for the Contaminant Class was obtained as 37.19 mg/kg. Figure 1 presents the observed probability distribution and Classes of BM.

For FMDM, the soil samples were classified into three parts according the degree of Cu contamination - background, anthropogenic and contaminant with respective proportions of 87.78%, 10.66% and 1.56%; arithmetic mean Cu concentrations of 5.77, 12.06 and 27.47 mg/kg, and arithmetic standard deviations of 4.21, 2.75 and 0.34 mg/kg. Equations (2) and (3) yield the thresholds of 12.00 and 26.29 mg/kg for Anthropogenic and Contaminant Classes, respectively. Figure 2 plots the observed and predicted probability distributions and Classes.

For the goodness of fit, the χ^2 value and the degrees of freedom were 4.48 and 9, respectively. Then, the p-value of FMDM was 0.8894, exceeding 0.05, which indicated that FMDM did not differ significantly from the sample distribution and was appropriate in representing the Cu concentration in farmland soils in Taiwan.

Nine hundred and eighteen samples were classified by applying the thresholds of BM and FMDM. Table 1 presents the results. BM yielded

Tab. 1. The numbers of samples and proportions of Background, Anthropogenic, and Contaminant Classes by BM and FMDM

	Method	Background Class	Anthropogenic Class	Contaminant Class	Total
BM	Number of samples	908	*	10	918
	Proportion (%)	98.91	*	1.09	100.00
FMDM	Number of samples	797	104	17	918
	Proportion (%)	86.82	11.33	1.85	100.00

* There was no Anthropogenic Class in BM.

Tab. 2. The calculated proportions of background, anthropogenic, and contaminant soils in Background, Anthropogenic, and Contaminant Classes

	Background Class (%)	Anthropogenic Class (%)	Contaminant Class (%)	π (%)
background soil	81.29	6.12	0.37	87.78
anthropogenic soil	5.71	4.95	0	10.66
contaminant soil	0	0	1.56	1.56
Total	87.00	11.07	1.93	100.00

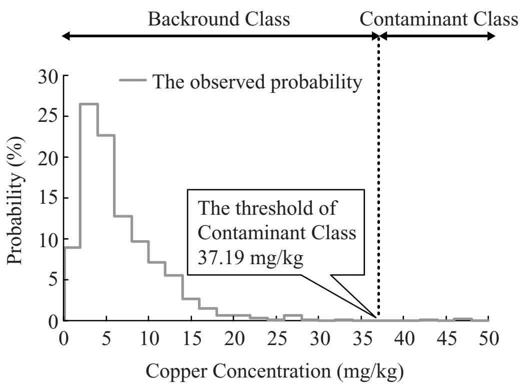

Fig. 1. Observed distribution and threshold of BM

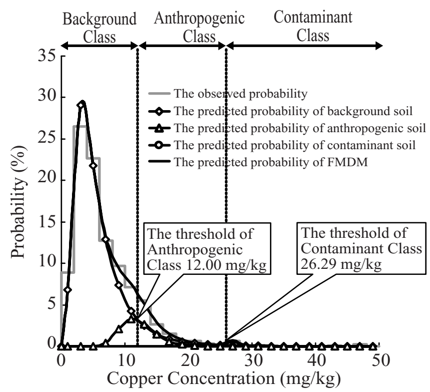

Fig. 2. Observed and predicted probability distribution of FMDM

10 and 908 samples in the Background and Contaminant Classes, respectively. FMDM yielded 17, 104 and 797samples in Background, Anthropogenic and Contaminant Classes, respectively.

Table 2 reveals that the proportion of soil samples classified in the Contaminant Class by BM was 1.09%, which was lower than that, 1.56%, determined by FMDM. Almost one-third of the contaminant soil was misclassified as background, because when the contaminant soils were included in the soil samples, GM and GSD were overestimated, so the BM was overestimated and the proportion classified in the Contaminant Class was underestimated.

According to Table 2, the proportions of Background, Anthropogenic and Contaminant Classes, according to FMDM, were 87.00%, 11.07% and 1.93%, respectively. Table 3 presents the proportions of soils in the various classes, according to Eq. (4). The calculated proportions of Background, Anthropogenic and Contaminant Classes were 86.82%, 11.33% and 1.85%, respectively, which data were tallied with the former results.

According to Table 3, the hit rates of Background, Anthropogenic and Contaminant Classes were separately calculated as 93.43%, 44.70% and 80.90% using Eq. (5). The high hit rate of the Background Class shows that most of the background soils were classified in the Background Class. The hit rate of the Anthropogenic Class was only 44.70%, because the predicted probability distribution distributions of the background and anthropogenic soils were very close to each other in the range 7 to 20 mg/kg. This finding indicated the difficulty of identifying the background site as sampling (Gough, 1993; Portier, 2001). The hit rate of the Contaminant Class reached 80.90%, because some of the background soil was classified in the Contaminant Class. However, the most important result was that all of the contaminant soil was classified in the Contaminant Class. From the perspective of environmental risk assessment, this fact ensures public health.

CONCLUSIONS

This work supports the following conclusions. (i) FMDM was more accurate and flexible than BM in describing the concentration distribution of elements in soil. The goodness of fit of FMDM can be evaluated by the calculated proportion of each soil in each Class and the hit rates, but that of BM cannot. (ii) BM must satisfy the constraint that the sample sites must not have been disturbed or affected by human activities; however, this is a difficult requirement to satisfy for sampling in developed areas. The proportions of background, anthropogenic and contaminant soils could be estimated by FMDM from field samples, even when the soil samples could be mixed from natural and anthropogenic areas because of the difficulty of defining background sites. (iii) From the viewpoint of environmental risk assessment, BM underestimates the proportion of the contaminant soil, when the anthropogenic and contaminant soils were sampled; however, FMDM separates all of the contaminant soils from the samples to ensure public health.

ACKNOWLEDGMENTS

The authors thank Environmental Protection Administration, Executive Yuan for providing the data. We wish to thank the unknown reviewer for helpful suggestions and comments.

REFERENCES

Alloway, B. J., Heavy Metals in Soils, 2nd ed, 368pp, Blackie Academic & Professional, 1995.

Ames, K. C., and D. B. Hawkins, "Statistical analysis and areal trends of background concentration of metals in soil of Clark County, Washington," *USGS Water Resour. Invest. Rep. 96-4252 X*, U.S. Geol. Slirv., Tacoma, WA, 1997.

Ames, K.C., and E.A. Prych, "Background concentrations of metals in soils from selected regions in the Slate of Washington," *USGS Water Resour. Invest. Rep. 95-401 X*, U.S. Geol. Slirv., Tacoma, WA, 1995.

Chen, M., L. Q. Ma, and W. G. Harris, "Baseline concentrations of 15 trace elements in Florida surface soils," *J. Environ. Qual.*, **28**, 1173-1181, 1999.

Chen, M., L. Q. Ma, and W. G. Harris, "Arsenic concentrations in Florida surface: influence of soil type and properties," *Soil Sci. Soc. Am. J.*, **66**, 632-640, 2002.

Chow, Wayne S., Multivariate Statistical Analysis: with Application of SAS/ STAT, 965pp, Best-Wise Press, 2002. (in Chinese)

Du, J., "Combined algorithms for constrained estimation of finite mixture distributions with grouped data and conditional data," 124pp, *M. S.*

Thesis, McMaster University, Hamilton, Canada, 2002.

Dudka, S., "Baseline concentrations of As, Co, Cr, Cu, Ga, Mn, Ni, and Se in surface soils, Poland," *Appl. Geochem*., **2**, 23-28, 1993.

Dudka, S., R. Ponce-Hernandez, and T.C. Hutchinson, "Current levels of total element concentrations in the surface layer of Sudbury's soils," *Sci. Total Environ.*, **162**, 161-172, 1995.

Environmental Protection Administration, Executive Yuan, Republic of China, *The investigation of heavy metal concentration in Taiwan*, Part I~IV, Taipei, 1989 (in Chinese).

Everitt, B. S., "An introduction to finite mixture distributions," *Stat. Meth. Med. Res.*, **5**, 107-127, 1996.

Fergusson, J. E., The Heavy Elements: Chemistry, Environmental Impact and Health Effects, 614pp, Pergamon Press, 1990.

Gough, L. P., "Understanding our fragile environment, lessons from geochemical studies," *USGS Circular 1105*, U. S. Gov. Print. Office, Washington, DC, 1993.

Gough, L. P., R. C. Severson, and H. T. Shacklette, "Element concentrations in soils and other surficial materials of Alaska," *USGS Prof. Pap. 1458*, U. S. Gov. Print. Office, Washington, DC, 1988.

Kabata-Pendias, A., and H. Pendias (ed.), *Trace Elements in Soils and Plants*, 3rd ed., CRC Press, Boca Raton, FL, 2001.

Krebs, C. J., *Ecological Methodology*, 2nd ed., 620 pp, Benjamin Cummings, Menlo Park, California, 1999.

Liu, C. M., L. J. Zhang, C. J. Davis, D. S. Solomon, and J. H. Gove, "A finite mixture model for characterizing the diameter distributions of mixed-species forest stands," *For. Sci*., **48**(4), 653-661, 2002.

Portier, K. M., "Statistical Issues in Assessing Anthropogenic Background for Arsenic," *Environmental Forensics*, **2**, 155-160, 2001.

Yang, S.-Y., and W.-L. Chang, "Use of Finite Mixture Distribution Theory to Determine the Criteria of Cadmium Concentrations in Taiwan Farmland Soils," *Soil Sci.*, **170**(1), 55-62, 2005.

收稿日期：民國 94 年 11 月 1 日
修正日期：民國 95 年 2 月 10 日
接受日期：民國 95 年 2 月 13 日