

Delimitation of arsenic-contaminated groundwater using risk-based indicator approaches around blackfoot disease hyperendemic areas of southern Taiwan

Cheng-shin Jang · Chen-wuing Liu ·
Kuang-liang Lu · Ching-chieh Lin

Received: 18 October 2006 / Accepted: 23 January 2007 / Published online: 25 April 2007
© Springer Science + Business Media B.V. 2007

Abstract This work determined scopes of arsenic (As)-contaminated groundwater using risk-based indicator classification approaches in blackfoot disease hyperendemic areas of southern Taiwan. Indicator kriging was first used to establish a conditional cumulative distribution function at each cell. Three approaches – the p -quantile estimate, the E-type estimate and the minimization of the expected loss – were then adopted to delimit contaminated regions for a regulated standard of As concentrations in groundwater. According to a risk assessment model established in our previous research, the standard was set to 250 $\mu\text{g/l}$ for aquacultural use, corresponding to the 77.1th percentile of observed concentrations. Misclassification risks and uncertainty were examined for the classification approaches. The analyzed results reveal that contaminated areas are the largest using the 0.771-quantile estimate, whereas they are the smallest using the minimization of the expected loss. Proportions of credible polluted areas with low risks to false

positives maintain a constant, 12.9–13.2%, for the classification approaches. To reduce a great impact on human health, As-polluted groundwater should be strictly prohibited to cultivate fish in credible polluted zones and monitored persistently in polluted zones with high risks to false positives.

Keywords Blackfoot disease · Indicator kriging · Misclassification · Uncertainty · Risk

Introduction

Arsenic has been well documented to be a major risk factor for blackfoot disease (BFD) (Chen et al. 1994). Blackfoot disease was once epidemic on the southwestern coast of Taiwan (Tseng 1977). The residents used artesian well water with a high As content for over 50 years. Large-scale investigations on the association between As complexes in well water and age-adjusted mortality from several diseases (Chen et al. 1995; Lai et al. 1994) and cancers (Chen and Wang 1990; Wu et al. 1989) yield mutually supporting findings. According to statistical data, 2,758 persons suffered from BFD in Taiwan until 1997. Tseng (1977) reported that As concentrations in groundwater were 10–1,840 $\mu\text{g/l}$ in these areas and most of them ranged from 400–600 $\mu\text{g/l}$. Jang et al. (2006) used As concentrations of <10–1,470 $\mu\text{g/l}$ in groundwater around BFD hyperendemic areas to

C.-s. Jang
Department of Leisure and Recreation Management,
Kainan University, Luzhu, Taoyuan,
Taiwan 338, Republic of China

C.-w. Liu (✉) · K.-l. Lu · C.-c. Lin
Department of Bioenvironmental Systems Engineering,
National Taiwan University, Taipei,
Taiwan 106, Republic of China
e-mail: lcw@gwater.agec.ntu.edu.tw

spatially analyze potential carcinogenic risks associated with ingesting As in aquacultural tilapia. Groundwater is used abundantly for several utilities and is an alternative to surface water in coastal regions of southwestern Taiwan, where surface water resources are severely deficient. Nowadays, most inhabitants in this region do not drink well water directly since many epidemiological evidences prove that the As exposure is strongly related to the incidence of diseases and cancers. However, very large quantities of groundwater are used to farm fish and shellfish. Abnormal high As concentrations in groundwater indirectly enters the food-chain via various paths and bio-accumulates in humans (Han et al. 1998; Huang et al. 2003).

In Taiwan, monitoring and regulation standards of As in non-drinking groundwater are 250 and 500 $\mu\text{g/l}$, respectively. Using a model of risk assessment (US EPA 2001) and arithmetic or geometric average values of experimental parameters in tilapia documented in Jang et al. (2006), As concentrations of approximately 340 $\mu\text{g/l}$ in groundwater exceed an accepted risk standard (one millionth) via consuming tilapia farmed in the BFD area. Therefore, the As concentration of 250 $\mu\text{g/l}$ in groundwater is suitable for delineating polluted scopes of aquacultural use. The range from 340 to 250 $\mu\text{g/l}$ is regarded as a factor of safety for individual discrepancy.

Spatial distributions of polluted groundwater quality are commonly heterogeneous. However, only a small proportion of in-situ data can be analyzed in a field investigation owing to time and cost constraints. Consequently, sparsely measured data contain considerable uncertainty. Geostatistics is widely used in modeling the spatial variability and distribution of field data with uncertainty. Indicator kriging (IK), which is a frequently employed non-parametric geostatistical method, makes no assumption on distributions of variables, and a 0–1 indicator transformation of data is used to make the predictor robust to outliers. At an unsampled location, the values estimated by IK represent the probability that does not exceed a particular threshold. The indicator-based approaches have been frequently applied to the soil pollution by heavy metals and probabilistically delineated polluted extents (Juang and Lee 1998; Juang et al. 2004). Liu et al. (2004), Goovaerts et al. (2005) and Jang et al. (2006) used the indicator-based approaches to evaluate the As exposure probability in groundwater. Saisana et al. (2004) adopted IK to delineate zones polluted with nitrogen dioxide in air.

Owing to sparsely measured data, various estimates may produce great difference, resulting that wrong management and remediation policies are implemented. Correct delimitation of As-contaminated groundwater is vital for managing groundwater use and sustaining human health. The objective of this work was to determine extents of As-contaminated groundwater using risk-based indicator approaches in BFD hyperendemic areas of southern Taiwan. Indicator kriging was first used to establish a conditional cumulative distribution function at each cell for several thresholds. Three approaches – the p -quantile estimate, the E-type estimate and the minimization of the expected loss – were adopted to delimit polluted regions and to examine likely misclassification and uncertainty.

Study area

Geography and hydrogeology

The study area is located in the plain region of Yunlin, Chiayi and Tainan counties (Fig. 1). This area is enclosed by the Taiwan Strait to the west, the Central Mountain to the east, the Choushui river to the north and the Erhjin river to the south. The northern study area, Yunlin county, is the southern part of the Choushui River alluvial fan, and the central and

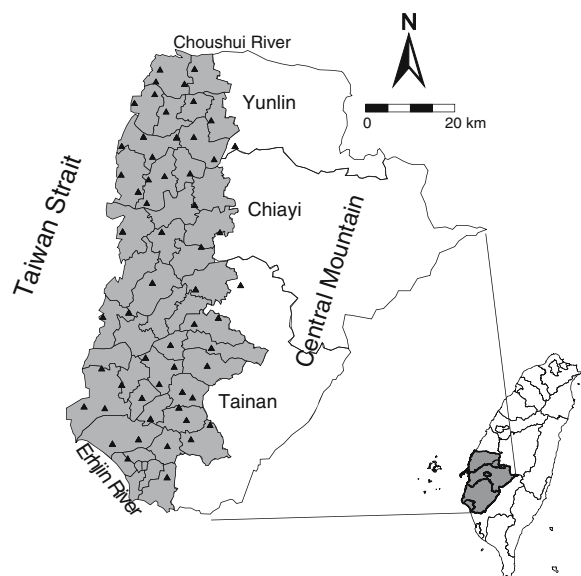


Fig. 1 Study site in southwestern Taiwan. The triangle mark denotes hydrogeological stations

southern study area, Chiayi and Tainan counties, covers the Chianan plain. This region is approximately 2,700 km², and extends 30 km from east to west and 90 km from north to south. The average annual precipitation is around 1,400 mm. Rainfall is concentrated in a five-month period from May to September. Aquaculture and agriculture are the primary source of revenue for the inhabitants in the coastal and plain regions, respectively. A large amount of groundwater has been extracted from aquifers in this area to supply fishponds and irrigation.

Between 1992 and 2001, the Taiwan government implemented a long-term comprehensive groundwater monitoring network plan to explore the hydrogeological characteristics of the main groundwater system in this region. The hydrogeological study demonstrated that the area was formed in the late Quaternary age. Subsurface hydrogeological analysis to a depth of approximately 300 m shows that the formation can be clearly divided into six interlayered sequences in the coastal area of the Choushui River alluvial fan, including three marine sequences and three non-marine sequences (Central Geological Survey 1999). The non-marine sequences with high permeability can be considered to be aquifers. The marine sequences with fine sediments can be regarded as aquitards. However, the formation in the Chianan plain cannot be clearly partitioned into marine and non-marine sequences because of the presence of many small and discontinuous aquitards. Liu et al. (2006) confirmed that the As distributions in groundwater were correlated more highly with those in the clay soil (marine sequence) than those in the sand and gravel soil (non-marine sequence). The marine sequences between the southern part of the Choushui River alluvial fan and the Chianan plain were formed by a similar sedimentation process (Wang et al. 2007). Thus, this study combined the two regions to assess As-polluted extents in groundwater.

The new groundwater-monitoring network in the Choushui River alluvial fan and Chianan plain was quickly implemented. It consists of 56 hydrogeological investigation stations and 168 groundwater-monitoring wells in aquifers of different depths. Most investigation stations have 2–5 monitoring wells, which are located in aquifers. Number of wells at a station depends on number of aquifers. A large number of monitoring wells are distributed in the northern and southern regions. Nevertheless, few

observed data are located in the central region which is the interface between the Choushui River alluvial fan and the Chianan plain. Few aquifers are present in the region, causing few monitoring wells established.

Arsenic concentrations in groundwater

From 2000 to 2001, the Taiwan Sugar Company (2001 and 2002) undertook groundwater quality surveys in the Choushui River alluvial fan and in the Chianan plain. The surveys analyzed 31 water quality items, including the concentrations of As. Abnormally high As content was the only carcinogenic trace element among these water quality items. Figure 2 showed cumulative frequencies of measured 168 As concentrations in this study area. The As concentration of 250 µg/l responded to the 77.1th percentile of measured As concentrations. Jang et al. (2006) thoroughly explored statistical features of these As concentrations.

Materials and methods

Variogram analysis

A geostatistical method is based on the regionalized variable theory which states that variables have both random and spatial structures in an area. A variogram of the data should first be determined. An experimental variogram is computed to quantify the spatial variability of variables. The experimental variogram

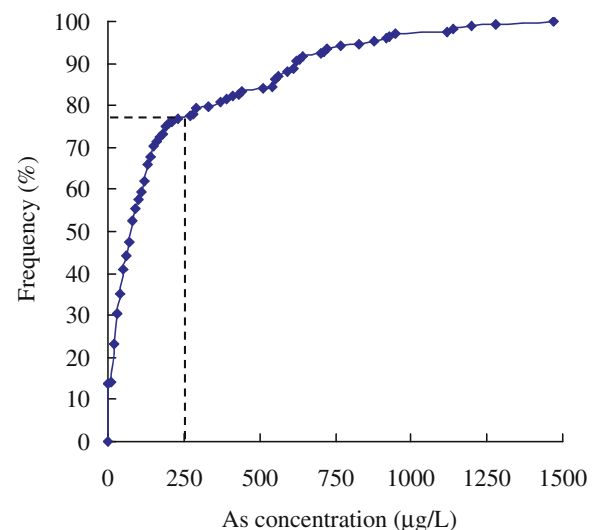


Fig. 2 A cumulative distribution function of observed As concentrations

is fitted by a theoretical model, $\gamma(h)$, which can be spherical, exponential or Gaussian model, and three parameters of the fitted model – the nugget effect, the sill and the range – are determined (Isaaks and Srivastava 1989). In the study area, the spatial scale in the horizontal direction differs greatly from that in the vertical direction. A three dimensional (3D) anisotropic variogram was used to analyze the spatial variability (Deutsch 2002). This study set eight thresholds of As concentrations – the 16th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles of measured concentrations – and analyzed their variograms. Jang et al. (2006) discussed in detail the 3D anisotropic variability of As concentrations in groundwater for the thresholds. This study followed results of variogram analysis reported in Jang et al. (2006).

Indicator kriging

Indicator kriging is a non-parametric geostatistical method for estimating the probability that the attribute value is no greater than a specific threshold, z_k , at a given location \mathbf{u} (Goovaerts 1997). In IK, the spatial variable, $Z(\mathbf{u})$, is transformed into an indicator variable with a binary distribution, as follows.

$$I(\mathbf{u}; z_k) = \begin{cases} 1, & \text{if } Z(\mathbf{u}) \leq z_k, k = 1, 2, \dots, m \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The expected value of $I(\mathbf{u}; z_k)$, conditional on n surrounding data, can be expressed as,

$$\begin{aligned} E[I(\mathbf{u}; z_k | (n))] &= \text{Prob}\{Z(\mathbf{u}) \leq z_k | (n)\} \\ &= F(\mathbf{u}; z_k | (n)) \end{aligned} \quad (2)$$

where $F(\mathbf{u}; z_k | (n))$ is the conditional cumulative distribution function (ccdf) of $Z(\mathbf{u}) \leq z_k$. Indicator kriging is an estimation technique which is based on an estimator that is defined as,

$$I^*(\mathbf{u}_0; z_k) = \sum_{j=1}^n \lambda_j(z_k) I(\mathbf{u}_j; z_k) \quad (3)$$

where $I(\mathbf{u}_j; z_k)$ represents the values of the indicator at the measured locations, \mathbf{u}_j , $j=1, 2, \dots, n$, and $\lambda_j(z_k)$ is a weighting factor of $I(\mathbf{u}_j; z_k)$ used in estimating $I^*(\mathbf{u}_0; z_k)$.

A linear interpolation yields a continuous ccdf within each class of threshold values (z_{k-1} , z_k). The continuous ccdf at the lower tail is extrapolated toward a zero using a negatively skewed power

model with $\omega=2.5$. The continuous ccdf at the upper tail is extrapolated toward an infinite upper bound using a hyperbolic model with $\omega=1.5$ (as recommended by Goovaerts (1997) and Deutsch and Journel (1998)). This work used the gamv and ik3d codes in GSLIB (Deutsch and Journel 1998) to perform the experimental variogram and IK, respectively.

Risk-based classification of contaminated and safe regions

Based on the aforementioned ccdf established using IK, three criteria for classifying a location as contaminated follows (Goovaerts 1997):

1. The 0.771-quantile estimate is used to determine an As-contaminated concentration based on a ccdf each cell. That is, the cumulative probability of 0.771 in a ccdf each cell responds to the As-polluted concentration. A site is considered as contaminated if its estimated concentration exceeds 250 $\mu\text{g/l}$. Meanwhile, a site is considered as safe if its estimated concentration is under 250 $\mu\text{g/l}$.

2. The E-type estimate is an expected value of As-contaminated concentrations based on a ccdf each cell. The contaminated or safe locations are determined as well as the previous classification approach.

3. The minimization of the expected loss is to evaluate the impact of the two possible decisions. The loss associated with classifying a location \mathbf{u} as safe could be modeled as

$$L_1(z(\mathbf{u})) = \begin{cases} 0 & \text{if } z(\mathbf{u}) \leq z_c \\ w_1[z(\mathbf{u}) - z_c] & \text{otherwise} \end{cases} \quad (4)$$

where w_1 is the relative cost of underestimating the toxic concentration and z_c is the threshold. The loss associated with classifying a location \mathbf{u} as contaminated could be modeled as

$$L_2(z(\mathbf{u})) = \begin{cases} 0 & \text{if } z(\mathbf{u}) > z_c \\ w_2 & \text{otherwise} \end{cases} \quad (5)$$

The remediation cost is modeled as a constant value w_2 . Based on the ccdf each cell, the expected loss attached to the two types of classification is determined as $\varphi_1(\mathbf{u}) = E[L_1(z(\mathbf{u})) | (n)]$

$$\begin{aligned} &= \sum_{k=1}^{K+1} L_1(\bar{z}_k) [F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))] \\ &\quad (6) \end{aligned}$$

$$\begin{aligned}\varphi_2(\mathbf{u}) &= E[L_2(z(\mathbf{u}))|(n)] \\ &= \sum_{k=1}^{K+1} L_2(\bar{z}_k)[F(\mathbf{u}; z_k|(n)) - F(\mathbf{u}; z_{k-1}|(n))]\end{aligned}\quad (7)$$

where K is the threshold number of ccdf and \bar{z}_k is the mean of classes. The location \mathbf{u} is then declared safe or contaminated so as to minimize the resulting expected loss:

If $\varphi_I(\mathbf{u}) > \varphi_2(\mathbf{u})$, \mathbf{u} is classified as contaminated.

If $\varphi_I(\mathbf{u}) < \varphi_2(\mathbf{u})$, \mathbf{u} is classified as safe.

Misclassification risks

Two misclassification risks are assessed based on a ccdf model $F(\mathbf{u}; z_k|(n))$. The risk $\alpha(\mathbf{u})$ of incorrectly classifying a location \mathbf{u} as hazardous (false positive) is defined as follows (Goovaerts 1997).

$$\begin{aligned}\alpha(\mathbf{u}) &= \text{Pr ob}\{Z(\mathbf{u}) \leq z_c | z_L^*(\mathbf{u}) > z_c, (n)\} \\ &= F(\mathbf{u}; z_c|(n))\end{aligned}\quad (8)$$

where $z_L^*(\mathbf{u})$ is the estimator. The risk $\beta(\mathbf{u})$ of wrongly classifying a location \mathbf{u} as safe (false negative) is expressed as follows (Goovaerts 1997).

$$\begin{aligned}\beta(\mathbf{u}) &= \text{Pr ob}\{Z(\mathbf{u}) > z_c | z_L^*(\mathbf{u}) \leq z_c, (n)\} \\ &= 1 - F(\mathbf{u}; z_c|(n))\end{aligned}\quad (9)$$

The misclassification risks were applied to examine three classified results and provided potential misjudged information.

Results and discussion

A spatial pattern of high As-contaminated groundwater at the shallow layer is apparently different to those at the middle and deep layers. Three layers at depths of 50, 150 and 250 m under the ground surface represented the use of groundwater for aquaculture at the shallow, middle and deep layers, respectively. Additionally, each layer was discretized into a grid of 44×100 cells, with a spacing of 1 km.

Classification based on different estimates

A corresponding As concentration was computed using a linear interpolation or extrapolation with a

power model or a hyperbolic model based on the 0.771-quantile in a ccdf at each cell. Figure 3a displays estimated As concentrations using the 0.771-quantile estimate at three depths. A maximal estimated As concentration is 1,060 $\mu\text{g/l}$ occurred at the 150 m depth. Figure 3b presents black and dark gray areas where estimated As concentrations exceed 250 $\mu\text{g/l}$. Approximately 54.4% of the regions are declared as contaminated (Table 1). Figures 3c and d show risks to false positives and false negatives, respectively. Generally, more credible contaminated areas declared have lower misclassification risks. A categorized threshold of 0.5 was utilized to determine misclassification with a high or a low risk. In borders of the polluted zones, risks of more than 0.5 are wrongly defined as contaminated, but they do not exceed 0.771. Dark gray zones in Fig. 3b, roughly 41.2% of total areas (Table 1), are high risks to false positives. In borders of the safe regions, risks of 0.1–0.229 are wrongly classified as safe.

An expected value of As concentrations was computed using the E-type estimate based on nine classes in a ccdf at each cell. Figure 4a exhibits the expected values of As concentrations at three depths. A maximal expected value of As concentrations is 576.8 $\mu\text{g/l}$, occurred at the 150 m depth. The maximal As concentration estimated by the E-type estimate is considerably smaller than that by the 0.771-quantile estimate. Figure 4b describes black and dark gray areas where estimated As concentrations exceed 250 $\mu\text{g/l}$. Approximately 40.6% of the regions are declared as contaminated (Table 1). Figures 4c and d exhibit risks to false positives and false negatives, respectively. Potential misclassification zones, which have high risks to false positives, are marked as dark gray in Fig. 4b. They are 27.4% of total areas with a maximal risk probability of 0.737 (Table 1). In borders of the safe regions, risks of 0.2–0.413 are wrongly classified as safe. They rise significantly compared with those using the 0.771-quantile estimate, but do not surpass the potential misjudged threshold of 0.5.

The remediation cost is the expense which uses surface water to replace groundwater for aquacultural needs. The cost of the potential ill health is the loss that residents suffer from cancers and diseases associated with ingesting As. Unfortunately, field survey data on potential ill health (w_1) and remediation cost (w_2) in the BFD areas are lack. This work adopted parameters of $w_1=1$ and $w_2=2.5$, used by

Fig. 3 **a** Distributions of As concentrations using the 0.771-quantile estimate; **b** classification of contaminated sites; **c** risks to false positives; and **d** risks to false negatives

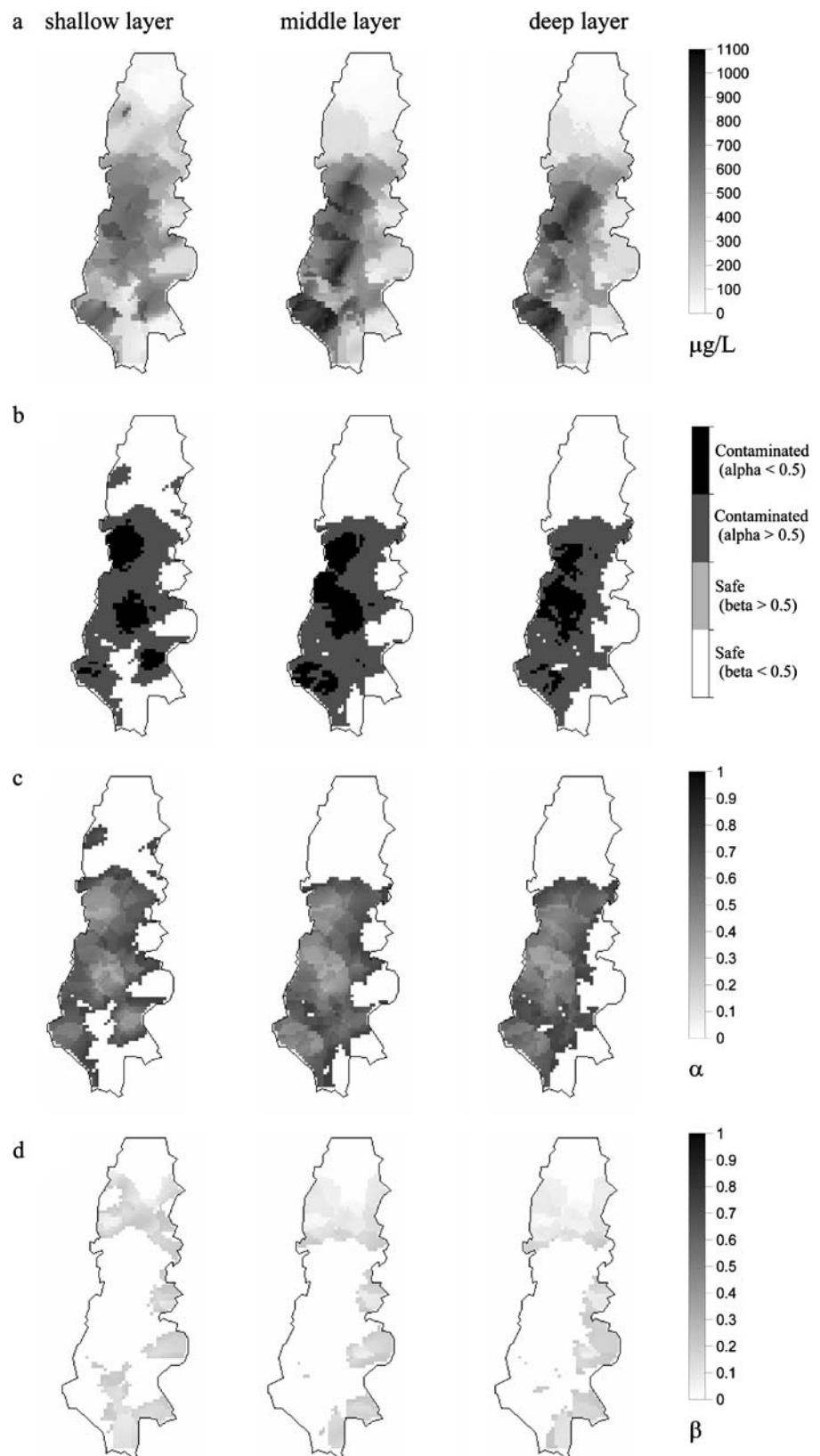


Table 1 Summary and comparison of estimated results

Approaches	0.771-quantile	E-type	Minimization of the expected loss ($w_2/w_1=2.5$)
Polluted area (%)			
Total	54.4	40.6	35.4
$\alpha > 0.5$	41.2	27.4	22.5
$\beta > 0.5$	0	0	0.2
Uncertainty assessment of small deviations: 1. Variations of polluted area (%)			
Total	16.1 ^a	— ^c	16.0 ^b
$\alpha > 0.5$	16.1 ^a	— ^c	15.3 ^b
$\beta > 0.5$	0 ^a	— ^c	0.8 ^b
Uncertainty assessment of small deviations: 2. Variations of misclassification risks			
Maximal α	−0.1 ^c (decrease)	— ^c	0.086 ^d (increase)
Maximal β	0.1 ^c (increase)	— ^c	−0.086 ^d (decrease)
Merit/demerit			
Merits	1. small polluted areas with risks of $\beta > 0.5$ 2. applicable to estimating extreme values 3. simple calculation	1. middle polluted areas with risks of $\alpha > 0.5$ 2. simple calculation	1. small polluted areas with risks of $\alpha > 0.5$
Demerits	1. large polluted areas with risks of $\alpha > 0.5$	1. expected values determined from ccdf easily smooth out extreme values	1. large polluted areas with risks of $\beta > 0.5$ upon $w_2/w_1 > 3$ 2. large maximum risks of α and β 3. difficulty in obtaining a field w_2/w_1 value.

^a Variation of polluted areas computed from the 0.8-quantile to the 0.7-quantile.

^b Variation of polluted areas computed from the $w_2/w_1=2$ to the $w_2/w_1=3$.

^c Variation of misclassification risks computed from the 0.8-quantile to the 0.7-quantile.

^d Variation of misclassification risks computed from the $w_2/w_1=2$ to the $w_2/w_1=3$.

^e Not evaluated.

Goovaerts (1997). Expected values of potential ill health and remediation cost were estimated according to Eqs. 6 and 7 (Figs. 5a and b, respectively). If an expected value of potential ill health exceeds that of remediation cost at a cell, the cell is regarded as a contaminated site. Meanwhile, if an expected value of remediation cost exceeds that of potential ill health at a cell, the cell is regarded as a safe region (Fig. 5c). Thus, roughly 35.4% of the regions are declared as contaminated locations (Table 1). Figures 5d and e elucidate risks to false positives and false negatives, respectively. Figure 5c presents polluted zones with low and high misclassification risks. Contaminated zones with high risks to false positives are 22.5% of total areas and possess a maximal risk of 0.94 (Table 1). Few classified safe regions have high risks to false negatives (0.2% of total areas).

Typically, the 50th percentile of measured As concentrations (median value) used to delineate pollution areas can balance and minimize misclassification risks. Misclassification risks to false positives are more serious than those to false negatives in this study owing to the 77.1th percentile of measured As concentrations used. Although proportions of total contaminated areas and contaminated areas with high risks to false positives vary as the different estimates, proportions of contaminated areas with low risks to false positives keep stationary, about 12.9–13.2%. The patterns of declared contaminated areas with low risks to false positives are alike for the three classification approaches. Furthermore, Jang et al. (2006) documented that 16.3 and 70.9% of total areas are hazardous for human health at the 75th percentile and the 95th percentile of risks, respectively. Their

Fig. 4 **a** Distributions of As concentrations using the E-type estimate; **b** classification of contaminated sites; **c** risks to false positives; and **d** risks to false negatives

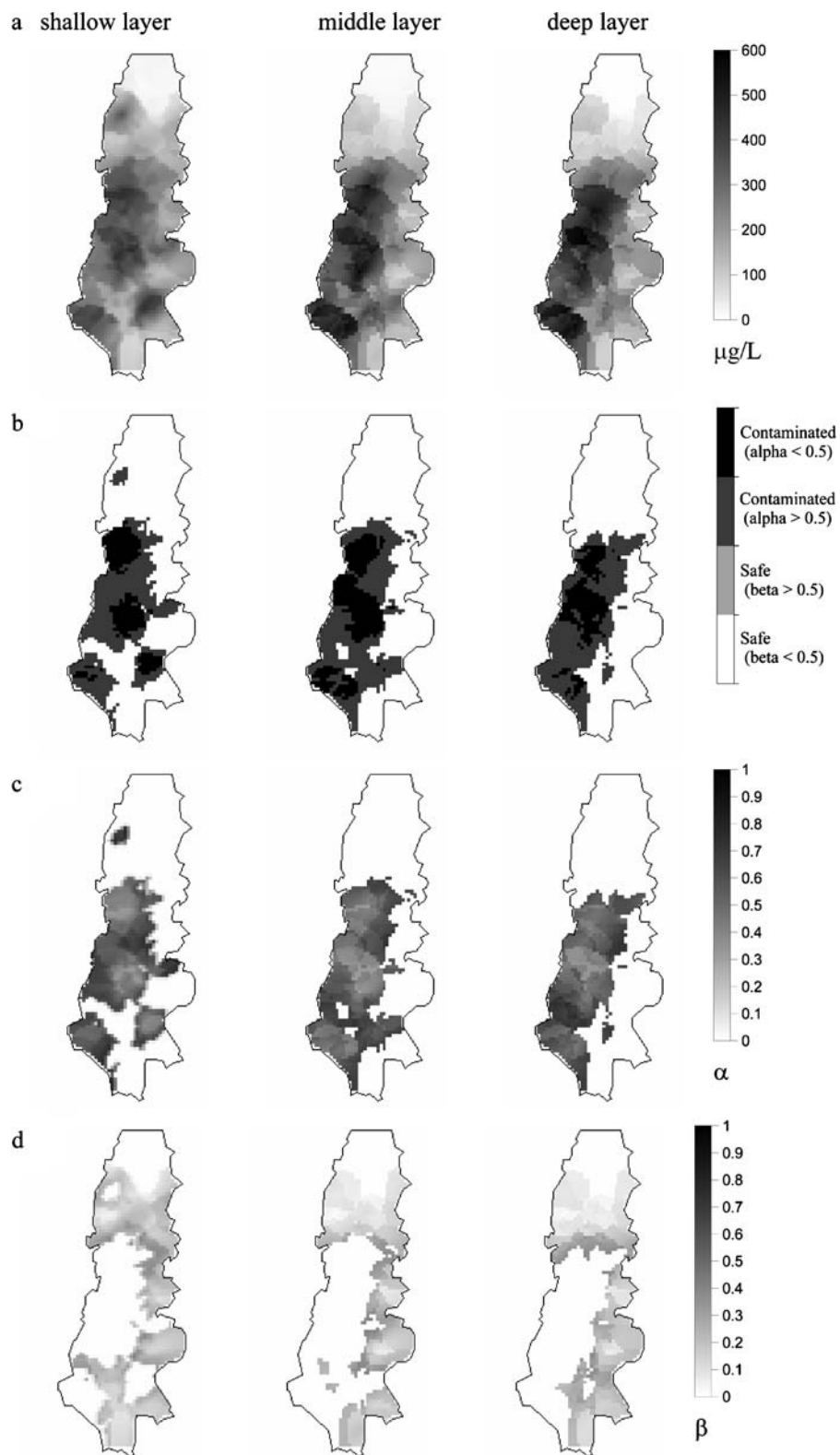
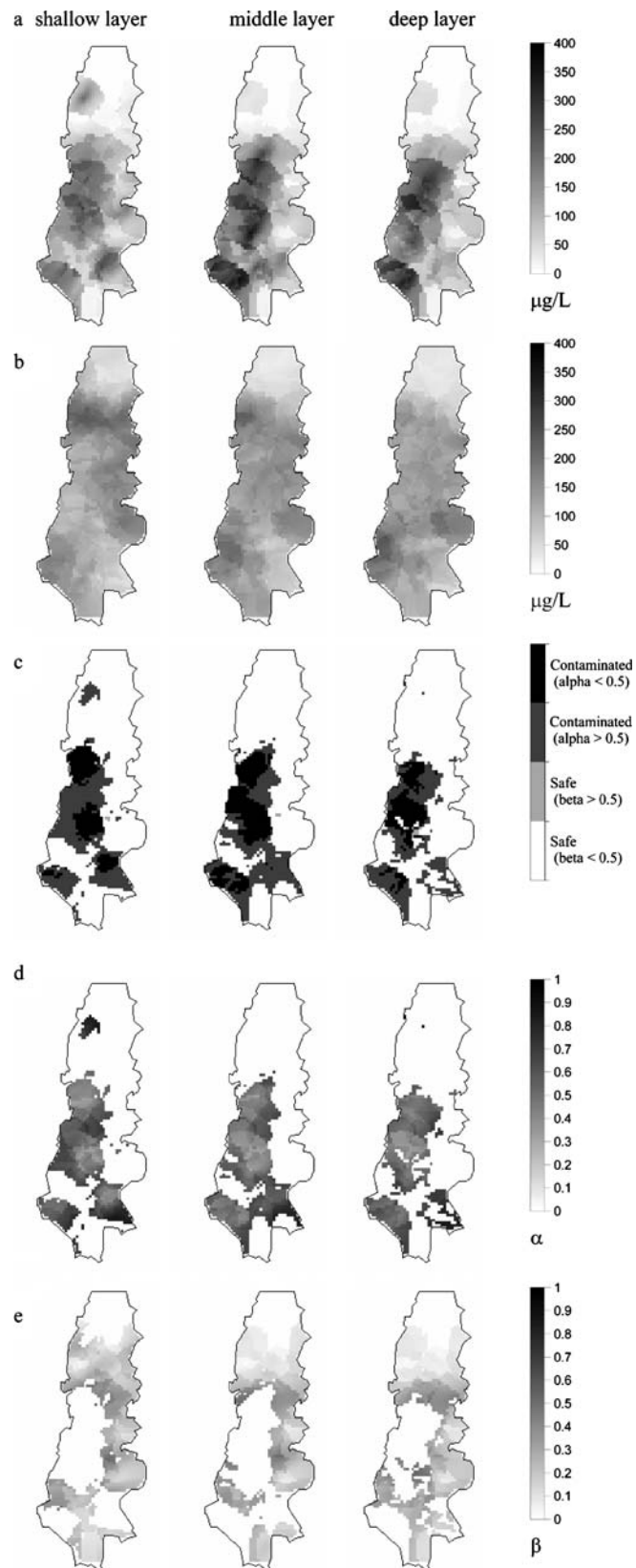


Fig. 5 **a** Expected values of potential ill health; **b** expected values of remediation cost; **c** classification of contaminated sites using the minimization of the expected loss; **d** risks to false positives; and **e** risks to false negatives



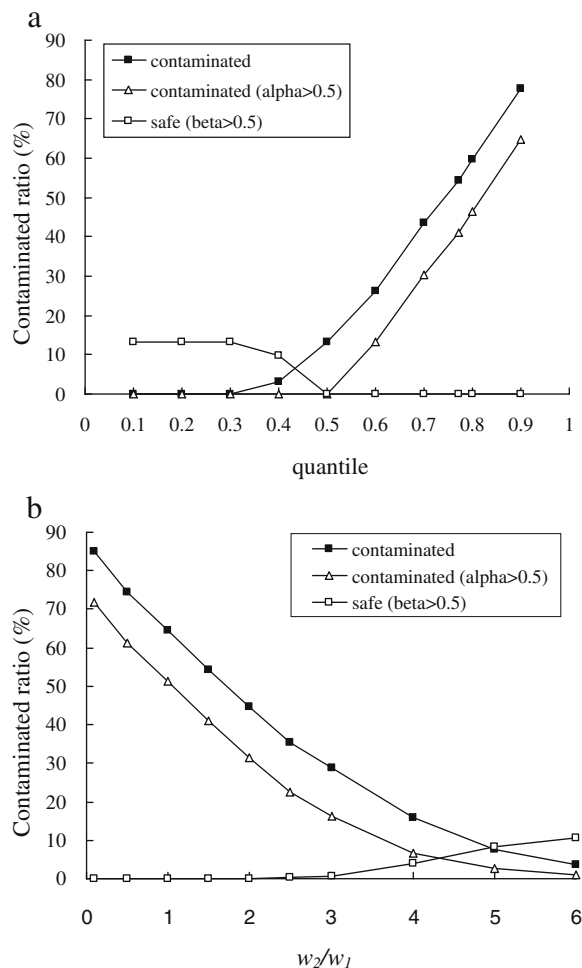


Fig. 6 **a** Relationship between various quantiles and corresponding contaminated ratios, and **b** relationship between various w_2/w_1 and corresponding contaminated ratios

work suggested that the 75th percentile of risks is considered in formulating a remedial strategy for the aquacultural use of groundwater in the preliminary stage. The contaminated sites estimated at the 75th percentile of risks to human health are compared with those with low classification risks. The proportion of contaminated sites analyzed by risk assessment of human health is slightly larger than that estimated by the three indicator-based approaches. Great difference between two occurs at the shallow layer. Only northern contaminated sites at the shallow layer are declared in Jang et al. (2006) and are considerably smaller than those declared in this work. Meanwhile, contaminated sites at the middle and deep layers declared in Jang et al. (2006) are larger than those declared in this work.

Classification uncertainty

The 77.1th percentile of As-polluted concentrations corresponding to the regulated level of 250 $\mu\text{g/l}$ does not guarantee to be obtained the same data set in other surveys. Moreover, because of lacks of field data, a ratio of $w_2/w_1=2.5$ is considerably uncertain for classifying polluted scopes. This study applied the sensitivity analysis to explore their uncertainty. However, the E-type estimate is not considered with parameter uncertainty because an expected value smoothes out extreme values and the E-type estimate is less sensitive to estimated parameter variation than the remaining two estimates.

Figure 6a demonstrates ratios of total contaminated areas, and contaminated areas with high risks to false positives and false negatives against various quantiles. No regions are delineated as contaminated upon the quantiles of less than 0.2. Contaminated areas with high misclassification risks are null using the 0.5-quantile estimate (the median estimate). Over the 0.5-quantile, ratios of total contaminated areas and contaminated areas with high risks to false positives rise as increasing quantiles, and ratios of contaminated sites with low risks to false positives maintain a constant, 13.2%. Under the 0.5-quantile, ratios of contaminated sites with high risks to false negatives increase as decreasing quantiles, and they have a maximum value of 13.2%.

Figure 6b illustrates ratios of total contaminated areas and contaminated areas with high risks to false positives and false negatives against various w_2/w_1 . The low w_2/w_1 increases ratios of contaminated areas with high risks to false positives. Ratios of contaminated sites with low risks to false positives still keep a constant, 13.2%, as the w_2/w_1 of less than two and decrease as the w_2/w_1 of more than two. Ratios of contaminated sites with high risks to false negatives significantly increase when the w_2/w_1 surpasses three.

Figure 7a accounts for distributions of false positives and false negatives against different quantiles. The distributions of the misclassification risks systematically vary: distributions of false positives grow and distributions of false negatives reduce as increasing quantiles. Figure 7b presents distributions of false positives and false negatives against various w_2/w_1 . The distributions of false positives reduce as increasing w_2/w_1 , whereas the distributions of false negatives grow as increasing w_2/w_1 . The distributions of false positives

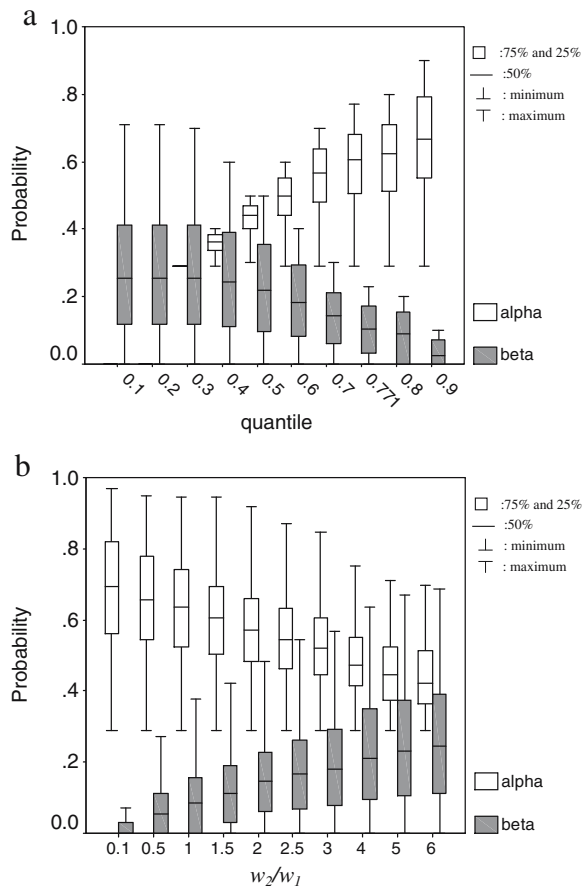


Fig. 7 Distributions of risks to false positives and false negatives using **a** the *p*-quantile estimate, and **b** the minimization of the expected loss

using the minimization of the expected loss are typically wider than those using the *p*-quantile estimate.

When small deviations estimated at the 0.771-quantile and the w_2/w_1 of 2.5 (deviation values of 0.1 and 1, respectively) occur, total contaminated areas and contaminated areas with high risks to false positives vary 16.1–15.3% (Table 1), but contaminated areas with low risks to false positives insignificantly change. Risk distributions of false positives and false negatives associated with the small deviation alter 0.1 for the 0.771-quantile estimate and 0.086 for the minimization of the expected loss (Table 1). Consequently, the aforementioned parameters with small imprecision used in this study do not influence on delineating contaminated areas with low risks to false positives but considerably impacts on classifying the others. The risk-based indicator approaches not only are used to delimit polluted regions but also provide some useful information

associated with their misclassification risks and uncertainty measurement. Moreover, the risk-based indicator approaches adopted to classify hazardous extents for policy-making are simpler and more practical than risk assessment approach used in Jang et al. (2006).

Conclusions

This work determined extents of As-polluted groundwater using risk-based indicator approaches in BFD hyperendemic areas of southern Taiwan. Three approaches – the *p*-quantile estimate, the E-type estimate and the minimization of the expected loss – were used to delimit contaminated regions for a specific regulated standard of As concentrations in groundwater. The standard was set to 250 $\mu\text{g/l}$ for aquacultural use, representing no adverse health effects on human. Some useful information associated with misclassification risks and uncertainty can provide a reference for policy-makers. Because of several advantages of groundwater utilization, such as convenient acquisition, low expense of withdrawal and elevating temperature in pond water in winter, As-contaminated groundwater is used inevitably in BFD hyperendemic regions. Delineation of polluted regions is critical for managing groundwater uses to farm fish. If being declared as contaminated, regions are not proper for aquaculture. For the three classification approaches, total contaminated areas are the largest using the 0.771-quantile estimate, whereas they are the smallest using the minimization of the expected loss. Table 1 lists some merits and demerits of using the three classification approaches. None of the approaches possess all the advantage of delimiting polluted regions and misclassification risks. Nevertheless, although total contaminated areas and contaminated areas with high risks to false positives vary for the three classification approaches, credible contaminated areas with low risks to false positives approach to a constant, 12.9–13.2% of contaminated areas. Arsenic-polluted groundwater should be prohibited strictly to farm fish in the credible contaminated zones and monitored persistently in contaminated zones with high risks to false positives.

Acknowledgements The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contracts Nos. NSC90-2313-B-002-322 and NSC91-2313-B-002-270.

References

- Central Geological Survey (1999). *Project of groundwater monitoring network in Taiwan during first stage-research report of Choushui river alluvial fan* (pp. 383). Taiwan: Water Resources Bureau, in Chinese.
- Chen, S. L., Dzen, S. R., Yang, M. H., Chiu, K. H., Shieh, G. M., & Wai, C. M. (1994). Arsenic species in groundwaters of the blackfoot disease area, Taiwan. *Environmental Science & Technology*, 28, 877–881.
- Chen, C. J., Hsueh, Y. M., Lai, M. S., Shyu, M. P., Chen, S. Y., Wu, M. M., et al. (1995). Increased prevalence of hypertension and long-term arsenic exposure. *Hypertension*, 25, 53–60.
- Chen, C. J., & Wang, C. J. (1990). Ecological correlation between arsenic level in well water and aged-adjusted mortality from malignant neoplasms. *Cancer Research*, 50, 5470–5474.
- Deutsch, C. V. (2002). *Geostatistical reservoir modeling* (pp. 124–152). New York: Oxford University Press.
- Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation* (pp. 259–368). New York: Oxford University Press.
- Goovaerts, P., Avruskin, G., Meliker, J., Slotnick, M., Jacquez, G., & Nriagu, J. (2005). Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. *Water Resources Research*, 41, 10.1029/2004WR003705.
- Han, B. C., Jeng, W. L., Chen, R. Y., Fang, G. T., Hung, T. C., & Tseng, R. J. (1998). Estimation of target hazard quotients and potential health risks for metals by consumption of seafood in Taiwan. *Archives of Environmental Contamination and Toxicology*, 35, 711–720.
- Huang, Y. K., Lin, K. H., Chen, H. W., Chang, C. C., Liu, C. W., Yang, M. H., et al. (2003). Arsenic species contents at aquaculture farm and in farmed mouthbreeder (*Oreochromis mossambicus*) in blackfoot disease hyperendemic areas. *Food and Chemical Toxicology*, 41, 1491–1500.
- Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to applied geostatistics* (pp. 278–322). New York: Oxford University Press.
- Jang, C. S., Liu, C. W., Lin, K. H., Huang, F. M., & Wang, S. W. (2006). Spatial analysis of potential carcinogenic risks associated with ingesting arsenic in aquacultural tilapia (*Oreochromis Mossambicus*) in blackfoot disease hyperendemic areas. *Environmental Science & Technology*, 40, 1707–1713.
- Juang, K. W., Chen, Y. S., & Lee, D. Y. (2004). Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environmental Pollution*, 127, 229–238.
- Juang, K. W., & Lee, D. Y. (1998). Simple indicator kriging for estimating the probability of incorrectly delineating hazardous areas in a contaminated site. *Environmental Science & Technology*, 32, 2487–2493.
- Lai, M. S., Hsueh, Y. M., Chen, C. J., Shyu, M. P., Chen, S. Y., Kuo, T. L., et al. (1994). Ingested inorganic arsenic and prevalence of diabetes mellitus. *American Journal of Epidemiology*, 139, 484–492.
- Liu, C. W., Jang, C. S., & Liao, C. M. (2004). Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). *Science of the Total Environment*, 321, 173–188.
- Liu, C. W., Wang, S. W., Jang, C. S., & Lin, K. H. (2006). Occurrence of arsenic in groundwater of the Choshui river alluvial fan, Taiwan. *Journal of Environmental Quality*, 35, 68–75.
- Saisana, M., Dubois, G., Chaloulakou, A., & Spyrellis, N. (2004). Classification criteria and probability risk maps: Limitations and perspectives. *Environmental Science & Technology*, 38, 1275–1281.
- Taiwan Sugar Company (2001). *Water quality survey and analysis of groundwater monitoring network (3/5)* (pp. 67–76). Taiwan: Water Resources Bureau, in Chinese.
- Taiwan Sugar Company (2002). *Water quality survey and analysis of groundwater monitoring network (4/5)* (pp. 57–88). Taiwan: Water Resources Bureau, (in Chinese).
- Tseng, W. P. (1977). Effects and dose–response relationships of skin cancer and blackfoot disease with arsenic. *Environmental Health Perspectives*, 19, 109–119.
- US EPA. (2001). *Risk assessment guidance for superfund: Volume III-Part A, process for conducting probabilistic risk assessment*. EPA 540-R-02-002. Washington, DC: US Environmental Protection Agency.
- Wang, S. W., Liu, C. W., & Jang, C. S. (2007). Factors responsible for high arsenic concentrations in two groundwater catchments at Taiwan. *Applied Geochemistry*, 22, 460–476.
- Wu, M. M., Kuo, T. L., Hwang, Y. H., & Chen, C. J. (1989). Dose–response relation between arsenic concentration in well water and mortality from cancers and vascular diseases. *American Journal of Epidemiology*, 130, 1123–1132.