자원환경지질, 제36권, 제3호, 191-200, 2003
Econ. Environ. Geol., 36(3), 191-200, 2003

Optimization of Contaminated Land Investigation based on Different Fitness-for-Purpose Criteria

Jong-Chun Lee[1]* and Michael H. Ramsey[2]

[1]*National Institute of Environmental Research*
[2]*Centre for Environmental Research, University of Sussex, UK*

조사목적별 기준에 부합하는 오염부지 조사방법의 최적화 방안에 관한 연구

이종천[1]* · Michael H. Ramsey[2]

[1]국립환경연구원, [2]영국서섹스대학 환경연구센터

중금속으로 오염된 폐광산 주변부나 유류누출로 인한 토양오염 등과 같은 오염토양부지에 대한 조사를 위해서는 우선 조사의 목적과 상황에 알맞은 기준을 설정하고, 이를 토대로 조사방법을 설계하여야 한다. 이러한 상황별 조사방법은 조사결과의 질적 수준 뿐만 아니라 모방결합에 영향을 미치게 되는데 그 수준을 제한하는 요인으로는 보통 총 조사비용, 관련규제법이 요구하는 조사수준, 그리고 전문적 경험에 의한 주관적 기준 등이 그 예가 될 수 있다. 본 논문에서는 조사목적에 따라 달리하는 각 기준(fitness-for-purpose criteria)을 만족시키게끔 조사방법을 설계하는 예를 들어보았다. 먼저, 경제적 요소와 측정된 자료의 질적 수준을 참고하는 방법을 고안하여 보았다. 이를 위해 측정불확도(measurement uncertainty)로 평가되는 측정결과의 질적 수준이 의사결정의 오류에 영향을 미치는 확률을 확률적손실계산식을 이용하여 평가하고, 이를 이용하여 비용경제적인 측면에서 조사를 최적화하는 방안을 고안하였다. 이와 더불어 위해성 평가를 위한 부지의 평균오염도수준 산출의 예와 같이 의사결정의 오류에 의한 손실보다는 측정된 평균값의 오차와 측정에 요구되는 비용을 최소화 해야 하는 경우와, 또한 측정값에 대한 일정한 질적 수준이 요구되는 경우의 조사방법, 그리고 제한된 비용을 최대한 활용할 수 있는 시료수와 분석수의 최적비율 등을 도출하는 방법을 다루었다. 각 방법은 영국의 두 오염지역에 적용하여 그 타당성을 평가하여보았다.

주요어 : 오염부지조사, 측정불확도, 손실계산식, 시료채취방법

Investigations on the contaminated lands due to heavy metals from mining activities or hydrocarbons from oil spillage for example, should be planned based on specific fitness-for-purpose criteria(FFP criteria). A FFP criterion is site specific or varies with situation, based on which not only the data quality but also the decision quality can be determined. The limiting factors on the qualities can be, for example, the total budget for the investigation, regulatory guidance or expert's subjective fitness-for-purpose criterion. This paper deals with planning of investigation methods that can satisfy each suggested FFP criterion based on economic factors and the data quality. To this aim, a probabilistic loss function was applied to derive the cost effective investigation method that balances the measurement uncertainty, which estimates the degree of the data quality, with the decision quality. In addition, investigation planning methods when the objectives of investigations do not lie in the classification of the land but simply in producing the estimation of the mean concentration of the contaminant at the site(e.g. for the use in risk assessment), were also suggested. Furthermore, the efficient allocation of resources between sampling and analysis was also devised. These methods were applied to the two contaminated sites in the UK to test the validity of each method.

Key words : Contaminated land investigation, measurement uncertainty, loss function, sampling method

---

*Corresponding author: jclee@me.go.kr

1. Introduction

Decision-making on a potentially contaminated land, for either redevelopment or remediation, depends on the level and the extent of contaminant concentration estimated from the investigation of the land. The validity of the decision can be questioned if the process failed to assess the uncertainty on the measurements properly, or if the level of uncertainty exceeds a tolerable limit. This may eventually cause financial loss to the landowners or developers.

However, the uncertainty cannot be reduced without increasing cost for the sampling and/or analytical methods. In order to design a "cost-effective contaminated land investigation", the level of the uncertainty should therefore balance the potential financial loss from the decision error and the cost of the sampling and analysis: the former relating to "decision quality" and the latter to "data quality".

The level of the data quality required of a site investigation can be determined during the planning stage according to, for example, the total budget for the investigation, regulatory guidance or experts subjective fitness-for-purpose criterion (hereafter FFP criterion). If not restricted by such prescribed criteria, the investigation can be planned, at least in part, by considering economic factors. A "cost-effective investigation" can be designed to produce appropriate data quality at a minimum expenditure of total monetary cost. This data quality can be represented by the level of overall measurement uncertainty. In addition, the total cost should include not only the cost for sampling and analysis during the investigation but also for any financial loss incurred by a misclassification of the land.

The estimated degree of the measurement uncertainty, which affects both the data quality and the decision quality, can be converted to give both the cost for making the measurement at the particular uncertainty level and the economic loss from a decision error. The result can then be applied to the optimization of an investigation process on a contaminated land. The relationship between the uncertainty of the measured concentration and the expected monetary loss can be described by using the "loss function based on misclassification error". Consequently, the contaminated land investigation can be designed on the basis of the optimum uncertainty level that is determined by minimization of the loss function.

However, there are also cases when the objectives of investigations do not lie in the classification of the land but simply in producing the estimation of the mean concentration of the contaminant at the site(e.g. for the use in risk assessment). Moreover, the approach using the loss function based on misclassification error cannot handle the situation very well if the overall concentration of the land is well above or well below the threshold limit, so that there is virtually no chance of the misclassification error; or if the overall concentration is so close to the threshold that the cost of reducing misclassification error is too high. In these situations, the investigation should be carried out on the basis of different FFP criteria that minimize the resources for the investigation within the boundary of the limitation.

This paper therefore aims to apply the loss function based on misclassification error for the optimization of contaminated land investigation along with the other cost-effective investigation design methods, using data from the two contaminated sites in the UK with contrasting heterogeneity and level of contamination. The other loss function such as used in the Cochrans sampling theory, which calculates financial loss based on the error of the mean concentration estimation, is tested for the effectiveness on a different fitness-for-purpose criterion(FFP criterion). In addition, investigation at a target data quality is also designed with the further applicable implication of the optimal allocation method.

2. Integrating decision quality and data quality into optimized contaminated land investigation

2.1. Quantification of misclassification errors

"Uncertainty of a measured value(or measurement uncertainty, ISO 1993)" can be explained as an interval containing the true value with high probability(Thompson, 1995). The interval is usually expressed using standard deviation(*s*) estimated from an appropriate method. There has been no shortage of studies regarding the esti-

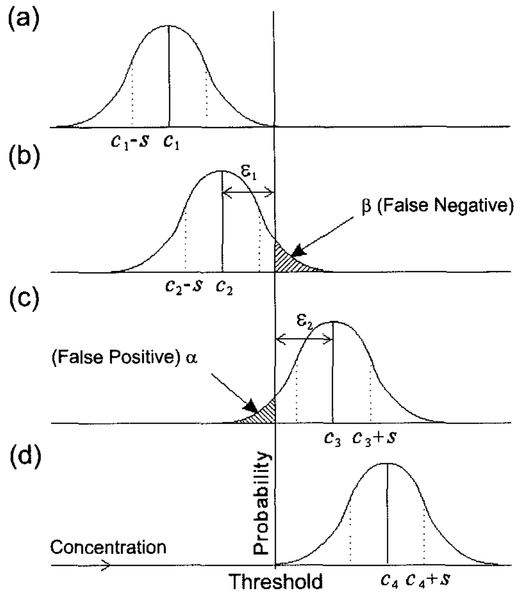

Fig. 1. Probabilities of making misclassification errors due to the measurement uncertainty around the measured concentration. Shaded area indicates the probability of true value lies across the threshold from the measured concentration. (a) Measured concentration is well below the threshold (i.e. uncontaminated) so the probability of making a misclassification error is virtually zero. (b) Probability liable to misclassification error (a false negative error, β) on a measured concentration (c_2) while the true concentration is assumed to be above the regulatory threshold value (T). (c) Probability liable to misclassification error (a false positive error, α) on a measured concentration (c_3) while the true concentration is assumed to be below the regulatory threshold value (T). (d) Measured concentration well above the threshold (i.e. contaminated) so the probability of making a misclassification error is virtually zero. Note: s= parameter for an uncertainty on measured concentration expressed in standard concentration, c=measured concentration, ε = |Threshold-measured concentration|.

mation of measurement uncertainty originating from the investigation of contaminated land(Miesch, 1967; Garrett and Goss, 1979; Ramsey, 1997; Ramsey, 1998; Lee and Ramsey, 2001). The probability of making misclassification error applying the estimated measurement uncertainty can be calculated using the "standard normal cumulative distribution function", Φ(z)(Equation 3). For example, the probabilities of making false negative error (β, Fig. 1b) and false positive error(α, Fig. 1c) can be calculated as,

$$\beta=1-\Phi(z_1) \qquad (1)$$

$$\alpha=1-\Phi(z_2) \qquad (2)$$

$$\Phi(z)=\int_{-\infty}^{z}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\right)dz \qquad (3)$$

$$z_1=|\varepsilon_1|/s$$
$$z_2=|\varepsilon_2|/s \qquad (4)$$

where ε=|Threshold-measured concentration|

The diagrams in Fig. 1 illustrate that the error on the measured concentration does not necessarily incur loss if the measured concentration value is well below(Fig. 1a), or well above(Fig. 1d), the regulatory threshold value(T).

2.2. Probabilistic loss functions based on misclassification error

One of the major objectives of a site investigation is to classify the potentially contaminated land. In this case, the error of classification(or misclassification error) cannot be estimated simply by the difference between the measured value and the true value, as were the cases for the loss functions in the previous sections. In other words, the misclassification error is to be estimated as a "probability" that makes reference to a threshold value for the classification.

An estimate of concentration has an uncertainty that can be expressed as a probabilistic distribution on the measured concentration. When the uncertainty on a measured value is compared with the threshold value, the probability of a misclassification can then be estimated. This quantification of misclassification error(Section 2.1) can be applied to develop a loss function that estimates an expected financial loss due to the misclassification error(Thompson and Fearn, 1996).

The "expectation of total loss", E(L), consists of the two terms(Equation 5): loss of the end-user of the investigation result incurred by the misclassifying the land(L_{end}) and the cost of sampling and analysis(L_{meas}).

$$E(L)=L_{end}+L_{meas} \qquad (5)$$

The "loss from the misclassification error(L_{end})" is a product of the "misclassification error(α or β)" and the "consequence cost(C_c)" that refers to the maximum loss caused by either a false negative error($C_{c\beta}$) or by a false positive error($C_{c\alpha}$).

The "cost of sampling and analysis(L_{meas})" can be calculated based on a "cost rule" that describes the inversely proportional relationship between the variance and the number of measurements. Hence, the cost of sampling and analysis(L_{meas}) can be calculated as

$$L_{meas}=\frac{D}{S^2_{meas}} \tag{6}$$

Where

$D=(\sqrt{A}+\sqrt{B})^2$: optimised cost for unit measurement variance

A: sampling cost for unit sampling variance
B: analytical cost for unit analytical variance
S^2_{meas} : measurement variance=S^2_{samp} (sampling variance)+S^2_{anal} (analytical va riance)

As a result, the loss function on misclassification error that calculates the expectation of total loss, E(L), of the Equation 5 can be restated as,

$$E(L)=L_{end}+L_{meas}=C_c\left[1-\Phi\left(\frac{\varepsilon}{s_{meas}}\right)\right]+\frac{D}{s^2_{meas}} \tag{7}$$

where Φ stands for standard normal cumulative distribution function(Equation 3).

2.3. Application of the probabilistic loss function as a FFP criterion

The quality objective of the investigation can be set during the designing stage according to the budget, or the regulatory compliance without reference to a loss function. In addition, an experts experience is usually sought to set a FFP criterion with which the quality of the investigation can be planned. The cost-effective design of the investigation of potentially contaminated land can then be built on the basis of the subjective FFP criterion. However, an optimization of the investigation in terms of a financial FFP criterion can only be achieved by determining a level of data quality from which a minimum total loss is incurred.

The expectation of total loss(Equation 7) that combines the loss function on misclassification error and the cost of sampling and analysis can also be used for the optimization of investigation. The each function quantifies the financial loss in terms of overall uncertainty on measured values. Minimization of the combined function(Equation 8) provides the optimum level of uncertainty that can be used as a criterion for the investigation optimization process. The uncertainty level derived from the Equation 8 incurs the minimum amount of the total expected loss(Fig. 2). The contaminated land investigation design can then be optimized to address this target uncertainty level for the measurement.

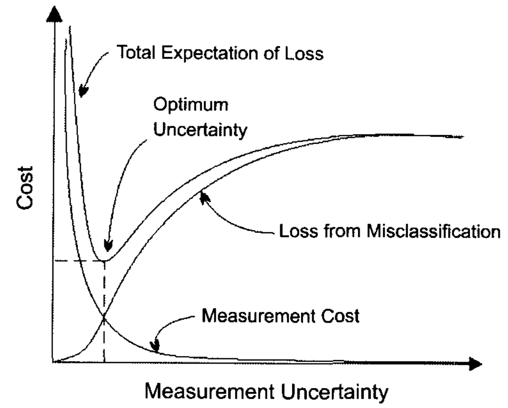

Fig. 2. Illustration for the loss function on the misclassification error and measurement cost (Equation 7). Minimisation of this function enables to locate the optimum level of measurement uncertainty for the optimisation of con-taminated land investigation.

$$\mathrm{Min}[E(L)]=\mathrm{Min}\left\{C_c\left[1-\Phi\left(\frac{\varepsilon}{s_{meas}}\right)\right]+\frac{D}{s^2_{meas}}\right\} \tag{8}$$

3. Investigation design related with mean estimation

3.1. Application of Cochrans sampling theory to investigation design

The optimization method based on the loss function on misclassification error may not be appropriate if the objective of a site investigation does not lie in the classification of the land, but simply requires the estimation of the mean concentration of the land(for example, for a site-specific risk assessment). In which case, the mean value should be estimated with accuracy and the financial loss from this investigation is therefore incurred from the sheer magnitude of error on the estimated concentration. Hence, a symmetric loss function that can describe the relationship needs to be applied for a cost-effective investigation.

Cochrans sampling theory(Cochran, 1977), which is constructed on such a symmetric loss function, allows the calculation of the optimum number of measurements to be made at a minimum total financial loss caused by the error on the estimate as well as the cost for the measurement.

Consider that the objective of the investigation on a site is to determine the site mean concentration of Pb. The investigators should, therefore, decide on the number of measurements that are needed to determine the mean concentration of the site at a minimum possible financial loss. In other words, the number of the measurements can be considered optimum if the monetary cost incurred by taking *n* samples and the financial loss arising from the error on the estimate together is at a minimum. This can be derived by using the Cochrans sampling theory as in Equation 9.

$$n'_{meas}=\sqrt{\lambda\cdot(s^2_{total}/C_{meas})} \qquad (9)$$

where

n'_{meas}:optimum number of measurements

λ: constant

s^2_{total}: variance of the site

C_{meas}: unit cost for one measurement=unit sampling cost(C_{samp})+unit analytical cost(C_{anal})

3.2. Optimal allocation between sampling and analysis

The optimal allocation method for a stratified sampling can be derived using the equation for the optimal sampling size(Equation 9), which determines the total sample size *n* and the sample size in each stratum n_h. If there is more than one stratum in a population, as is the case for a stratified sampling strategy, the total sampling cost would be $C=C_0+\sum C\cdot n_h$, where C_0 refers to the overhead cost.

The sum of the loss and the sampling cost for all the sampling strata is at a minimum when n_h is proportional to $W_hs_h/\sqrt{C_h}$(from the Equation 9). (where W_h is the stratum weight; see below equation for the explanation). Hence, when the total sample size *n* is set, the optimum sample size in the stratum *h* can be determined as in Equation 10, which is called the optimal allocation equation (Cochran, 1977).

$$n_h=\frac{W_hs_h/\sqrt{C_h}}{\sum(W_hs_h/\sqrt{C_h})}\cdot n \qquad (10)$$

where,

h : stratum

n_h : sample size for stratum h

N_h : total number of sampling units in stratum h

W_h: N_h/N, stratum weight

s_h : standard deviation of a stratum h

C_h : unit sampling cost of a stratum h

The Equation 10 can then be applied to determine the optimal allocation between the field samples and chemical analyses if each can be considered as a separate stratum. The standard deviation values(*s*) values of each stratum can be substituted for by the sampling and analytical uncertainties.

The stratum weights of W_{samp} and W_{anal} can be set as equal since the concentration value of one field sample can only be estimated through one chemical analysis and if the two strata are considered to be independent of each other. By applying the same stratum weights, the optimal ratio of field samples to chemical analysis is calculated as Equation 11.

$$\frac{n'_{samp}}{n'_{anal}}=\frac{s_{samp}\sqrt{C_{anal}}}{s_{anal}\sqrt{C_{samp}}} \qquad (11)$$

n_{samp}:optimum number of samples

n_{anal} :optimum number of analyses

s_{samp} :standard deviation of a sampling method

s_{anal} :standard deviation of an analytical method

C_{anal} :unit analytical cost

C_{samp}:unit analytical cost

3.3. Investigation design with a target data quality

Consider a situation where the target quality of the concentration estimate for a contaminated site is pre-determined by, for example, an experts fit-for-purpose criterion(FFP criterion). The expected variance as a target quality(v_{target}) is inversely proportional to the number of measurements(n_{meas}), and the relationship can be expressed as in Equation 12 where s^2_{total} stands for total variance of the measured values across the site.

$$n_{meas}=s^2_{total}/v_{target} \qquad (12)$$

If the estimated mean value from these measurements needs to be expressed with confidence, the number of the measurements(therefore number of the samples when without duplicate analysis) should be multiplied by a value of the standard

Table 1. Prior information for the optimization process for the two study sites in terms of Pb.

(unit: $\mu g\ g^{-1}$)	s_{samp}	s_{anal}	s_{meas}	s_{total}	$\bar{n}$	T	*e*	C_{samp}	C_{anal}
Nottingham	5.78	1.93	6.09	20.9	113	500	387	13	25
Coseley	173	23.9	175	359	818	500	318	15	25

s_{meas}=measurement uncertainty, $s_{meas}=\sqrt{(s^2_{samp}+s^2_{anal})}$, $s_{total}=\sqrt{(s^2_{meas}+s^2_{geo})}$, Threshold (T)=Trigger value for domestic gardens, (ICRCL, 1987), $\varepsilon=|T-\bar{n}|$, C_{samp}=unit sampling cost in £ (cost per field sample), C_{anal}=unit analytical cost in £ (cost per analysis).

normal variant(z=1.96 with a 95% confidence level). This calculation with z, however, does not consider the number of samples from which the variability is estimated. Hence, t-statistics can be introduced to address the designated confidence level(Keith *et al.*, 1996) as in Equation 13.

$$n_{meas}=(t_{1-\alpha/2})s^2_{total}/v_{target} \qquad (13)$$

Since the *t* value can only be estimated at a specified sample number(which is not yet known), iterative process is required starting with the approximated sample number generated from using *z* value at a confidence limit. This iterative process based on the Equation 13 has been incorporated in the DQO Process(USEPA, 2000).

4. Results and discussion

4.1. Prior information required for the optimisation

Different FFP criteria discussed in Section 3 are tested using the data obtained from the Coseley site and Nottingham site in the UK contaminated with heavy metals. Optimization process for a site investigation begins with gathering information about the site, upon which a new design can be built. The reliability of the process, therefore, depends on the quality of the prior information. The information consists not only of the knowledge of the site, such as the small-scale heterogeneity represented by the sampling uncertainty, but also the analytical uncertainty, and the cost of each method of measurement. This information should vary with specific contaminant, based on which the optimization for the whole investigation can be limited.

In this regard, Pb was selected as the contaminant for the application of the optimization method using the probabilistic loss function since the measured Pb concentration values of the Nottingham site were susceptible to false negative errors, and those from the Coseley site were more likely to give rise to false positive errors, based on the ICRCL trigger values(ICRCL, 1987) for domestic gardens(*i.e.* 500 $\mu g\ g^{-1}$ of Pb). The estimated uncertainty values are listed along with the mean concentrations in Table 1.

The sampling cost at each site was calculated from the actual total sampling cost divided by the number of samples(Table 1). This unit sampling cost was calculated differently for each site because of the difference in sampling conditions caused by, for example, the distance from the site to the lab facility, and the accessibility at the site and the difficulty of sampling, which eventually affected the overhead cost of sampling.

4.2. Optimization applying the probabilistic loss function

The constants in the Equation 7 such as C_c(consequence cost) and D(optimized cost per unit measurement variance) are site specific. This is because the constants reflect site characteristic features that can be derived from a preliminary investigation of the site or estimated from pre-existing information. The consequence cost(C_c), defines maximum possible loss caused by either a false negative or a false positive error. It could reflect the valuation of the property, type and concentration level of the contaminant in question. This therefore relates to the seriousness of the risk to human and/or the environment, which could be derived from the records of compensation determined by litigations on similar subject, cost of remediation for the contaminant or similar factors. In addition, the determination of the optimized cost per unit measurement variance(D) requires information on sampling and analytical cost as well as the uncertainty values produced from the sampling and analytical methods.

4.2.1.Optimization on false positive error: Coseley site

Even though the mean concentration of Pb at the Coseley site(818 $\mu g\ g^{-1}$) exceeds the threshold value(500 $\mu g\ g^{-1}$), there is the probability of making a false positive error for the use of the value. This is because the level of overall measurement uncertainty value verified from the preliminary investigation is rather high because of the heterogeneous nature of the Coseley site(Table 1). Decision-making on the basis of the data without consideration of the uncertainty of the measurement can therefore lead to a false positive misclassification error. Hence, the investigation can be optimised by balancing the uncertainty of the measurement against the expectation of total loss from the investigation using the Equation 7. Required constants for the equation are determined based on the information from the preliminary investigation (Table 1).

The consequence cost(C_c) for the false positive error defines the financial loss arising from the unnecessary remediation. The cost can be estimated for a unit area that one measured concentration represents, which is, in this case, 25 m^2 from the 5 m by 5 m grid dimension. If the dig and dump method is to be used for the removal of contaminated area at a depth of 1m, approximately £3,750 is required at the cost of £50 per ton of contaminated soil removal. Therefore, the Equation 7 can be restated after substituting the derived values for the constants.

$$E(L)=C_c\left[1-\Phi\left(\frac{\varepsilon}{s_{meas}}\right)\right]+\frac{D}{s_{meas}^2}$$

$$=3750\left[1-\Phi\left(\frac{318}{s_{meas}}\right)\right]+\frac{623351}{s_{meas}^2}$$

This equation expresses expectation of total financial loss as a function of measurement uncertainty(s_{meas}). The relationship can be plotted for varying s_{meas} values, thereby the optimum value of measurement uncertainty, which minimizes the expectation of total loss, can be estimated by inspection of the plot(Fig. 3).

This plot suggests that the actual measurement uncertainty level(s_{meas}=175 $\mu g\ g^{-1}$) derived from the preliminary investigation has room for improvement towards the optimum uncertainty level(s_{meas}'=115 $\mu g\ g^{-1}$) in order to incur the minimum total loss. In addition, the plot shows that most of the expectation of total loss comes from the loss from misclassification error. Althou-gh there is a rather large gap between the threshold and overall concentration value(ε=318 $\mu g\ g^{-1}$), the relatively high level of heterogeneity of the Coseley site could easily cause a false positive error. Consequently, even a small reduction of measurement uncertainty will reduce the misclassification error drastically. As a result of optimization of the investigation design fitted to this optimum uncertainty level, the expected loss for the unit area of one measurement represents (*i.e.* 25 m^2 from the 5 m by 5 m grid dimension) can be reduced by 61% from £150 to £57.8.

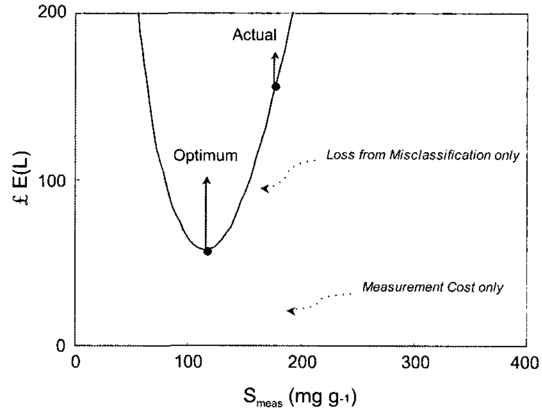

Fig. 3. Plot of the expectation of total loss for Coseley site with the individual plots of the loss from the false positive error and the measurement cost. All plots are the functions of the measurement uncertainty (s_{meas}). The actual measurement uncertainty level derived from the preliminary investigation can be compared with the optimum measure-ment uncertainty level that minimise the expectation of total loss.

4.2.2. Optimisation on false negative error: Nottingham site

There is not a significant probability of a false negative error from the investigation of Nottingham site that can be made using the mean concentration(113 $\mu g\ g^{-1}$) of the site which is far below the threshold limit(500 $\mu g\ g^{-1}$). This is because of the large gap between the two values(ε =387 $\mu g\ g^{-1}$) along with the small measurement uncertainty(6.09 $\mu g\ g^{-1}$) originating from the rather homogeneous characteristics of the contaminant distribution at the site.

In order to find the optimum measurement uncertainty level for an investigation on this site, all the required constants for the Equation 7 are determined applying the site-specific prior information listed in Table 1.

For the Nottingham site, the consequence cost (C_c) is the maximum loss arising from the false negative error incurred by, for example, compensation to shareholders settled from a possible litigation and/or by the loss from the cancellation of the project. This constant is again site and case-specific and often hard to estimate(M'Gonigle *et al.*, 1944). Nevertheless, a value of £5,000 per 25 m^2 is assumed, which is higher than the consequence cost for the false positive error of £3,750 for the Coseley site. The rationale of which is that the false negative error usually causes a more serious loss from posing harm to human health than the case of false positive where the loss can be incurred by unnecessary remediation. As a result, the Equation 7 is completed after replacing all the constants.

$$E(L)=C_c\left[1-\Phi\frac{\varepsilon}{s_{meas}}\right]+\frac{D}{s^2_{meas}}=5000\left[1-\Phi\left(\frac{387}{s_{meas}}\right)\right]+\frac{929.64}{s^2_{meas}}$$

This relationship between the expectation of total financial loss, E(L) and the measurement uncertainty, s_{meas}, is plotted in Fig. 4.

The measurement method has a plenty of room to be relaxed as the uncertainty of the measurement(s_{meas}) is relaxed considerably from 6.09 µg g^{-1} to 85 µg g^{-1}. As a result, the measurement cost reduces even without increasing the loss from the misclassification. Consequently, the expected total loss for a unit area represented by one measurement(25 m^2) could be reduced from £25.07 to £0.14.

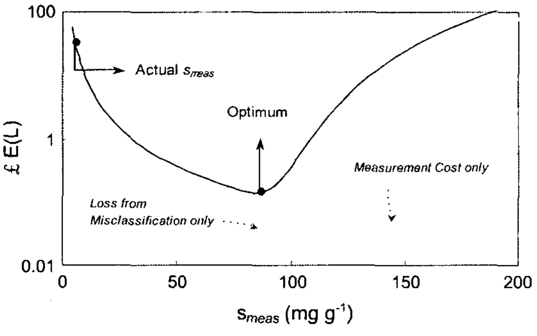

Fig. 4. Plot of the expectation of total loss for Nottingham site with the individual plots of the loss from the false negative error and the measurement cost. All plots are the function of the measurement uncertainty (s_{meas} µg g^{-1}). The actual measurement uncertainty level derived from the preliminary investigation can be compared with the optimum measurement uncertainty level that minimise the expectation of total loss.

4.3. Optimum investigation design related with mean estimation

4.3.1. Result of applying symmetrical loss function

By substituting the variables for the data from the preliminary investigation of the Coseley site (Table 1), the optimum number of measurement is calculated as

$$n'_{meas}=\sqrt{\lambda\cdot(s^2_{total}/C_{meas})}=\sqrt{\lambda(359.4^2/40)}=56.8\sqrt{\lambda} \quad (14)$$

However, in order to complete the calculation, the value for the constant λ should be determined. This can be derived from the following symmetrical loss function that the Cochrans sampling theory adopts(as described in Section 3.1).

$$L(n)=\lambda E(\varepsilon^2) \quad (15)$$

Since this equation converts the amount of error (as variance) into monetary unit, the constant(λ) thus has a unit such as £·(µg g^{-1})$^{-2}$. This constant is site specific and can be derived empirically depending on the quality objective of the investigation. However, by assuming monetary loss of £ 25 for unit variance(*i.e.* λ=25 £·(µg g^{-1})$^{-2}$), the Equation 14 can be calculated to be

$$n'_{meas}=\sqrt{\lambda\cdot(s^2_{total}/C_{meas})}=\sqrt{25(359.4^2/40)}\approx 284$$

Therefore, the mean concentration of the site determined from 284 measurements incurs minimum total loss from the estimate.

4.3.2. Investigation design with a target data quality

Assume for example that the target variance (v_{target}) on the mean(as standard error) is set at 100^2 (µg g^{-1})$^{-2}$ for the investigation of Coseley site, which amounts to about 28% of the total variability. Then the number of measurements to achieve the target quality is calculated as,

$$n_{meas}=s^2_{total}/v_{target}=359^2/(100)^2=12.9\approx 13 \quad (16)$$

Furthermore, in order to express the estimated mean value with confidence, the number of the sample should be multiplied by a value of the standard normal variant(z=1.96 with a 95% confidence level). For this, the iterative process to calculate the t value based on the Equation 13 can be made by using the Enviro-Calc computer program, which is a part of the DOQ-PRO(Keith *et al.*, 1996). The number of measurements to achieve the same data quality as in Equation 16, is then recalculated with 95% confidence to be 52.

4.3.3. Optimal ratio of sampling increments for a composite sample

The sampling uncertainty arising mainly from the small-scale heterogeneity is often larger than the analytical uncertainty. The composite sampling is therefore recommended to reduce the influence of the small-scale heterogeneity. The optimum number of increments for a composite sample can be calculated applying the Equation 11. Since analysis on each composite sample composed with different number of increments constitutes one measurement for a contaminant concentration, the cost-effective ratio between the sampling increment and chemical analysis is now calculated as,

$$\frac{n'_{samp}}{n'_{anal}}=\frac{s_{samp}\sqrt{C_{anal}}}{s_{anal}\sqrt{C_{samp}}}=\frac{173\sqrt{25}}{23.9\sqrt{15}}=9.3$$

where

n'_{samp}: optimum total number of field samples

n'_{anal} : optimum total number of analysis

This result implies that approximately nine sampling increments should be taken for each composite sample, which is the cost effective way of a measurement that creates a minimum measurement uncertainty.

5. Conclusions

Specific criteria or conditions can be applied in designing a contaminated land investigation, including, for example, a target level of data quality, a budget, or a standard method. In the absence of those specific requirements, a cost-effective investigation can be considered as an alternative for the design. A cost-effective investigation design should incur a minimum financial loss from the investigation balancing the quality of the data against the total loss.

The loss function based on misclassification error was applied to the study sites in order to test the feasibility for the cost-effective investigation design. As a result, the mean overall concentration of the Coseley site exceeding the threshold limit coupled with relatively high level of measurement uncertainty due to the heterogeneity of the site, can lead to a false positive error in classification. When the total expectation of loss curve is plotted against the level of measurement uncertainty, the degree of measurement uncertainty that needs to be reduced to lower the misclassification error can be derived. This optimum level of measurement uncertainty can be achieved by improving the sampling method, rather than analytical method. This is mainly due to the highly heterogeneous nature of the Coseley site.

By contrast, the investigation on the Nottingham site, which is relatively homogeneous and low in contamination, can produce a false negative error. Nevertheless, the result of the optimization indicates that the original measurement method is costly compared to the optimum level that does not require a high precision measurement method. It is because of the homogeneous nature of this site, which contributes towards lowering the sampling uncertainty.

In addition, the application of the loss function on misclassification error to the design of site investigations needs the determination of the consequence cost(C_c), which is the maximum loss from the misclassification. This value cannot be derived directly, as is the case with other environmental valuation. Nevertheless, the results from the optimization process will provide a valuable insight into the cost-effectiveness of the investigation design, to say the least.

Other investigation can be carried out to estimate the mean concentration of contaminant at the site, rather than classifying it. In this case, the cost-effective investigation design requires a different loss function, such as a symmetrical loss function, that calculates the total loss on the basis of the error on the mean estimation. Cochrans sampling theory, which is built on a symmetrical loss function, was applied to derive the optimum number of measurements that produce minimum expectation of total loss. However, this loss function requires a constant(*i.e.* λ, which converts the error term into monetary units) that cannot be derived directly. Consequently, it also requires a subjective or an empirical approach to the determination of the constant.

Investigations design based on a target data quality(set as an error on the estimated mean concentration of the site) is relatively simple. The number of measurements can be easily derived to achieve the target quality. Nevertheless, if the result from the measurements is to be produced with statistical significance, the number of measurement needs to be recalculated with additional t-statistic that requires an iterative process. Software developed for a part of USEPAs DQO-Process(*i.e.* DQO-PRO) was found to be fast and effective for the iteration.

Finally, the optimal allocation method has been applied to the optimization of composite sampling method. This calculation of the ratio has been derived from the cost-rule and similarly from the Cochrans stratified sampling theory. As a result, the optimal ratio was calculated for the composite sample increments that are subjected to one chemical analysis. Consequently, each measurement can be carried out at a minimum measurement uncertainty in a cost-effective way.

In conclusion, all optimization process in designing an investigation requires prior knowledge or information on the site(and/or assumptions in establishing constants used in the equations), upon which the successful investigation design lies. Nevertheless, the application of these methods offers more objective FFP criteria for the design of the investigation than the experts subjective judgment of fitness-for-purpose.

References

Cochran, W.G. (1977). Sampling techniques. 3rd ed. Wiley series in probability and mathematical statistics. Wiley, New York.

Garrett, R.G. and Goss, T.I. (1979). The evaluation of sampling and analytical variation in regional geochemical surveys. Geochemical exploration 1978: Proceedings of the seventh International Geochemical Exploration Symposium (Golden, Colo). Watterson, J.R. and Theobald, P.K. Eds. Association of Exploration Geochemists, c 1979, Rexdale, Ontario, p. 371-383.

ICRCL (1987). Guidance on the assessment and redevelopment of contaminated land. Interdepartmental committee on the redevelopment of contaminated land, DOE, UK, London.

ISO (1993). Guide to the expression of uncertainty in measurement. 1st ed. International Organisation for Standardisation, Geneva, Switzerland.

Keith, L. H., Patton, G. L. Lewis, D.L. and Edwards, P.G. (1996). Determining what kinds of samples and how many samples to analyze. Principles of Environmental Sampling. Keith, L.H. Ed. American Chemical Society, Washington, DC.

Lee, J.C. and Ramsey, M.H. (2001). Modelling measurement uncertainty as a function of concentration: an example from a contaminated land investigation. Analyst, v. 126. p. 1784-1791.

M'Gonigle, R., Jamieson, T.L. McAllister, M.K. and Peterman, R. (1944). Taking uncertainty seriously: from permissive regulation to preventative design in environmental decision making. Osgoode Hall Law Journal v. 32, p. 99-168.

Miesch, A.T. (1967). Theory of error in geochemical data. U.S. Geological Survey Professional Paper Part A, 574p.

Ramsey, M.H. (1997). Measurement uncertainty arising from sampling: Implications for the objectives of geoanalysis. Analyst. v. 122, p. 1255-1260.

Ramsey, M.H. (1998). Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources. Journal of Analytical Atomic Spectrometry v. 13, p. 97-104.

Thompson, M. (1995). Uncertainty in an uncertain world. Analyst v. 120, p. 117N-118N.

Thompson, M. and Fearn, T. (1996). What exactly is fitness for purpose in analytical measurement?. Analyst v. 121, p. 275-278.

USEPA (2000). Guidance for the data quality objectives process, EPA QA G-4. EPA/600/R-96/055. United States Environmental Protection Agency, Washington DC.

2003년 5월 26일 원고접수, 2003년 6월 19일 게재승인.