Incorporation of Gy's theory of sampling in the investigation of contaminated soil in China

J. Song[1], R. Brewer[2], W. Tang[1], R. Liu[1], and Y.M. Luo[1,3]

[1]University of Chinese Academy of Sciences, China
[2]Hawaii Department of Health, United States of America
[3]State Key Laboratory of Soil & Sustainable Agriculture, China

With a legacy of rapid industrialisation, economic growth and urbanisation over the past 40 years, contamination of agricultural soils and industrial sites has become a major public concern in relation to food safety, human and ecosystem health and China's sustainable development. Characterisation of contaminated soil at industrial sites traditionally involved the collection and testing of discrete grab samples from a single location or a sample composed of 5-9 increments collected from within a targeted area and volume of soil. The sample data are then interpolated using various geostatistical tools to derive soil contamination distribution maps. While seemingly cost-effective, this approach very often led to incorrect predictions of the lateral and vertical extent of contamination and significant uncertainty in estimation of mean contaminant concentrations for assessment of risk. Pierre Gy's pioneering work on the Theory of Sampling (TOS) for particulate matter offered a pathway forward to address these errors and laid the foundation for the development of Decision Unit and Multi Increment Sampling methods (DUMIS) in the environmental industry. Several environmental agencies have published policy or technical guidance on the use of TOS for sampling of contaminated soils. The technical document published by the Hawaii Department of Health (HIDOH) in the United States is perhaps the most comprehensive and up-to-date. In collaboration with HIDOH, the Institute of Soil Science, Chinese Academy of Sciences (ISSCAS) has conducted field research at dozens of agricultural and industrial sites in China. Based on the experience gained in both the US and China, ISSCAS successfully proposed and drafted Chinese national guidelines for DUMIS investigation methods, which came into force in October 2023.

Keywords: Contaminated site, spatial variability, Decision Unit, Multi-Increment Sample, data quality, Theory of Sampling, TOS

INTRODUCTION

In order to understand the evolution of environmental sampling methods in China over the past 30 or more years, it is necessary to consider the economic and social evolution of the country itself. Over the past 40 years, China has experienced a booming economy, rapid industrialisation and urbanisation. This has brought great economic and social improvements to society and created a strong and growing middle class. However, the economic growth was partially achieved at the cost of resource depletion and environmental damage. As a consequence, contamination of industrial sites and agricultural soils has become a major public concern for human health, food safety, ecosystem health and sustainable development of China.

Serious concern over health and environmental impacts required development of guidance to investigate, assess and remediate the contamination. The Chinese Law on Soil Pollution Prevention and Control mandated conducting a nationwide soil pollution survey every ten years. The first nationwide soil pollution survey was conducted between 2005 and 2013. It not only revealed the overall soil contamination status across China, but also laid the foundation for the development of a series of technical guidelines in relation to environmental site investigation (HJ 25.1-2019)(MEE, 2019a), monitoring(HJ 25.2-2019)(MEE, 2019b), human health risk assessment (HJ 25.3-2019)(MEE, 2019c), remediation of soil (HJ 25.4-2019)(MEE, 2019d) and groundwater (HJ 25.6-2019)(MEE, 2019e) and verification of contaminated sites (HJ 25.5-2018)(MEE, 2018). The national guidelines were released in 2014 and updated in 2019.

Laws and guidance adopted by countries that had earlier passed through this stage of economic development were quickly adopted and assumed to be scientifically sound. Early guidelines for the characterisation of soil contamination at industrial sites called for the collection of discrete grab samples or 'composite' samples with 5-9 increments, based in part on guidance published by the United States Environmental Protection Agency (USEPA) in the 1980s and 1990s (e.g., USEPA 1985, 1986, 1989, 1991, 1992a,b) . A review of the origin of USEPA soil sampling guidance is provided by Brewer *et al*., (2017a,b). The data were then interpolated using various geostatistical tools to generate soil contamination distribution maps. While seemingly cost effective, the approach has been found to have poor data repeatability and very often led to incorrect prediction of the boundaries and soil concentrations in the contaminated area. This is evidenced by numerous case studies where field duplicate samples failed the sample variability test; multiple isolated 'hot' and 'clean' spots exist on the site and sample exceedances occur at the boundaries of cleanup areas (e.g., Brewer *et al*., 2017a,b).

PROBLEMS WITH EXISTING SOIL SAMPLING METHODS

USE OF FIELD REPLICATE SAMPLES

In the Chinese technical guidelines for monitoring during risk control and remediation of soil contamination of land for construction (HJ 25.2-2019) (MEE, 2019b), field duplicate samples are defined as samples collected from the same location and individually sealed and analysed. This definition is similar to that in a document published by the Nevada Department of Conservation & Natural Resources in 2008, where a field duplicate is defined as a distinct, discrete grab sample collected from the same point in time and space as the first sample, or as near to the same time and place as possible (NDEP, 2008). In the field quality control standard operating procedure (SOP) published in 2007 by the Maine Department of Environmental Protection (MDEP), duplicate grab samples of gasoline range organics and volatile organic compounds (VOCs) should be co-located (duplicate sample immediately adjacent to the original sample) and other types of samples may be combined ("composited"). The guidance states that the combined soil should be "mixed thoroughly and placed in an appropriate sample container) (MDEP, 2007). Samples that are subdivided into separate portions for testing by different laboratories are referred to as "split samples" in other states' SOP documents.

Apparently in these documents the purpose of collecting the soil sample and duplicate sample from the same location or from co-located locations is to confirm an acceptable level of spatial variability in sample data. However, in practice, duplicate samples do not always meet the required relative deviation criteria for intra-laboratory duplicates, for example $\leq$25% for inorganics, $\leq$40% for semi-volatile organic compounds (SVOCs) and $\leq$65% for (VOCs) (MEE, 2022). Variability and potential data error for a given point is likewise impossible to assess with the small number of co-located samples collected, typically just two. In some cases, original and duplicate samples can lead to opposite conclusions, with data for one sample above the soil standard and data for another below the soil standard. In such cases, the sample with the highest concentration of the contaminant is normally used for decision making, even though the degree of error in the data (and decisions) is unknown. Most importantly, soil standards are typically developed to address chronic exposure and health risk and apply to the mean of the contaminant concentration for a targeted exposure area in question as a whole, rather for individual points.

To address this dilemma, additional technical regulation have been introduced in China. As part of the field quality control measures, the new regulation requires the collection of triplicate samples (collected from the same location or co-located points) to be analysed by two independent laboratories. The quality of field sampling is assessed in two steps, namely category assessment and relative deviation assessment. As long as the same conclusion (compliance or non-compliance with the soil standard) can be drawn from replicate samples, field sampling quality is considered acceptable and no relative deviation assessment is needed. Otherwise, relative deviation (RD) limits for intra- and inter-laboratory replicates should be calculated and compared against the allowable relative deviation (MEE, 2022). The two-step assessment method simplifies the quality control of field sampling, as the investigators are no longer required to evaluate possible reasons for potentially large variations between field replicates as long as the replicate data fall into the same category (contaminated or uncontaminated). This issue was recently discussed in relation to the 'Assay Exchange' paradigm, Esbensen & Vogel (2023), which brings important, critical context to the issue of over-emphasising analytical result comparisons only.

Although the new regulation eases off the quality control of field sampling, the underlying problem with the overall unreliability and representativeness of the discrete grab sample data remains. A field study by Brewer *et al.,* (2017a) found that random variability of contaminant concentrations within a single sample at the scale of a typical laboratory aliquot as well as random variability between discrete grab samples collected less than one-meter apart can be in the range of 1-3 orders of magnitude, depending on the chemistry and form of the contaminant. One could not help but doubt the representativeness of a single discrete grab sample. This has significant implications for the use of discrete grab sample data to delineate the lateral and vertical extent of contamination at a site. The study also demonstrated significant variability in the mean concentration of a contaminant estimated for a targeted area based on comparison of replicate sets of discrete grab sample collected from the area. This calls into question the reliability of discrete grab sample data in environmental risk assessments.

In the case of so-called "composite" samples, as mentioned above, split samples are often collected to assess the quality of the field sampling. Again, what the split sample can't say is the representativeness of the final sample itself. This can only done be comparing the method in which the sample was collected to Theory of Sampling requirements and comparison of replicate samples prepared by combining and testing independent set of increments.

LACK OF QUALITY CONTROL FOR LAB SUBSAMPLING

In the existing Chinese guidelines for environmental monitoring of farmland soil (NY/T 395-2012) (MOA, 2012) or soil in general (HJ/T 166-2004) (SEPB, 2004), no attempts are made to quantify the errors introduced during laboratory subsampling process. For heavy metal analysis (Hg excluded), field samples are air dried, sieved to 2 mm and subsampled using the quartering method to pass through 0.15 mm. Errors during quartering are rarely assessed, which may well be highly problematic, as the error associated with coning-and-quartering is known to be considerably higher than the analytical error. This is especially true if the particle-size distribution varies between different quarters, as contaminants are known to be concentrated in fine particles. For analysis of non-volatile organic contaminants, a second sample is used and no pulverisation is required. A random scoop of moist soil (i.e. grab sampling) is weighed and extracted and another portion is weighed and oven dried to measure moisture content. Sample processing methods to ensure that subsamples are representative of the original sample as a whole are not clearly described or required in the guideline. The mass of the subsample is based on the mass required by the laboratory for analysis, rather than by such concepts as the fundamental sampling error (FSE). In other words, the mass of the subsample should be based on the acceptable error associated with the collection of a subsample. The latter is at the core of the rationale of correct sampling as described in the extensive literature on TOS (e.g. Gerlach *et al*., 2002, 2003, Minnitt and Esbensen, 2017; Minkkinen and Esbensen, 2019; Pitard, 2019; Dubé and Esbensen, 2022).

Testing for SVOCs offers an example. Due to the hydrophobic nature of SVOCs, large variability in SVOC concentration can exist within the soil in the sample container. This is evidenced by potentially

large differences in soil SVOCs data analysed multiple times on customers' request. Unfortunately, evaluation of the errors associated with discrete grab sampling is not required in the existing Chinese technical guidelines. In general, the primary sampling errors are significantly larger than the secondary sub-sampling errors, with the analytical errors very nearly always significantly small. It is concerning that the preceding sampling and sub-sampling errors still receives so little attention.

ERROR IN INTERPOLATION OF DATA FROM DISCRETE SAMPLING

In site investigation, spatial analysis tools are often used to interpolate discrete data and produce isoconcentration maps. Methods commonly used for spatial interpolation include inverse distance weighting, kriging, Thiessen polygon and trend surface. The fundamental hypothesis of spatial analysis is that the closer the points are in space, the more likely they are to have similar values. However, due to the differences in the algorithms used, different interpolation methods can produce significantly different mapping results, which causes tremendous uncertainties for remediation projects.

The selection of an interpolation method for different types of soil contaminants depends on factors such as the spatial characteristics of the data, the presence of spatial correlation, and the specific objectives of the analysis. The reliability of isoconcentration mapping programs is based in particular on discrete sample data depending on two key assumptions (Brewer *et al.,* 2017b): 1) Data for a single location are reasonably reliable for the immediately surrounding area in general and 2) Contaminant trends between points is linear. Random variability between co-located discrete samples highlights potential error associated with both of these assumptions. The problem is the scale or the spatial resolution used in a typical site investigation is typically 10 to 20 metres. In contrast, the variability in soil contaminant concentrations exists at scales of millimetres or centimetres.

Within the community of professional samplers, these critical issues are being addressed by systematic use of variographic data analysis, allowing consideration of all scales simultaneously (from back-to-back duplicate samples to half the largest scale characterising the decision unit adopted.

Potentially significant errors in the use of sampling methods presented in earlier guidance have been warned by Brewer *et al*., (2017a) who showed in detailed field studies that the differences in contaminant concentrations in discrete samples placed a few inches apart can be 1-3 orders of magnitude. Therefore, discrete grab sampling data are not reliable for predicting small-scale distribution patterns of soil contamination at a site. This is due to the inherent susceptibility of discrete grab sample data to potentially large fundamental error (FE) and grouping and segregation error. Furthermore, from a risk assessor's perspective, it is the average concentration in the exposure area that determines the exposure and associated health risk, not the point concentration. The use of the so-called maximum concentration for site investigation and risk assessment can lead to unnecessary remediation of clean areas or to the approval of a falsely clean site only to find it heavily contaminated at a later stage. Many of these deficiencies can be overcome by diligent use of geostatistics.

APPLICATION OF THEORY OF SAMPLING

The first step in addressing the unreliability of sample data collected in the environmental industry was a methodical review of the origin of these errors. As part of this process, the importance of understanding contaminant heterogeneity in terms of Pierre Gy's TOS quickly became evident (e.g. Gerlach *et al*., 2002, 2003, Minnitt *et al.*, 2007; Minnitt and Esbensen, 2017; Minkkinen and Esbensen, 2019; Pitard, 1993, 2019; Dubé and Esbensen, 2022; see also Brewer *et al*., 2019, 2022). Primary sources of error in environmental data were determined to include (after Minnitt *et al*., 2007):

- Nugget effects
- Fundamental sampling error
- Grouping and segregation error
- Long-range heterogeneity error
- Increment collection error
- Sample processing error; and

- Laboratory subsample collection error.

The first three factors are associated with variability between individual particles, referred to as 'compositional heterogeneity.' The fourth factor is associated with heterogeneity between different areas of a targeted area and volume of soil or other particulate matter. The latter factors consider the collection of a sample in the field and the collection of a subsample for analytical testing in the laboratory. Errors associated with these factors are hinted at in early environmental guidance but not adequately addressed.

Collection of a Representative Field Sample

Collection of a representative sample of soil in the field requires measures to adequately capture and represent 'long-range heterogeneity,' critical to the mining industry, as well as measures to minimise error associated with the collection of sample increments. Long-range, 'distributional' heterogeneity is addressed by the collection of a large number of 'increments' ("multiple increments") of soil within a targeted Decision Unit (DU) area and volume of material in a systematic, random manner and combination of the increments to prepare a single 'Multi Increment' sample. (Multi Increment is a registered trademark of EnviroStat, Inc.).

A DU is an area and volume of soil or other particulate media about which a decision is to be made. DUs are designated based on two criteria: 1) Characterisation of areas frequented by human or other receptors and assessment of health risk ('exposure area' DUs) or 2) Isolation of suspected areas of heavy contamination in order to optimise remediation ('source area and boundary' DUs). A detailed review of DU designation schemes for different site scenarios is provided by HIDOH (2023). Field studies and comparison of replicate sample data have demonstrated that a minimum of 30 to 75 increments are necessary to adequately capture distributional heterogeneity within a targeted area and volume of soil (Brewer *et al.*, 2017a,b; HIDOH, 2023). Preparation of a 50-increment Multi Increment sample is recommended as a default. The collection of a minimum one- to three-kilogram sample in the field is recommended in order to address both FE and increment collection error.

This approach allows the inherent, distributional heterogeneity of contaminant distribution within a targeted DU area and volume of soil to be captured by the sample collected. The concept of distributional heterogeneity was overlooked in early USEPA soil sampling guidance (USEPA, 1985): The… level (of contamination) is assumed to be uniform within (a contaminated area) and zero outside it. This ultimately flawed assumption is repeated in subsequent guidance documents, for example (USEPA, 1989): When there is little distance between points it is expected that there will be little variability between points. Such assumptions were never tested in the field. It is believed that the authors of the original USEPA guidance were experienced in testing of dissolved-phase contaminants in industrial wastewater (refer to Brewer et al, 2017b). Contaminant concentrations in wastewater can be normally be assumed to be relatively uniform provided that the process inside of the facility does not change. The authors prematurely applied this same concept to the distribution of contaminants in soil that had been contaminated by liquid wastes.

Actual random variability of contaminant concentrations both within and between individual grab samples imposed unseen but potentially significant error in interpretation of the extent of contamination (Brewer et al., 2017a,b). Multiple warnings regarding the unreliability of the proposed sampling methods were given in USEPA as well as other independently published papers at the time but in large part ignored, for example (USEPA, 1992a): Grab samples or judgmental samples lack the component of correctness; therefore, they are biased. The so-called grab sample is not really a sample but a specimen of the material that may or may not be representative of the sampling unit. Great care must be exercised when interpreting the meaning of these samples. These and other warnings throughout the 1990s and early 2000s were, however, largely ignored.

Collection of a Representative Laboratory Subsample

Guidance for processing and testing of discrete grab samples at a laboratory recognises the likely presence of distributional heterogeneity within a sample but simplistically recommends mechanical

'homogenisation' of the sample and selection of a random, one- to ten-gram subsample for testing (USEPA, 2018). Testing of replicate subsamples is recommended to test the reproducibility of the sample data. Grinding is recommended to further 'homogenise' the sample if the precision of replicate subsample data is unacceptable. No mention is made of the use of a sectoral splitter or similar equipment to improve subsample data collection. In practise, samples are rarely ground and the higher of two duplicate subsamples is simply used for decision making.

Concepts of FE and associated requirements for minimum analytical subsample mass are also not discussed. Subsample mass is instead based on the minimum mass necessary to carry out the requisite analysis. Competition between environmental commercial laboratories to reduce cost leads to testing of a subsample mass as small as half a gram. Pitard (1993) highlights the error in this approach, especially when small, high-concentration nuggets (grains) might be present in the sample: As samples (i.e., laboratory subsamples) become too small, the probability of having one of these (high-concentration) grains present in one selected sample diminishes drastically; furthermore, when one grain is present, the estimator … of the true unknown average… becomes so high that it is often considered as an outlier by the inexperienced (sic) operator.

This fact did not go unrecognised by other researchers in the USEPA but was largely ignored (USEPA, 1999): True and complete homogeneity is a matter of scale and is impossible to achieve for particulates because many factors, including gravity, work against it. But the extent of heterogeneity and its effect on environmental sampling can be minimised. Established methods from the mining industry are applicable to the sampling of soils. The work of George Matheron, father of geostatistics, and Pierre Gy, sampling expert, can provide useful insights for environmental scientists who are faced with sampling a complex matrix for trace contaminants. Brewer *et al*., (2017a,b) similarly identified the potential high variability and randomness of subsample data when nuggets of contamination are present in the sample.

Errors associated with compositional heterogeneity and correct collection of a representative, analytical subsample at the laboratory are addressed by considering the FE and requirement of a minimum ten-gram subsample mass for testing. The latter in part reflects a target maximum FE and maximum particle size of two millimetres (default maximum particle size for assessment of direct exposure and incidental ingestion of soil). Consideration of FE and compositional heterogeneity of individual particles comprehensively takes into account such factors as particle shape, particle density, contaminant liberation and related factors. Grinding of samples is recommended when acceptable for project-specific data quality objectives. Use of a sectoral splitter is recommended, when possible, for the collection of analytical subsamples. Manual collection of a subsample from a ground sample or wetted, unground subsample using a two-dimensional slab cake and Multi Increment collection methods is otherwise recommended, given the current lack of sectoral splitters at most commercial environmental laboratories (refer to HIDOH, 2023).

<u>Estimation of Contaminant Mean</u>

Consideration of TOS allows for the direct estimation of the mean (true) concentration of a contaminant in a targeted area and volume of soil. Reliance on discrete grab sample data, in contrast, introduced significant error into estimation of mean contaminant concentrations for assessment of exposure risk. Error in the representativeness of a single set of sample data was not routinely tested through the collection of replicate sets of data. Misguided recommendations to exclude 'outlier' data in risk assessment guidance for estimation of a mean concentration exacerbated potential error (USEPA, 2013): The inclusion of outliers in the computation of the various decision statistics tends to yield inflated values of those decision statistics, which can lead to incorrect decisions. Often inflated statistics computed using a few outliers tend to represent those outliers rather than representing the main dominant population of interest (e.g., reference area). Outliers represent observations coming from populations different from the main dominant population represented by the majority of the data set. Outliers distort most statistics (e.g., mean, UCLs, UPLs, test statistics) of interest. Therefore, it is desirable to compute decision statistics based upon data sets representing the main dominant

population and not to compute distorted statistics by accommodating a few low probability outliers (e.g., by using a lognormal distribution).

Earlier authors had clearly warned against excluding high-concentration data from estimates of the contaminant mean (USEPA, 1989): This document recommends that all data not known to be in error should be considered valid… High concentrations are of particular concern for their potential health and environmental impact. And (USEPA, 2002): There are a variety of statistical tests for determining whether one or more observations are outliers. These tests should however be used judiciously. It is common that the distribution of concentration data at a site is strongly skewed so that it contains a few very high values corresponding to local hot spots of contamination. The receptor could be exposed to these hot spots, and to estimate the exposure point concentration correctly it is important to take account of these values. Therefore, one should be careful not to exclude values merely because they are large relative to the rest of the data set.

The importance of including high-concentration points for estimate of a mean within a subject volume of particulate media was, in contrast, well recognised in the mining industry (Pitard, 2009): A common error has been to reject 'outliers' that cannot be made to fit the Gaussian model or some modification of it as the popular lognormal model. The tendency, used by some geostatisticians, has been to make the data fit a preconceived model instead of searching for a model that fits the data… It is now apparent that outliers are often the most important data points in a given dataset. Such misconceptions delayed the advancement of sampling methods in the environmental industry until relatively recently.

DEVELOPMENT OF RISK-BASED SOIL SAMPLING METHODS OUTSIDE CHINA

Misconceptions of a hypothetical uniform distribution of contaminants in soil and recommendations to omit 'outlier' sample data for calculation of a mean led to significant project delays and cost overruns for site investigation, remediation and land redevelopment as well as uncertainty in the remaining risk of human health and the environment. This lapses in judgement were due to multiple factors, including (see Brewer *et al*., 2022): 1) The lack of third-party confirmation of sample data representativeness; 2) Requirements for the collection of discrete samples in existing regulatory guidance; and 3) The lack of economic or legal repercussions for entities granted the authority for determining sample collection methods (i.e., government regulatory agencies). This contrasts sharply with the mining industry, where the estimate mean concentration of a commodity in a targeted volume of ore is routinely tested by physical extraction of the commodity and where poor sampling methods directly affect the economic viability of the company.

Brewer *et al*., (2017b) provides an in-depth review of the slow but progressive development of environmental sampling guidance. TOS-based tools such as DUs and Multi Increment samples (MIS) can be thought of as 'risk-based' investigation methods (Brewer *et al*., 2022). A DU is defined as an area and volume of soil or sediment about which a decision regarding risk and/or remediation is to be made based on sample results. MIS refers to a sample prepared by the collection and combination of multiple increments of soil from a single DU. An MIS is required to have a minimum mass and must be processed in accordance with Gy's sampling theory to produce representative data (HIDOH, 2023). In 1999, the USEPA issued a memo stating that TOS, developed for the mining industry, can provide environmental scientists with guidance for the correct sampling and subsampling of soil (USEPA, 1999). The memo noted that TOS is also applicable to some sampling events at hazardous waste sites and to the successful subsampling of those samples at the analytical laboratory. In 2006, Appendix A of the USEPA Methods 8330B for the determination of nitroaromatics, nitramines, and nitrate esters recommended the use of MIS investigation methods to obtain a representative field sample and laboratory analytical subsample for the determination of energetic material residues. The use of discrete sampling methods is discouraged due to the inability to control error in collection and processing of samples (USEPA, 2006). This reflects a better understanding of the inherent distributional and compositional heterogeneity of contaminants in soil and the need to consider the correctness of sample and subsample collection methods as well as mass. In 2011, the USEPA recommended what they referred to as 'incremental

sampling' for the investigation of dioxin-contaminated sites (USEPA, 2011). In 2019, the USEPA approved the use of such sampling approaches at polychlorinated biphenyl clean-up sites (USEPA, 2019). The US Army Corps of Engineers (USACE) also recommended the use of multi increment-type for the investigation of military sites contaminated with munitions and metal residues (Walsh, 2009; Clausen, *et al*., 2013). In addition to federal agencies, several US states have also developed DUMIS-based soil sampling methods since 2008. Among these, the technical guidance developed by the Department of Health, Hawaii is the most systematic and compliant with Gy's TOS (HIDOH, 2023).

DEVELOPMENT OF TOS-BASED SAMPLING METHODS IN CHINA

The development of DUMIS in China started in 2017 as a collaboration between the Institute of Soil Science, Chinese Academy of Sciences (ISSCAS) and the Department of Health, Hawaii, USA. During the period of 2017-2018, dozens of field studies were conducted to test the applicability and optimisation of DUMIS for soil (and crop) sampling in agricultural fields and industrial sites. These field studies cover the following scenarios:

(1) Sampling to derive the environmental background concentration of elements in pasture soils on the Qinhai Plateau;
(2) Sampling to derive the environmental background concentration of elements in natural and cultivated soils of Shenzhen, Guangdong Province;
(3) Investigation of metal contamination in the soil-plant system at contaminated agricultural fields in Tongling, Anhui Province;
(4) Monitoring of long-term effects of soil additives on the plant and soil quality;
(5) Investigation of contamination of farmland adjacent to a steel work;
(6) Sampling and testing of soil for arbitration of lawsuit;
(7) Post-remediation confirmation sampling in agricultural fields contaminated by heavy metals (e.g. Mao *et al*., 2021);
(8) Post-remediation confirmation sampling at industrial sites contaminated by metals (e.g. Zhao *et al*., 2021) and organic contaminants (e.g. PAHs, TPH, BTEX and chlorinated solvents).

In 2018 and 2019, training courses on sampling for defensible decision-making were organised, which were well attended by interested parties from academia and the soil remediation industry across China. In 2019, a consortium led by ISSCAS was established to draft the national standard Soil Quality Decision Unit Multi Increment Sampling tasked by the Standardization Administration of China (SAC). The consortium consisted of research institutes from the Chinese Academy of Sciences, Ministry of Ecology and Environment, Ministry of Agriculture and Rural Affairs, Ministry of Land Resources, provincial environmental and agricultural monitoring centres, commercial laboratories, as well as consultants and remediation companies. Concerns focused on the expense and lack of a clear endpoint for investigation, assessment and remediation of contaminated land in the US and other western countries. This included uncertainty and unknown error in estimation of mean contaminant concentrations to assess risk and guide remedial actions. Concern also arose over reports of properties previously declared 'clean' that were later found to be still contaminated, so additional testing was carried out. Research and training in TOS confirmed that such errors and uncertainty were related to the collection of samples in the field and processing at the laboratory, rather than in the laboratory analysis itself or statistical tests employed in an attempt to make sense of the discrete sample data generated.

Experiences and lessons learned from the US and more recently China were incorporated into the Chinese DUMIS guidelines. During preparation of draft guidance and public consultation, misunderstandings and concerns about DUMIS were noted and carefully addressed. For example, some reviewers incorrectly believed that discrete sample data were required to identify 'hot spots' of heavy contamination. This reflected several misconceptions. The first was that published, risk-based screening levels for contaminants applied to any given mass of soil within a targeted exposure area. Published risk assessment guidance clearly states, however, that risk is to be assessed based on the mean

contamination concentration for the total area and volume of soil included in a pre-designated exposure area (e.g., USEPA 1988, 1989, 1991, 1992a).

In other cases, risk assessors versed traditional statistical methods and with no training in TOS were initially concerned that MIS methods would mask ('dilute) high-concentration areas of contamination and underpredict environmental risk. This highlights the need to establish well-thought-out 'exposure area' DU areas and volumes of soil prior to the collection of samples in the field. Remediation experts expressed concern that exposure area DUs could be very large and result in unnecessarily high remediation costs if a significant volume of otherwise clean soil was inadvertently included in remedial actions. This is addressed in the DUMIS guidance by placing an emphasis on a thorough background study of the site prior to sample collection and designated of smaller, targeted 'source area' and 'boundary' DUs to help isolate and characterise suspect high-concentration areas of contamination. This allows identification and remediation of localised areas of heavy contamination in a manner that both optimises the time and effort required for remediation and provides high confidence that concerns regarding the risk to human health and the environment have been addressed. A large variance in the relative standard deviation (RSD) of field triplicates in the DUMIS sampling can be indicative of significant distributional heterogeneity within a DU being sampled.

Other reviewers were concerned about the time and cost involved in accrediting laboratory SOPs adapted to DUMIS samples. It is critical that standards for processing and testing samples reflect sound science, however. Otherwise the effort taken to collect representative samples in the field is wasted. Research institutions including the ISSCAS have basic laboratory SOPs in place and assisting commercial laboratories in preparation of SOPs for their work. This includes methods for management and processing of large samples as well as for the collection and preservation of samples to be tested for VOCs. This will allow laboratories to progressively become accredited in DUMIS sampling methods and help assure entities submitting samples to the laboratory for testing that the data generated are reliably representative of the samples provided.

DUMIS investigation methods provide a more robust, reliable, cost-effective and defensible approach for the risk-based site investigation, assessment, remediation and project completion. In March 2023, the Chinese DUMIS national standard was officially approved by SAC and became effective in October 2023 (SAC, 2023). In 2022, DUMIS was also recommended in the Jiangsu Provincial Standard for Investigation of Soil Contamination at Lead-Acid Battery Site (DB32/T 4426-2022)(MRBJS, 2022). It is now being recognized that DUMIS environmental investigation methods are critical for reliable assessment of all types of contaminants in soil and other particulate media in China. These methods are being used to address a wide range of environmental contaminants, including heavy metals, pesticides, PCBs, PAHs, solvents, explosives and petroleum. Alternative collection methodologies have also been developed for the collection of a sample to be tested for volatile contaminants, including placement of increments (refer to HIDOH 2023). Examples scenarios where DUMIS investigation methods are being used to characterise and remediate contaminated soil as well as sediment include:

- Decommissioned textile manufacturing mega-complexes;
- Metal processing facilities;
- Asphalt batch plants;
- Former munitions ranges;
- Investigation of soil contamination with BTEX and chlorinated solvents at an industrial site;
- Operating lead acid battery site;
- Long-term monitoring of the effects of soil additives;
- Investigation of soil contamination in terraced fields;
- Restoration of contaminated agricultural lands;
- Assessment of natural background concentrations of heavy metals in soils; and
- Assessment of area-wide, anthropogenic contamination in densely populated urban areas.

Although DUMIS-based soil sampling protocol is now available in China, training and widespread use of the methods across the environmental industry in China will take some time. Stronger efforts are needed to disseminate Pierre Gy's TOS and DUMIS guidance to academia, practitioners in the industry and competent authorities. It is of vital importance to reach a national-level, scientific consensus with regard to conversion from discrete sampling to DUMIS.

From our experience with DUMIS presentations and training courses in China, academia and remediation companies are the people who easily understand and willingly accept DUMIS. It is no surprise that remediation companies are supporting DUMIS as they are on the front lines dealing with uncertainties and associated technical, economic and legal risks caused by reliance on discrete grab sample data. Environmental authorities who may not have a strong background in science and lack practical experience in the field, and who oftentimes do not directly face repercussions for erroneous data and decisions regarding health risk and remedial actions, as well as time and cost overruns faced by the party paying for the investigation and cleanup, are in some cases more reluctant to adopt DUMIS. The main concerns about adopting DUMIS are not so much technical as regulatory and agency manager's fear of calling into question past regulatory decisions and a requirement to expend funds to retest sites that have already been declared "clean.". The same lack of training and fear of diverging from status quo has plagued USEPA upper management for over thirty years, as attested by a reluctance to hold high-level discussions of mistakes in past guidance and issuance of policy documents requiring nationwide training of USEPA staff and a rapid transition to DUMIS investigation methods. This highlights the need for concurrent training of regulatory authorities in China, as well as private environmental organisations and professionals in the legal and financial industry in order to hasten the transition to more science-based sampling.

Responsible parties are worried about the increased cost and time spent on the collection of large numbers of increments across the decision unit, especially the increased cost of drilling more boreholes during subsurface sampling. This will change as the inefficiencies and financial and legal costs associated with delays in site cleanup and the discovery of additional contamination during or after redevelopment of a property becomes more apparent. Recently, there have been several environmental incidents and lawsuits in China due to soil contamination discovered prior to the transfer of land/apartments to schools or residents. Environmental authorities are becoming more resilient mainly because of the potential management consequences and possible liability for their agencies. Concerns over the need to re-investigate sites that were previously declared 'clean' by a regulatory agency based on past sampling methods needs to be addressed. The introduction of a transition period by the state of Hawaii in the USA and other regulatory agencies that have adopted TOS methods can be referenced in this regard.

CONCLUSIONS

The DUMIS protocol has been available in China since 2023. This represents a first step forward for the wider acceptance of TOS and the application of DUMIS in the environmental industry in China. Examples of DUMIS methods serve as demonstrations of scientifically sound, fast, reliable and relatively inexpensive approaches for investigating, assessing and remediating contaminated industrial lands. Expanded training and consensus needs to be reached among academia, stakeholders and relevant authorities on when and how to switch from discrete sampling to DUMIS. Our review of the delayed acceptance of TOS-based sampling methods by the environmental industry in comparison to the commodities industry has identified multiple factors that require tailored approaches to address, including:

1) Lack of training in TOS and sampling approaches employed in the commodities industry.
2) Unquestioned reliance of outdated USEPA guidance documents by inadequately trained and experienced environmental regulatory staff.
3) Reluctance of government agency attorneys and upper management to call into question regulatory mandates for past or ongoing investigations based on discrete grab sample data.

4) Lack of the degree of third-party verification of sample data prevalent in the commodities industry (e.g., extraction and determination of true mean concentration and mass of a targeted metal from crushed ore).
5) Lack of financial and legal repercussions for what subsequently turns out to be erroneous data and a failure to adequate address potential health risks on the part of government agency directing work or repercussions for significant time and cost overruns necessary to complete a project.

Circumstances are perhaps changing more rapidly in China than in western countries due the fact that most land is owned by the government, most environmental investigations and remedial actions as well as property redevelopment are funded by the government. Many regulatory agencies in China are also beginning to require independent, third-part confirmation of cleanup. This motivates agencies funded by the government to oversee investigation and remedial actions to ensure that the sample data results are defensible and reproducible. A complete transition to TOS-based sampling methods will take time. The fact that a government entity ultimately retains responsibility as well as repercussions for all stages of a project will, however, help to ensure that the environmental industry in the country continues to evolve in a progressive and scientifically-oriented manner.

REFERENCES

Brewer, R., Peard, P. and Heskett M. (2017a). A Critical Review of Discrete Soil Sample Data Reliability: Part 1—Field Study Results. *Soil and Sediment Contamination: An International Journal*, DOI: 10.1080/15320383.2017.1244171.

Brewer, R., Peard, P. and Heskett M. (2017b). A Critical Review of Discrete Soil Sample Data Reliability: Part 2—Implications. *Soil and Sediment Contamination: An International Journal*, DOI: 10.1080/15320383.2017.1244172.

Brewer, R., Song, J. and Ramsey, C. (2019). Advanced Sampling Methods to Expedite the Reliable Characterization, Remediation and Redevelopment of Contaminated Industrial Lands in China: *Proceedings of the 9th World Conference on Sampling and Blending*, Beijing. pp. 292-307.

Brewer, R., Ramsey, C., Heskett, M. and Song, J. (2022). Fake Data? The Need for Theory of Sampling Concepts in Environmental Research and Investigations: *Proceedings of the 10th World Conference on Sampling and Blending*, Kristiansand. pp. 193-206.

Clausen, J. L., Georgian, T., Bednar, A. (2013). Incremental Sampling Methodology (ISM) for Metallic Residues. US Army Corps of Engineers Environmental Science and Technology Certification Program (Project ER-0918).
https://usace.contentdm.oclc.org/digital/collection/p266001coll1/id/4546/

Department of Health, State of Hawaii, USA (HIDOH). (2023). Technical Guidance Manual for the Implementation of the Hawai'i State Contingency Plan, Section 3.0, Site Investigation Design and Implementation: Overview.
https://health.hawaii.gov/heer/files/2023/09/TGMSection3HDOHJuly2023.pdf

Dubé, J.S., and Esbensen, K.H. (2022). Revisiting Pierre Gy's formula (TOS) e A return to size-density classes for applications to contaminated soils, coated particular aggregates and mixed material systems. *Analytica Chimica Acta*. 1193, 339227.

Esbensen, K. and Vogel, D. (2023). Unsubstantiated complacency re. the "assay exchange" paradigm: sampling uncertainties with hidden economic consequences. *Spectroscopy Europe*. DOI: https://doi.org/10.1255/sew.2023.a2.

Gerlach, R.W., Dobb, D.E., Raab, G.A., Nocerino, J.M. (2002). Gy sampling theory in environmental studies. 1. Assessing soil splitting protocols. *Journal of Chemometrics*. 16:321-328.

Gerlach, R.W., Nocerino, J.M., Ramsey, C.A., Venner, B.C. (2003). Gy sampling theory in environmental studies. 2. Subsampling errors estimates. *Analytica Chimica Acta*. 490:159-168.

Maine Department of Environmental Protection (MDEP). (2007). Standard Operating Procedure: Field Quality Control, SOP Number: TS 003, Revision: 4.
https://www.maine.gov/dep/spills/petroleum/documents/sop/fieldqualitycontrol.pdf

Mao, J., Huang Y.J., Song J., Zhao, X.F., Tang, W. (2021). Application of Decision Unit-Multi Increment Sampling in synchronized soil and crop sampling in heavy metal-contaminated farmland. *Journal of Agro-Environment Science (in Chinese)*, 40(3), 535–543.

Market Regulation Bureau of Jiangsu Province (MRBJS). (2022). Technical specifications for investigation on soil contamination of lead-acid battery site. DB32/T 4426-2022.

Ministry of Agriculture (MOA), P.R. China. (2012). Technical rules for monitoring of environmental quality of farmland soil. NY/T 395-2012.

Ministry of Ecology and Environment (MEE), P.R. China. (2018). Technical guideline for verification of risk control and soil remediation of contaminated site. HJ 25.5-2018.

Ministry of Ecology and Environment (MEE), P.R. China. (2019a). Technical guidelines for investigation on soil contamination of land for construction. HJ 25.1-2019.

Ministry of Ecology and Environment (MEE), P.R. China. (2019b). Technical guidelines for monitoring during risk control and remediation of soil contamination of land for construction. HJ 25.2-2019.

Ministry of Ecology and Environment (MEE), P.R. China. (2019c). Technical guidelines for risk assessment of soil contamination of land for construction. HJ 25.3-2019.

Ministry of Ecology and Environment (MEE), P.R. China. (2019d). Technical guidelines for soil remediation of land for construction. HJ 25.4-2019.

Ministry of Ecology and Environment (MEE), P.R. China. (2019e). Technical guideline for groundwater remediation and risk control of contaminated sites. HJ 25.6-2019.

Ministry of Ecology and Environment (MEE), P.R. China. (2022). Technical Provisions for Quality Control of Soil Pollution Survey of Construction Land (Interim).

Minkkinen, P.O. and Esbensen, K.H. (2019). Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors: Fundamental Sampling Error and Grouping and Segregation Error, *Analytica Chimica Acta*. https://doi.org/10.1016/j.aca.2018.10.056.

Minnitt, R.C.A., Rice, P.M., Spangenberg, C. (2007). Part 1: Understanding the components of the fundamental sampling error: A key to good sampling practice. *J. South. Afr. Inst. Min. Metall*. 107, 505–511.

Minnitt, R.C.A. and Esbensen, K.H. (2017). Pierre Gy's Development of The Theory of Sampling: A Retrospective Summary With A Didactic Tutorial On Quantitative Sampling Of One-Dimensional Lots.

Nevada Division of Environmental Protection (NDEP). (2008). Statistical Analysis Recommendations for Field Duplicates and Field Splits. https://ndep.nv.gov/uploads/documents/field-duplicates-and-splits.pdf

Pitard, F.F. (1993). Pierre Gy's Sampling Theory and Sampling Practice. New York: CRC Press.

Pitard, F.F. (2009). Theoretical, practical and economic difficulties in sampling for trace constituents. *Proceedings of the 4th World Conference on Sampling and Blending*, Cape Town.

Pitard, F.F. (2019). Theory of Sampling and Sampling Practice. Chapman and Hall/CRC, Boca Raton, Florida, ISBN 9781138476486.

State Environmental Protection Bureau (SEPB), P.R. China. (2004). The technical specification for soil environmental monitoring. HJ/T 166-2004.

Standardization Administration of China (SAC). (2023). *Soil Quality – Decision Unit and Multi Increment Sampling Methods*: GB/T-42489-2023.

U.S. Environmental Protection Agency (USEPA). (1985). Office of Toxic Substances. Verification of PCB spill cleanup by sampling and analysis. Washington, DC (EPA-560/5-85-026).

U.S. Environmental Protection Agency (USEPA). (1986). Office of Toxic Substances. Field manual for grid sampling of PCB spill sites to verify cleanups. Washington, DC (EPA-560/5-86-017).

U.S. Environmental Protection Agency (USEPA). (1989).Office of Policy, Planning, and Evaluation. Methods for evaluating the attainment of cleanup standards, volume 1: Soils and solid media. Washington, DC (EPA/230/U2-89/042).

U.S. Environmental Protection Agency (USEPA). (1991). Office of Research and Development. Guidance for data usability in risk assessment (part A). Washington, DC (EPA/540/R-92/003).

U.S. Environmental Protection Agency (USEPA). (1992a). Office of Solid Waste and Emergency Response. A supplemental guidance to RAGS: Calculating the concentration term. Washington, DC (EPA 9285.7-081).

U.S. Environmental Protection Agency (USEPA). (1992b). Office of Research and Development. Preparation of soil sampling protocols: Sampling techniques and strategies. Washington, DC (EPA/600/R-92/128).

U.S. Environmental Protection Agency (USEPA). (1999). *Correct Sampling Using the Theories of Pierre Gy*. (1999)., Environmental Sciences Division Fact Sheet. https://clu-in.org/conf/tio/ism1/EPA-Gy-tech-memo-1999.pdf

U.S. Environmental Protection Agency (USEPA). (2002). Office of Emergency and Remedial Response, Calculating upper confidence limits for exposure point concentrations at hazardous waste sites. Washington, DC (OSWER 9285.6-10).

U.S. Environmental Protection Agency (USEPA). (2006). Method 8330b Nitroaromatics, Nitramines, And Nitrate Esters by High Performance Liquid Chromatography (HPLC). https://www.epa.gov/sites/default/files/2015-07/documents/epa-8330b.pdf

U.S. Environmental Protection (USEPA). (2011). User Guide Uniform Federal Policy - Quality Assurance Project Plan Template for Soils Assessment of Dioxin Sites. https://semspub.epa.gov/work/HQ/174547.pdf

U.S. Environmental Protection Agency (USEPA). (2013). Office of Research and Development. ProUCL Version 5.0.00, user guide. Washington, DC (EPA/600/R-07/041).

U.S. Environmental Protection Agency (USEPA). (2018).Hazardous Waste Test Methods, Chapter 3, Inorganic Analytes: U.S. Environmental Protection Agency, Revision 6, December 2018.

U.S. Environmental Protection Agency (USEPA). (2019). Incremental Sampling Methodology (ISM) at Polychlorinated Biphenyl (PCB) Cleanup Sites. https://www.epa.gov/sites/default/files/201908/documents/incremental sampling methodology at pcb cleanup sites.pdf

Walsh, M.R. (2009). User's Manual for the CRREL Multi-Increment Sampling Tool. Strategic Environmental Research and Development Program （Project ER-1481）. https://erdc-library.erdc.dren.mil/jspui/handle/11681/11813

Zhao, X.F., Song, J., Li, J., Tang W., Long J., Mao J., Lü, P.J., Pan, Y.Y. (2021). Application of Decision Unit-Multi Increment Sampling (DUMIS) in Effect Assessment of Pit Excavation at a Ni contaminated Site. *Research of Environmental Sciences (in Chinese)*, 34(6), 1449–1455.

Jing Song

Professor
Chinese Academy of Sciences

Professor in soil chemistry, whose main research interest includes risk-based soil environmental standards, soil sampling methods and mechanochemical destruction of POPs contaminated soil. Member of Soil Environmental Standard Expert Committee, Ministry of Ecology and Environment, Secretary of Soil Remediation Standardization Technical Committee of Jiangsu Province. Member of the Soil Environment Committee, Soil Science Society of China. Senior Associate Editor of International Journal of Phytoremediation. First author of the Chinese DUMIS national Guidance (GB/T 42489-2023 Soil Quality Decision Unit Multi-Increment Sampling Method).