Sampling and sampling strategies for environmental analysis

J. Zhang⁶ and C. Zhang⁷

⁶College of Environmental & Resource Science, Zhejiang University, Hangzhou 310028, China; ⁷Department of Environmental Sciences, University of Houston-Clear Lake, Houston, Texas 77058, USA

(Received 11 December 2010; final version received 8 April 2011)

Sampling errors are generally believed to dominate the errors of analytical measurement during the entire environmental data acquisition process. Unfortunately, environmental sampling errors are hardly quantified and documented even though analytical errors are frequently yet improperly reported to the third decimal point in environmental analysis. There is a significant discrepancy in directly applying traditional sampling theories (such as those developed for the binary particle systems) to trace levels of contaminants in complex environmental matrices with various spatial and temporal heterogeneities. The purpose of this critical review is to address several key issues in the development of an optimal sampling strategy with a primary goal of sample representativeness while minimizing the total number of samples and sampling frequencies, hence the cost for sampling and analysis. Several biased and statistically based sampling approaches commonly employed in environmental sampling (e.g., judgmental sampling and haphazard sampling vs. statistically based approaches such as simple random, systematic random, and stratified random sampling) are examined with respect to their pros and cons for the acquisition of scientifically reliable and legally defensible data. The effects of sample size, sample frequency and the use of compositing are addressed to illustrate the strategies for a cost reduction as well as an improved representativeness of sampling from spatially and temporally varied environmental systems. The discussions are accompanied with some recent advances and examples in the formulation of sampling strategies for the chemical or biological analysis of air, surface water, drinking water, groundwater, soil, and hazardous waste sites.

Keywords: sampling strategy; environmental sampling; sampling frequency; sampling design

1. Introduction

Environmental sampling is an integral part of the overall environmental data acquisition processes that include, but are not limited to, sample collection, preservation, preparation, instrumental analysis, field and laboratory quality assurance/quality control (QA/QC), and data assessment. As the first step in the chain, its importance in the overall data quality cannot be overstated [1–3]. It is self explanatory that if a sample is collected improperly, then all our subsequent careful lab work is useless. A poorly collected and
An unrepresentative sample will by no means generate any meaningful and reliable data. Rather, it may provide misleading information owing to its defect quality.

Despite its utmost importance, sampling is a subject analogous to the weather in a sense that much discussion has been made over the years but very little has been done about developing more scientific methods of sampling [4,5]. In an editorial note published over a half century ago, Murphy [4] pointed out the lack of scientific work on sampling in general, including sampling in metallurgy, pharmaceutical, food, petroleum industries as well as sampling in industrial hygiene and environmental science. In the field of environmental sampling, for example, only a few organized efforts were noted thus far, including the published work as a result of symposiums by the American Chemical Society (ACS) Committee on Environmental Improvement [6–9]. Research work on environmental sampling is very limited, as is evident from the number of papers in two primary ACS journals (Environmental Science and Technology and Analytical Chemistry) (Figure 1). A search in these two journals revealed that sampling papers account for far less than 1% of the total papers published, although there seems to be an overall increased number of papers published in recent years. In a recent study, Ort et al. [10] reported that only 11% of the 87 surveyed papers on pharmaceutical and personal care products (PPCPs) in wastewater systems provided justifications for sampling compared to 99% on analysis.

Some general misconceptions probably attribute to the lack of attention paid to environmental sampling as well as sampling in other industries. Oftentimes one may incorrectly assume that the quality of data is primarily determined by the nature of analytical measurement methods and accordingly the expensive state-of-the-art instrumentation rather than by sampling and sample preparation techniques [11]. Therefore, it is not uncommon to see that analytical results are reported to the third decimal point [4], whereas sampling uncertainties and associated errors are hardly mentioned. To many less

![Figure 1. Total numbers of ‘sampling’ papers per year in two primary American Chemical Society (ACS) journals: Environmental Science and Technology and Analytical Chemistry.](image-url)
skilled field crews, sampling can just be perceived as simple as the haphazard collection of grab samples, rather than a complicated science that depends on solid statistics and prior knowledge about the variability of the population being sampled.

It is apparent that the science of sampling is an enormous subject that deserves a multidisciplinary effort [5]. This paper will primarily address some important sampling strategies that are often overlooked in the analytical community who places more focus on laboratory sample preparation and instrumental analysis. It is beyond the scope of this review to cover many other important sampling issues such as new sampling devices (e.g. passive sampling), sample preservation, preparation, and sampling QA/QC [12–19]. We will first examine the errors of sampling relative to the errors of laboratory analysis, which is followed by the introduction of several sampling approaches commonly employed in environmental sampling as well as recent advances to formulate an optimal sampling strategy. Several key issues regarding, for example, the use of stratified random sampling, compositing, the effects of sample size and sampling frequency will be addressed with examples of sampling and sampling strategies applicable for chemical or biological analysis of air, water, soil, and hazardous waste sites.

2. Sampling errors versus analytical errors

The total error ($S_t$) of environmental analysis is the sum of two independent sources of errors, i.e. sampling error ($S_s$) and analytical error ($S_a$) [20]:

\[ S_t^2 = S_s^2 + S_a^2 \] (1)

Both sampling errors and analytical errors in Equation (1) are expressed in terms of the absolute values of variance or the relative % of the total variance. Note that errors expressed otherwise such as relative standard deviation (RSD) are not directly additive as shown in Equation (1). Although it is generally believed that sampling errors dominate the errors of analytical measurement (90% or higher, [21]), direct experimental measurements of error contributions from sampling compared to analysis have only been made available in a very few well-designed studies [22–24]. Lamé and Defize [22] determined sampling errors and analysis errors to be in the ranges of 91.0–98.6% and 0.3–5.5%, respectively, depending on the soil sample size in the range of 0.1 to 6,500 g for the determination of cyanide contents in soil samples. In measuring oil concentrations in Boston Inner Harbor surface water samples, Ahmed et al. [23] obtained similar results – sampling errors owing to variability of the number of oil-sorbed particles and the weight per particle to be 96% whereas analytical errors to be only 4%. Sampling errors for handling particulate materials were further divided into two sources, i.e. the fundamental error owing to variations of chemical compositions among different soil particles (clays, sands, etc.), and the segregation error owing to variations of locations (lots). The fundamental error was kept minimal by increasing sample size to >10 g, whereas segregation error was reduced by mixing soil samples from multiple locations (i.e. compositing). In an extreme case with a heterogeneous distribution of soil contaminants such as explosives compounds present in a nugget form in soils, increasing subsample size from 2 to 50 g did not reduce sampling errors [25].

Mar et al. [26] pointed out that in watershed monitoring, typical measurement errors (RSD) are within 25% for physical/chemical and 50% for biological entities, whereas sampling errors owing to natural variation are in the range of 100–400% of observed mean
values for physical, chemical, and biological characteristics. Figure 2 depicts how the total errors are correlated with sampling errors and analysis errors based on 1000 iterations of the Monte Carlo simulations. It is evident from this simulation that given the predominant nature of sampling errors, the total error is significantly correlated to the sampling errors \( R^2 = 0.97 \) whereas the effect of analytical errors on the total errors is not significant \( R^2 = 0.04 \). This result has important implications in that a further reduction in analytical uncertainty (even if it is possible) is probably of little importance to cost-effectively improve the overall data quality [27–29]. In fact, it would be more cost-effective to employ a less precise in-situ measurement technique (such as portable X-ray fluorescence for metal analysis), rather than the more traditional but time-consuming ex-situ instrument (such as acid digestion followed by the analysis using inductively coupled plasma spectroscopy) [28,30].

Unlike analytical errors which can be easily quantified by spiking and standard addition, sampling errors are difficult to quantify, because it is impossible to spike the analyte in the entire sampling population (e.g. a lake). This is a primary reason why analytical errors are frequently (yet improperly) reported to the third decimal point, and why sampling errors are rarely documented in environmental analyses. Such a discrepancy also attributes to the lack of a theoretical basis in quantifying sampling errors owing to various types of spatial and temporal heterogeneities. Gy’s sampling theories [31,32] addressed sampling errors related to granular materials for geological and metallurgical applications. According to Gy, the types of sampling errors (sources) can be grouped into seven categories, including fundamental error (inherent sample characteristics such as particle sizes and compositions), segregation error (sampling lots, particle density induced stratifications), long-range heterogeneity error (spatial variations), periodic heterogeneity error (temporal and spatial fluctuations), delimitation error (inappropriate sampling design and the wrong choice of equipment), extraction error (failure of sampling procedure

Figure 2. Monte Carlo simulation of total error vs. analytical error and sampling error (the results as shown are based on a total of 1000 iterations assuming analytical errors and sampling error are in the range of 0–50% and 100–400%, respectively).
to precisely extract the intended increment), and separation error (loss, contamination, and alteration of a sample or subsample). However, such sampling theories developed for particulate systems cannot be directly applied to complex environmental matrices for trace chemicals. In assessing temporal uncertainties of monitoring trace levels of PPCPs in wastewater, Ort et al. [33] were able to determine errors associated with different sampling modes and frequencies (continuous vs. discrete) using model simulated data against results experimentally measured at a 5-min sampling interval. Their results indicated that sampling errors owing to relatively long sampling intervals and inadequate sampling modes can lead to over-interpretation of measured data and ultimately wrong conclusions. This Monte Carlo simulation approach was also used to compare different sampling strategies for estimating tributary loads [34]. Model simulations suggested that flow stratified sampling provided the best estimate of tributary load among all sampling strategies examined. It should be noted that the magnitudes of sampling errors for trace contaminants cannot be readily quantified unless, for example, in-situ real-time monitoring data can be acquired and then data are retrospectively assessed with the aid of modelling approaches [10].

The general strategies to reduce sampling errors are either through replicate samples for analysis or through an increased sample size (volume or weight) [23,28]. The best sample size is certainly to use all the samples after homogenization. Practically, this is impossible for environmental sampling. In determining petroleum in Boston Inner Harbor water samples, Ahmed et al. [23] showed that increasing sample size (e.g. 1 L to 1 gallon) is not worthwhile, whereas keeping a sample size of 1 L but having four replicates shows an improved precision. Their study indicated that increasing the number of replicates \( n \) improved overall precision by a factor of \( n^{1/2} \) over that of a single analysis. In many circumstances, however, increasing both replicates and sample volume is apparently not preferred. For example, an increased number of replicate analyses may be cost-prohibitive for certain difficult-to-analyse trace contaminants. An aliquot extracted from an excessive amount (volume) of a sample will make subsequent sample clean-up harder owing to the matrix effect. In the following sections, we confine our discussions to the development of an optimal and practical sampling strategy that considers primarily sample representativeness while minimizing the total number of samples hence the total cost of sampling and analysis.

3. Biased and unbiased sampling approaches

Among a variety of sampling strategies, judgmental sampling and haphazard sampling are both non-probability based. Judgmental sampling is the subjective selection of sampling locations based on information on the sampling site, visual inspection, and personal knowledge and experience. While judgmental sampling can reduce sampling number, it is biased (if done improperly) and does not support any statistical inference. In studying the eutrophic status of a large number of lakes (over 11,000 lakes with size \( \geq 1 \) ha) in the northeastern US, Peterson et al. [35] revealed that judgmental sampling biased toward the subjective selection of larger lakes which resulted in an underestimation of eutrophic status (i.e., greater depths of Secchi disk transparency of the lake water). Haphazard sampling claims ‘any sampling location will do’ and hence it encourages taking samples consciously at convenient locations or times for spatial or temporal sampling, respectively [1,28]. Mistakenly employed by many as a random approach, haphazard sampling is also
biased and legally not defensible. The integration of prior knowledge into a random sampling procedure is possible through a statistically-based double sampling approach such as ranked set and weighted double sampling [36]. Both approaches employ a frugal measurement (e.g. visual observation of biological habitants) which is correlated to a more costly but accurate measurement (e.g. measuring tape) in a subsequent sampling stage. In ranked set sampling, a frugal measurement is used to rank a subset of samples followed by the actual measurement of one sample from each set until all sample sets are completed. In weighted double sampling, prior information from a frugal measurement is used to categorize samples into groups (strata), and the overall mean is calculated from the weighted average of all strata.

Figure 3 shows three common probability-based sampling designs for two-dimensional spatial sampling as well as one-dimensional temporal sampling [1,37]. With the assumption of insignificant variability within the sampling medium, simple random sampling is applicable mostly for relatively homogeneous populations. It will allow for statistical verification of clean-up, but typically result in more samples which is not as cost-effective as other sampling designs in large-scale sampling campaigns [38]. This drawback can be overcome by a systematic sampling design either through systematic grid or systematic random approaches. In mapping spatial patterns of environmental pollution, systematic sampling approach can be used to delineate spatial patterns such as airborne and soil contaminants within a geographical region [e.g. 39, 40]. In monitoring temporal variations, systematic random sampling gave a better representativeness for atmospheric PCBs with diurnal variations corresponding to daily temperature changes [41]. The grid sizes (e.g. 2 km x 2 km in space, 24 hr in time) vary depending on how well the resolution of a particular spatial or temporal pattern need to be defined. The spatial grid shape is typically square, but in hazardous waste site spatial sampling, Parkhurst [42] illustrated that triangular grids can give more spatial coverage and result in 23% fewer groundwater sampling wells than required with square grids.

Stratified random sampling is an extensively used sampling approach with a promise of various environmental applications. This sampling approach can significantly reduce standard deviation (hence improved sampling precision) as compared with other approaches particularly if the strata are quite different from one another. Scott [43] developed a computerized stratified random site-selection approach to delineate spatial pattern of groundwater quality in relation to land uses and hydrogeological settlings.

Figure 3. Three common probability-based sampling designs for sampling in (a) two-dimensional space domain, and (b) one-dimensional time domain: 1. simple random sampling; 2. stratified random sampling (three strata); 3. systematic grid sampling; 4. systematic random sampling.
The reduction of sample number and cost is achieved by collecting more samples in relatively more heterogeneous strata, or fewer samples in strata where sampling costs are relatively higher than other strata (e.g. saturated vs. vadoze zone). As shown in Figure 4, the types of strata in environmental analysis vary depending on the differences in geographical, hydrological, meteorological, demographical, chemical, or biological features. For example, six sampling strata in the combinations of three contaminant categories (heavily oiled, moderately oiled, and not oiled) and two beach lengths (<100 m, >100 m) were chosen to estimate the total contaminated area and mass remaining in Prince William Sound, Alaska after the Exxon Valdez oil spill [44]. In a subsequent study on the vertical distribution of shoreline oil residues, six other strata in the combination of three regions (Herring Bay, Lower Pass, and Bay of Isles) and two contaminant categories (heavily and moderately oiled) were designated [45]. Another well-designed sampling study worthy of note was conducted by El-Shamy for the estimation of the population of fish impinged on the intake screens of power plants [46]. Stratified random sampling using optimum allocation, i.e. sampling most intensively during months of peak fish abundance and variation, resulted in a significantly improved precision as compared to a routine method with evenly distributed sampling events of every four days throughout the year (RSD 4% vs. 32% for a representative fish species). It should be noted that, regardless of various uses of stratified random sampling reported to date, sample representativeness and sampling precision are seldom justified in the majority of literature.

4. Reducing sample number and sampling frequency

The preceding section presents ‘where’ and ‘when’ to take samples, an equally important component of sampling strategy is to estimate ‘how many’ samples and ‘how frequent’ of the sampling events. In a broad sense, a sampling frequency also implies a corresponding sample number with respect to temporal sampling – both target at a minimal sampling effort for the best statistical representation of the population. Unfortunately, there is no

Figure 4. Stratified random sampling: example types of strata in environmental applications.
universal guideline to achieve such a goal without considering site-specific information. Determining a minimally required sample number and sample frequency is often a challenging task for any environmental project. The key to the determination of sample number or sample frequency is to gather any available data on the spatial and temporal variabilities of the population for which samples are to be collected. The variabilities in turn depend on the properties of the contaminants and more importantly the types of environmental matrices, such as groundwater, surface water (flowing vs. stagnant), atmosphere, soils, or domestic sewage effluents. An example of such a spatial variability is when certain contaminants (e.g. explosives compound) exist in the ‘nugget’ form whereas a soil sample collected a few centimeters away may be totally depleted of this compound [25].

The minimal sample number to be representative of a population at a specified tolerable error limit can be determined as follows:

\[ n = \frac{S^2 t^2}{e^2} \]  

where \( s \), \( t \), and \( e \) are standard deviation, Student’s \( t \) value, and allowable error, respectively. Equation (2) clearly defines the dependence of sample number on variance \( (s^2) \), and the inverse relationship between sample number and the square of the tolerable error level. Stated in another way, the required sample number \( (n) \) is directly proportional to the square of the ratio of standard deviation to the tolerable error \( (s/e) \). As the ratio of \( s/e \) and confidence level increase at the high end, the required sample number increases at a more significant rate (Figure 5). Equation (2) can also be written with a standard normal variate \( Z \) if the number of samples or replicates is larger [47]. Additionally, in regulatory compliance monitoring, the allowable error \( (e) \) can be substituted by the deviation of contaminant level from its regulatory threshold \( (RT - \bar{x}) \), where \( RT \) is a regulatory threshold (e.g., emission standard) and \( \bar{x} \) is the estimate of mean (actual concentration of

![Figure 5. Required number of samples as a function of the ratio of standard error to tolerable error \( (s/e) \) at three confidence levels (80%, 90%, and 95%).](image)
the contaminant of interest). This implies more samples are required when the estimated sample mean approaches the regulatory limit. Conversely, fewer samples are needed at the same confidence level if contaminant concentrations are far below or above the regulatory limit.

Since the $t$-value depends on sample number and confidence level, a trial-and-error calculation is needed to estimate $n$ [1,6]. Equation (2) also indicates that an optimal sampling design is at the expense of a prior knowledge regarding the natural variability. This becomes challenging because preliminary data on spatial or temporal variabilities are typically not available. For example, a dataset for a period of two years was considered to be essential to determine an optimal groundwater sampling frequency [48]. Nevertheless, this need may be justified for long-term monitoring programs with a large number of monitoring stations and costly analysis, such as the case in groundwater monitoring in Superfund remediation and post closure of landfill sites. Efforts have been made to reduce the cost by reducing the sampling frequency based on site-specific scenarios, such as the cost-effective sampling (CES) algorithm considering trend, variability, magnitude statistics, periodicity, and autocorrelation [49,50]. The underlying principle for CES algorithm is that sampling frequency should be determined primarily by the rate of change in concentrations observed in the recent past. Monthly and quarterly sampling frequency represents a good initial choice for the network design of surface water and groundwater monitoring, respectively [48,51]; however, the adequacy should always be re-evaluated on a site-to-site basis. For example, groundwater sampling frequency of the existing monitoring scheme can be downgraded (e.g. semiannual to annual, or quarterly to semiannual) when contaminant concentration changes in groundwater wells become less insignificant over time [52].

It should be noted, however, that the efforts to achieve optimal sampling should be exercised with cautions as problems may arise when sampling becomes too infrequent or even too frequent. For example, the estimates of annual mean herbicide concentrations in Midwestern rivers owing to ‘spring flush’ during spring and early summer runoff events were underestimated by quarterly sampling scheme as required by the US EPA for municipalities using surface water as a source of drinking water [53]. A similar problem was noted recently by Ort et al. [33] in sampling PPCPs from sewage treatment plants. Their work indicated that an extremely short 5-min interval was needed to enable the positive detection of the peak concentrations of trace PPCPs as a result of a single toilet flush in a small community. An important implication of this finding is that results relied on a single ‘lucky’ grab sample may provide misleading information particularly for trace levels of emerging environmental contaminants such as PPCPs and hormonal substances. Similarly in atmospheric monitoring of particulate matter, current regulatory requirement of a low sampling frequency at one day every six days may have caused misclassifying a nonattainment area as being in compliance [54]. In contrast to infrequent sampling, a strategic oversampling (termed ‘sampling out’) was also noted in some drinking water treatment systems where oversampling of drinking water was intentionally employed to avoid violation of the Total Coliform rule proscribed by the US EPA [55]. This sampling out issue should be avoided from both technical and policy standpoints. The strategic oversampling is different from a conservative sampling campaign commonly seen in the monitoring of industrial effluents, where processes are being monitoring more closely to justify unnecessary potential emission charges from a regulatory agency.

There is an alternative strategy to mitigate the high cost associated with the analysis of a large number of environmental samples for the measurement of multiple parameters.
Compositing is such an appealing sampling strategy to reduce sample numbers when sampling cost is low relative to analytical cost and the primary interest is to estimate the mean at the expense of variations [1,56,57]. Compositing can be achieved by physical grinding of solid samples or mixing with volume-, time-, or flow-proportion mechanisms for liquid samples. These mixing devices have been made available commercially for wastewater sample collection, and storm water collections when manual collection is difficult because of a short storm runoff duration [58]. There are two basic environmental applications using composite sampling strategy. The binary classification is to detect the presence or absence of certain chemicals/biological indicators in the environment or biological tissues [59], such is the case in determining whether a Superfund site has been cleaned up or not, or the detection of certain pathogens in a drinking water supply. The mixing approach is a more widely used compositing approach for the determination of an average concentration of chemicals excluding volatile organic compounds (VOCs) since samples for VOCs cannot be mixed.

Figure 6 is a simplified schematic showing the cost-savings of using composite sampling if the analytical cost is assumed to be $200 per sample. In the case of binary classification, a reduction of sample number from 10 to 7 samples results in a 30% cost-saving. Here we also assume a 10% prevalence rate (detection rate) for a biological or chemical parameter. In Figure 6(b), where composting of every 5 samples is employed to estimate the mean of 10 original samples, an even greater 80% cost-saving can be achieved. The success of composting approach depends on several factors. The modelling work done by Johnson and Patil [59] indicated that the binary approach is cost-effective only for contaminants with a low prevalence (detection) rate of <10%. The cost-effectiveness of composting sampling is also strongly dependent on the cost of sampling ($C_s$) relative to the cost of analysis ($C_a$). The cost-effectiveness improves from $C_a/C_s = 1$ to 10, and the change in cost-effectiveness becomes negligible if the cost of analysis further increases beyond this range [59]. In a related study, Rohlf et al. [56] provided mathematical algorithms to estimate sample size and the means to optimizing composite sampling protocols. These algorithms were used to find the optimum sampling protocol that stays within a fixed

![Figure 6](http://example.com/figure6.png)

Figure 6. Composite sampling to reduce sample numbers: (a) to confirm clean up or not (binary classification), (b) to estimate the mean. Symbol ⊕ denotes absence of contaminants (below detection limit); ⊗ denotes presence of contaminants (above detection limit). Five samples are composited for both (a) and (b). In 6(a), if the composite is tested positive, each of the five samples is then analyzed. The total number of samples analyzed is seven. In 6(b), only two composited samples are analyzed from initial 10 samples.
budget, or to find the least costly sampling protocol that is still able to reliably detect a specified difference in means. Such algorithms are useful for determining composite sampling protocols when dealing with animal tissue samples, and other media such as water and sediment.

5. Concluding remarks
Some relevant guidelines for environmental sampling have existed for over a century [60], yet there is still an apparent need to raise the awareness of the importance of sampling during the entire environmental data acquisition process. It is crucial for us to recognize that all our careful lab work is wasted if a sample is not collected properly. Beginning professionals including college students in the environmental and analytical chemistry fields (and many other professions) should be trained to comprehend ‘there is no analysis without sampling’ [61] philosophy. Several hands-on training materials have been recommended for educational use in environmental and chemistry curriculum [1, 62–67]. Analytical chemists should pay more attention to the precision-limiting step of ‘sampling errors’ not just the state-of-the-art instruments. It is also crucial that all relevant research papers should present a clear justification for essential sampling details to document the sample representativeness and/or sampling errors.

In designing an optimal site-specific sampling strategy with regard to sampling location, time, frequency and sample number, one need to thoroughly examine the sources of temporal and spatial variability [57]. Several statistically-based sampling approaches capable of achieving cost-effectiveness of environmental sampling, such as stratified random sampling, systematic sampling, and compositing sampling are currently still underutilized. In compliance monitoring, it is also likely that the optimal sampling strategy may not agree with what is required by the current regulatory agency. These issues should be studied on a case-by-case basis since a universal sampling protocol is nonexistent. Sampling is complicated field of science, and much research is needed to revitalize it through some fundamental interdisciplinary and coordinated efforts among practitioner, researchers, regulators, and governmental agencies. Future research is needed to advance this field toward the development of a more scientifically sound theoretical framework as well as practical protocols for various environmental sampling scenarios.

References
[49] M. Ridley and D. MacQueen, presented at the American Society of Civil Engineers, Anchorage, AK, United States (2005).