59390

PIERRE GY'S SAMPLING THEORY *and* SAMPLING PRACTICE

Volume II
Sampling Correctness and Sampling Practice

Francis F. Pitard

CRC PRESS

10495

Pierre Gy's Sampling Theory and Sampling Practice

Volume II
Sampling Correctness and Sampling Practice

Author
Francis F. Pitard
President
Pierre Gy and Francis Pitard
Sampling Consultants
Broomfield, Colorado

CRC Press, Inc.
Boca Raton, Florida

Library of Congress Cataloging-in-Publication Data

Pitard, Francis F.
Pierre Gy's sampling theory and sampling practice / author,
Francis F. Pitard.
p. cm.
Includes bibliographies and indexes.
Contents: v. 1. Heterogeneity and sampling — v. 2 Sampling correctness and sampling practice.
ISBN 0-8493-6658-5 (v. 1) — ISBN 0-8493-6659-3 (v. 2)
1. Ores — Sampling and estimation. 2. Gy, Pierre. I. Title.
TN560.P56 1989
622'.7—dc19 89-911
CIP

This book represents information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Every reasonable effort has been made to give reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

All rights reserved. This book, or any parts thereof, may not be reproduced in any form without written consent from the publisher.

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431.

© 1989 by CRC Press, Inc.
Second Printing, 1991

International Standard Book Number 0-8493-6658-5 (v. 1)
International Standard Book Number 0-8493-6659-3 (v. 2)

Library of Congress Card Number 89-911
Printed in the United States

Chapter 21

SAMPLING OF LIQUID AND SOLID WASTES AND SAMPLING OF THE ENVIRONMENT

21.1. INTRODUCTION

Environmental protection agencies have been created in various countries around the world because, as an industrial society, we have failed to evolve and grow in harmony with the fragile heritage that was given to us: our environment.

Cost effective management of this extremely complex environment requires an integrated effort from many competent people such as geologists, metallurgists, hydrologists, biologists, chemists, statisticians, lawyers, and so on. All these people have one thing in common: the decisions they make heavily depend on estimates made by proxy on samples supposedly representative of a selected fraction of the environment.

Measuring the progression and the impacts of an environmental problem, and finding solutions, are goals that are successfully reached only if all sources of sampling errors are well understood. Therefore, in our following discussions we assume that the reader is already familiar with the content of the previous 20 chapters of these volumes.

There are two equally important sources of sampling errors:

— All materials studied in the environment are heterogeneous (the subject was studied in Parts 2 and 3).
— All sampling processes used to implement a sampling plan are either correct or incorrect (the subject was studied in Part 4).

The study of the various forms of heterogeneity is essential to the environmentalist who wants to detect and understand random fluctuations, trends, and cycles affecting the behavior of a given pollutant. Yet very little attention has been given to sampling errors such as:

— The fundamental error FE
— The grouping and segregation error GE
— The long-range heterogeneity fluctuation error CE_2
— The periodic heterogeneity fluctuation error CE_3

Only a good understanding of the heterogeneity, carried by a potential pollutant in a given environment, can lead to the development of an appropriate and effective sampling plan. To be cost effective, this sampling plan often consists of an iterative action.

Now, the implementation of a sampling plan can be correct or incorrect. Sampling errors due to incorrectness are numerous and very dangerous because they are difficult to detect and to quantify. These sampling errors are

— The increment delimitation error DE
— The increment extraction error EE
— The preparation error PE.

These errors always affect the accuracy of a sample. Yet, with the exception of PE, these errors are critical but they have not been given the attention they deserve. The development of the liquid waste sampler called "Coliwasa" is a good example of a tiny step made recently in the right direction to minimize errors such as DE and EE.

It should be clearly understood that it is entirely up to us to perform correct sampling. ***Correct sampling is often completely uncorrelated with sampling cost. Excuses to perform incorrect sampling cannot be justified by time and money limitations. They are often justified by the desire to comply with standards implemented by committees who know very little about sampling. It is a fact that, for environmental sampling, the technology of correct samplers remains to be developed.*** We hope that the contents of this chapter will be helpful to those in charge of developing sampling systems for the environment and will bring forth progressive new ideas.

21.2. KEY QUESTIONS AND ISSUES RELATED TO SAMPLING

Sampling of mineral ores, products, concentrates, coal, or chemicals presents many difficulties that were reviewed in the previous 20 chapters of these volumes. Sampling the environment presents exactly the same difficulties; however, there are a few peculiarities which could lead to important issues. These peculiarities need to be addressed.

When a given pollutant creates an environmental concern, the first questions the environmentalist asks himself are "At what level does this pollutant cause a health risk? Consequently, what should its appropriate regulatory threshold be?" Obviously, answers to these two questions have a tremendous impact on the development of an appropriate sampling plan. From previous chapters, we know how to prepare such a plan. So far there is no apparent difficulty but, as a direct consequence of the notion of heterogeneity, the environmentalist could ask himself another question: "Should the notion of threshold be relative to the sample size?" For example, let us assume that a pollutant is present at a level well below the threshold when large samples of weight M_{s2} are collected and analyzed. It could very well be that much smaller samples of weight M_{s1} (e.g., $M_{s1} = M_{s2}/100$), collected at the same locations than the large samples, will, most of the time, show lower pollutant amounts and, once in a while, will show higher levels than the regulatory threshold. This is observed when samples are small enough to skew the distribution of assay values until the distribution obeys a Poisson law: many values are much too low and a few are much too high.

This immediately leads to more questions: "If this pollutant behaves erratically because of its heterogeneity, is there a small chance that the same pollutant could, at times, become very hazardous because of a small scale clustering behavior, even though estimates from large samples are always below the threshold? Does a large sample cover the nature of the real problem? Is it necessary to completely redefine the notion of threshold or, at least, connect it somehow with the notion of heterogeneity?"

Again, this leads to more questons: "Is the strength of the bond between a pollutant and its support (e.g., sediment, mineral, vegetal, microorganism, colloid, and so on) an important criterion to define the value of a regulatory threshold? Is the heterogeneity carried by a pollutant a function of the strength of this bond? Is the amount of heterogeneity carried out by the pollutant correlated to the amount of heterogeneity carried out by its support?"

In this chapter we will not bring satisfactory answers to these delicate questions. However, the environmentalist may realize now that sampling the environment is not a simple matter. It also appears that a strong standardization of sampling plans may not be a great idea. Many are those that encourage standardization to ensure possible comparisons from one environmental site to another. The idea is legitimate. Indeed, a sampling strategy can be and should be standardized. The implementation of sampling plans should use standardized samplers whose primary quality is correctness. The standardization of an incorrect sampler, as it is often done, should be carefully prevented. However, the notion of heterogeneity being what it is, the standardization of sampling plans and sampling protocols is total nonsense. Each case is unique and each pollutant involved carries a particular amount of heterogeneity.

There are no possible ways to guess what this amount of heterogeneity will be without preliminary iterative testings. Then, and only then, can we standardize a sampling protocol for a local condition only. In other words the standardization of a strategy and of a correct sampler can be very general, while the standardization of a protocol can only be local. This distinction is important.

The standardization of a sampling strategy and of correct samplers can tremendously reduce lag time in updating or developing new methodologies as regulations change; but, and this is very important, enough flexibility should be allowed in those regulations to adjust sampling protocols with special local conditions characterized by the amount of heterogeneity carried by the pollutants of interest. This is the essence of a logical approach.

21.3. A LOGICAL APPROACH

The most difficult part of an environmental sampling problem is its formalization which requires solid knowledge in conventional and nonconventional statistics, knowledge of sampling theory, but also knowledge in the domain in which the problem takes place (e.g., geochemistry, hydrology, biochemistry, sedimentology, and so on). Therefore, the statistician alone cannot decide upon an appropriate sampling plan.

The environmentalist (e.g., geochemist, biologist, sedimentologist, chemist, metallurgist, and so on) should not only participate in the development of appropriate sampling plans, but should also initiate them in a logical way. Thus, the environmentalist should be familiar with the various parts of sampling theory.

An effective coordination between the environmentalist and those in charge of the necessary logistic support for the implementation of a sampling plan is a must. Special local conditions such as topography, temperature, water flows, meteorology, tides, and so on, are likely to play an important role. Indeed, there are numerous factors that may directly or indirectly affect the validity of a sampling plan. The environmentalist should ascertain that the sample collector will proceed exactly as indicated in the sampling plan, and that all the necessary and pertinent information will be logged.

All this is already the object of well-established standard procedures that are valid only if the environmentalist can quickly provide an effective sampling plan. It is impossible to prepare a valid plan without first characterizing the various kinds of heterogeneity carried by a pollutant in a particular waste. This part is often considered so superficially that the conclusions, reached after a thorough analysis of the data, are nothing short of being very ambiguous. Furthermore, the implementation of the sampling plan can be successful only when the environmentalist can quickly make a judgment on the correctness of a sampling system. This critical part is not well known by the public and rarely addressed. It is the largest source of sampling bias.

21.3.1. Structural Property of a Pollutant

The most relevant structural property of a pollutant is characterized by the amount of heterogeneity h it carries. We saw, in Part 1 of this book, that a one-dimensional waste h can be decomposed into three parts:

$$h = h_1 + h_2 + h_3 \tag{21.1}$$

where:

- h_1 is the small scale heterogeneity. It is the random fluctuations shown by the pollutant.
- h_2 is the large scale heterogeneity introduced by trends. Those are nonrandom fluctuations.
- h_3 is the heterogeneity introduced by cyclic phenomena. They are numerous.

We also have for a zero-dimensional waste:

$$h = h_1 \tag{21.2}$$

If we purposely consider a one-dimensional waste as a zero-dimensional waste, then h_3 cancels and h_2 becomes part of h_1. A one-dimensional waste can become a zero-dimensional waste if the chronological order of the samples collected to monitor a pollutant is voluntarily or involuntarily ignored. Conventional statistics often ignore the chronological occurrence of a pollutant. The trouble is that, most of the time, the occurrence of a pollutant is not a random phenomenon.

CONCLUSION 1: For zero-dimensional wastes we should characterize h_1 by calculating the constant factor of constitution heterogeneity IH_L. When we have an idea about the value of IH_L we can predict the variance of the fundamental sampling error FE, and the maximum of the variance of the grouping and segregation sampling error GE. Thus, we can select an appropriate sampling protocol with respect to the increment weight, sample weight, and subsample weight. We can also optimize the number of increments per sample.

CONCLUSION 2: For one-dimensional wastes we should characterize h_1, h_2, and h_3 by conducting variographic experiments, or by calculating variograms with existing data. With the variograms we can identify all sources of the pollutant fluctuations; therefore, we can decide upon the appropriate sampling protocol with respect to the increment weight, sample weight, and subsample weight, the optimum number of increments per sample, the spacing between increments and the most adequate sampling mode (i.e., systematic random, stratified random, or simple random).

21.3.2. Structural Property of Sampling Correctness

Sampling must be correct. It is often stated that sampling accuracy is usually achieved by some form of random sampling. It is true, and it would be indeed very easy to achieve sampling accuracy if randomness was the only factor on which it depends. Unfortunately, in the "real world", the implementation of probabilistic sampling is much more complicated. Accuracy depends mainly on sampling correctness which is a structural property of the sampling process.

CONCLUSION: If the sampling process is not correct, the sampling operation cannot be accurate, regardless of how good the sampling plan is. Furthermore, we cannot directly control accuracy; however, we can directly control correctness. We also know that a correct sampling operation is always accurate.

Many are those who are tempted to test an incorrect sampling system for its accuracy. The trouble is that an incorrect sampling system can be circumstantially accurate today, biased in one direction tomorrow, and biased in another direction the day after tomorrow. In sampling, there is no such thing as a constant bias. The desire of controlling accuracy without controlling sampling correctness is certainly the worst judgment error that a person can make. It is a direct departure from basic logic.

21.3.3. Interactions Between Sampling Cost, Accuracy, Precision, and the Regulatory Threshold

Intuitively, it is logical to think that the closer the expected level a given pollutant comes

to its regulatory threshold, the better the precision should be; therefore, the greater the sampling effort and sampling cost.

Contrary to widespread opinion, the same statement does not apply to accuracy. Accurate sampling mainly depends on the correctness of the sampling system. In many cases, a correct sampling system is less expensive than an incorrect one. In fact, there are many sophisticated on-line samplers and analyzers that are totally incorrect (on-stack emission monitoring for sulfur dioxide is an excellent example). Achievement of sampling accuracy is not a matter of cost, but rather a matter of choice. The trouble is that correct sampling systems for sampling wastes in the environment are almost nonexistent. These systems remain to be developed, tested, and implemented. This important aspect of the correctness of environmental sampling will be discussed later in this chapter.

21.3.4. Standardization of a Sampling Strategy

If we could standardize a sampling strategy for the sampling of liquid and solid wastes, the following four essential steps should be considered:

1. Modeling the waste:
   — Zero-dimensional (easy to solve)
   — One-dimensional (easy to solve)
   — Two-dimensional (difficult to solve)
   — Three-dimensional (often impossible to solve)
2. Characterizing the heterogeneity carried by the pollutant of interest in the waste:
   — Determine IH_L for zero-dimensional wastes.
   — Determine CE_1, CE_2, and CE_3 for one-dimensional wastes.
   — Two- and three-dimensional wastes can also be investigated as one-dimensional wastes, except that several directions or transects must be considered. This goes beyond the purpose of this book and enters the domain of geostatistics. We also should remember that sampling correctness is difficult to achieve with two-dimensional wastes, and often impossible to achieve with three-dimensional wastes.
3. Developing an appropriate sampling plan:
   — Transform a three-dimensional waste into a sum of several superposed two-dimensional wastes each time that it is possible.
   — Determine appropriate sample and subsample weights with respect to IH_L, CE_1, and targeted level of precision.
   — Determine the appropriate number of increments per sample and subsample.
   — Select an appropriate selection mode with respect to IH_L, CE_1, CE_2, and CE_3.
   — Optimize the laboratory subsampling plan.
4. Implementation of the sampling plan:
   — Thorough verification of the correctness of the sampling systems:
   - In the field.
   - At the laboratory.

Any of these above four major steps cannot and should not be overlooked in any sampling strategy concerning the sampling of liquid and solid wastes.

21.4. THE COMPONENTS OF THE OVERALL SAMPLING ERROR

This section is not a repetition of the contents of Chapter 3. Its objective is to make a clear correlation between our vocabulary and the usual vocabulary used by environmentalists.

An opening remark is necessary: when the environmentalist states that a waste is heterogeneous in terms of its chemical properties, for example, it should be clearly understood

that this statement is meaningless as long as he does not specify what kind of heterogene he is referring to. The reason is simple: causes, effects, and cures are different.

21.4.1. Errors Generated in the Sampling of Zero-Dimensional Wastes

FE: The fundamental sampling error is introduced by the variability of a pollutant fron one fragment to another. It is the constitution heterogeneity. Obviously this erro is negligible for liquids and gases without suspended solids. It is small for liquids and gases with fine suspended solids. It often becomes overwhelming with particulate solids. Subsampling protocols of particulate wastes used by certified laboratories are often disastrous. It should be mandatory to write on analytical logs both:

— The maximum particle size of a subsample.
— The weight of a subsample.

This should be done for all sampling and subsampling stages, including the final subsampling stage at the balance room.

GE: The grouping and segregation error. This error is generated by the distribution heterogeneity. It is the variability of a pollutant from one increment to another. We know that the maximum reached by GE is FE itself. The most effective way to minimize GE is to increase the number of increments. It should be clear that for a given sample weight, the number of increments has no effect on FE.

In this category we include all wastes that are considered essentially as statistical populations. This includes nearly all samples containing particulate solids that are submitted to a subsampling operation, either on the field or at the laboratory.

21.4.2. Errors Generated by the Sampling of One-Dimensional Wastes

CE_1: the short-range heterogeneity fluctuation error introduced by h_1. It is a small scale random error and we may write:

$$CE_1 = FE + GE \tag{21.3}$$

CE_2: the long-range heterogeneity fluctuation error introduced by h_2. It is a large scale nonrandom error. It is generated by the pollutant trends, whose origin often goes back to the waste producer.

CE_3: the periodic heterogeneity fluctuation error introduced by h_3. Cycles may be directly caused by the waste producer, or they may have natural origins (e.g., fluctuations of a pollutant regulated by intermittent discharges, or by the effect of tides, or by weekend activities, and so on). This sampling error may become important if sampling is implemented at regular time intervals in phase with the period of the cycle.

21.4.3. Errors Generated by Incorrect Sampling

FE, GE, CE_1, CE_2, and CE_3 are the type of sampling errors that can be minimized by the development of appropriate sampling plans. However, there are other kinds of sampling errors that can be introduced during the implementation of a sampling plan.

DE: the increment delimitation error. The sampling tool should be capable of delimiting the volume of an increment in such a way that all parts of the waste have exactly the same probability of being part of the model increment.