P 716

NA-59-1

SAMPLING FOR QUANTITATIVE ANALYSIS

PROCEEDINGS OF A SEMINAR FOR RECIPIENTS OF RESEARCH GRANTS

Date : 16 October 1987

Venue : Auditorium, NARESA

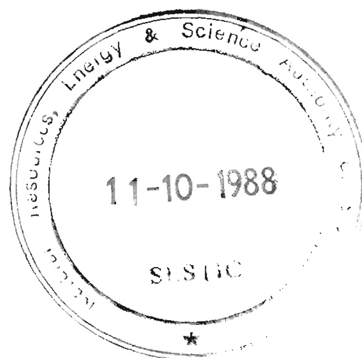
NA-59

Natural Resources, Energy & Science Authority of Sri Lanka

January 1988

SAMPLING FOR QUANTITATIVE ANALYSIS

Proceedings of a Seminar for Recipients of Research Grants, organized by the Statutory Working Committee on Chemical Sciences of NARESA, and held on 16th of October 1987.



Natural Resources, Energy & Science Authority of Sri Lanka

47/5, Maitland Place

Colombo 7

FOREWORD

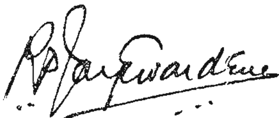
NARESA's scheme for award of grants for scientific research has many facets. Its main objective is to provide opportunities for young scientists to enhance research capabilities through what has been described as curiosity - oriented research.

As part of the training process for research, NARESA's Working Committees' organize seminars at regular intervals for grant recipients in order to expose them to peers and specialists. These seminars are generally discipline - oriented, at which grant recipients present the findings of research for critical review.

The Seminar on "Sampling for Quantitative Analysis" was a deviation. This was necessary to create a better understanding of the importance of collecting representative samples for quantitative studies.

Several specialists in the field accepted our invitation to present technical papers and also participate in the panel discussion.

I wish to thank the organizers, resource persons and participants for their contributions to the seminar, and for their assistance to bring out this publication.



Director-General

CONTENTS

		Page
	Table of Contents	iii
	Preface	iv
	Programme of Seminar	v
	Participants	vii
1.	An Overview - Sampling of Biological Materials M.A.T. de Silva	1
2.	Statistical Considerations in Sampling for Quantitative Analysis V. Abeywardena	9
3.	Practical Difficulties in Selecting a Representative Sample M.D. Paranavitana	23
4.	Sampling in Biology and Medicine S. Mendis	28
5.	Sampling for Air Particulates M. Ponnambalam	41
6.	Sampling of Water for Analysis of Cations and Anions H.D. Gunawardhana	52
7.	Sampling for Analysis of some Selected Food Materials N.M. Pieris	62

PREFACE

Although the main function of the Chemical Sciences Working Committee of Natural Resources, Energy and Science Authority of Sri Lanka (NARESA) has been its recommendation of award of Research Grants and the periodic monitoring of progress of these projects, the Committee also seeks to foster research in the Chemical Sciences by encouraging researchers to attain still higher standards.

One action in attempting to attain the above goal has been the organization of an Annual Seminar on a selected topic of importance where the target group has been the Research Grantees and their assistants.

Studying the reports of the Research Grantees it was felt that an in-depth seminar on sampling would be of special use to the younger researchers. Therefore a seminar with an 'out of the ordinary' topic of "Sampling for Quantitative Analysis" was organized.

The seminar had the following basic structure -

(a) Experts in statistics and sampling delivered lectures on the basic principles of statistical sampling;

(b) Researchers of repute shared their experiences on how they handled intriguing sampling problems that they had encountered in the process of their investigations.

The seminar produced a number of high quality presentations with outstanding case-studies that are bound to give a considerable insight into the type of sampling problems a researcher is likely to encounter and the methods of countering these problems. The quality of presentations was such that it was deemed important that the proceedings of the seminar to be published in order that its findings could be widely distributed among the scientific research community.

E.R. JANSZ
CHAIRMAN
CHEMICAL SCIENCES WORKING COMMITTEE
NATURAL RESOURCES, ENERGY & SCIENCE AUTHORITY OF SRI LANKA

SEMINAR ON 'SAMPLING FOR QUANTITATIVE ANALYSIS'

Date : Friday 16th October 1987

Time : 9.00 a.m. - 4.15 p.m.

Venue : Auditorium, NARESA

P R O G R A M M E

Inauguration	- Chairman	-	Addl. Director-General/ NARESA Mr L.C.A. de S.Wijesinghe
9.00 a.m.	- Welcome Address	-	Addl. D.G./NARESA
9.05 a.m.	- Overview	-	Mr M.A.T. de Silva NARESA
SESSION I	- Chairman	-	Dr E.R. Jansz
9.30 a.m.	- Statistical considerations in sampling for Quantitative Analysis	-	Mr V. Abeywardene Agroskills Ltd.,
10.00 a.m.	- T E A		
10.15 a.m.	- Practical difficulties in selecting a representative sample	-	Mr P. Paranavitana Sri Lanka Standards Inst.,
10.45 a.m.	- Sampling in Biology & Medicine	-	Dr (Mrs) S. Mendis Univ. of Peradeniya
11.15 a.m.	- Sampling air for particulates	-	Mr M. Ponnambalam Dept. of Labour
12.00	- L U N C H		

SESSION II	- Chairperson	-	Prof (Mrs) W.P.D. Pereira
2.00 p.m.	- Sampling of water for analysis of Cations & Anions	-	Dr H.D. Gunawardena Univ. of Colombo
2.30 p.m.	- Sampling for Analysis of some selected food materials	-	Dr (Mrs) N. Pieris - CISIR
3.00 p.m.	- **T E A**		
3.15 p.m.	- Panel Discussion	-	Dr E.R. Jansz (Chairman) Mr T. Kandasamy Dr U. Samarajeewa Dr A. Bamunuaarachchi Mrs M.J. de Silva
4.00 p.m.	- Summing up	-	Prof C.B. Dissanayaka

Participants

	Name	Institution
1.	Dr E.R. Jansz	CISIR
2.	Prof.(Mrs) W.P.D. Pereira	University of Colombo
3.	Prof. A.A.L. Gunatilaka	University of Peradeniya
4.	Prof. C.B. Dissanayaka	University of Peradeniya
5.	Mr M.A.T. de Silva	NARESA
6.	Mr V. Abeywardena	Agroskills Ltd.,
7.	Mr P. Paranavitana	Sri Lanka Standards Institute
8.	Dr(Mrs) S. Mendis	University of Peradeniya
9.	Mr M. Ponnambalam	Dept. of Labour
10.	Dr H.D. Gunawardhana	University of Colombo
11.	Dr(Mrs) N. Pieris	CISIR
12.	Mr T. Kandasamy	Retired, Government Analyst
13.	Dr U. Samarajeewa	University of Peradeniya
14.	Dr A. Bamunuaarachchi	University of Sri Jayewardenapura
15.	Mrs M.J. de Silva	City Analysts Laboratory
16.	Mrs Malinie Senanayake	Dept. of Minor Export Crops
17.	Dr(Mrs) C.M. Arewgoda	University of Kelaniya
18.	Dr G.S. Jayatileka	CISIR
19.	Mrs D. Rajapakse	CISIR
20.	Mr J. Ranatunge	CISIR
21.	Mrs T.D.Ranatunge	CISIR
22.	Miss M.J. Gunaratna	CISIR
23.	Mr P.G. Madawala	CISIR
24.	Dr P. Abeygunawardena	University of Peradeniya
25.	Dr A.L.T. Perera	University of Peradeniya
26.	Dr G. Keerthisinghe	University of Peradeniya
27.	Dr L.G.G. Yapa	University of Peradeniya
28.	Dr A.N. Jayakody	University of Peradeniya
29.	Mr Lalith Jayesinghe	University of Peradeniya
30.	Mr C.B. Ratnayake	University of Peradeniya
31.	Mr N.J. Liyanage	University of Peradeniya
32.	Mr B.V.R. Punyawardena	University of Peradeniya
33.	Miss J. Hettiaarachchi	University of Peradeniya

34.	Miss K.A.D.N. Perera	University of Colombo
35.	Dr(Mrs.) C.S. Wijewardene	University of Sri Jayewardenepura
36.	Miss S. Adikari	University of Sri Jayewardenepura
37.	Dr(Miss) H.M.K.K. Pathirana	University of Ruhuna
38.	Mr N.H. Dassanayake	NARA
39.	Dr(Miss) P. de Alwis	NARA
40.	Mr V. Perera	RRISL
41.	Miss U. Alwis	Labour Department

NA-59-2

AN OVERVIEW - SAMPLING OF BIOLOGICAL MATERIALS

M.A.T. de Silva

NARESA

One of the main objectives of NARESA's research grant scheme is to provide opportunities for young scientists to enhance research capability through the provision of supplementary resources. In doing so NARESA has also been mindful of the need to provide facilities for orientation, training and updating of knowledge through workshops and seminars, where opportunities are available for interaction with fellow workers and senior scientists. The present seminar is one such meeting organized by the Working Committee on Chemical Sciences.

The Working Committee on Chemical Sciences has been somewhat concerned about the lack of attention paid to sampling of materials, especially of biological materials, for scientific investigations. Both junior and senior researchers have been found to falter, resulting in data of doubtful validity. It is for this reason that the Working Committee decided to organize this seminar on the theme "Sampling for Quantitative Analysis". The main objective of the Seminar is to resurrect and review the subject of sampling of materials for experimental work. It will deal with both biological and non-biological material, and also develop the theoretical framework for sampling.

Biological materials as we all know, are highly complex substances, exhibiting specific physico - chemical properties, and existing in dynamic equilibrium with the environment. Therefore the composition of such material, at any given moment are determined by several factors, many of which are external (eg. diurnal and seasonal effects, moisture stress, supply of minerals etc.). Hence in developing a sampling technique for biological substances, a careful study should be undertaken initially, of the effects of all such factors, on the composition of such materials. Similar issues are also known to influence material of non-biological origin. However, the degree of precision required in sampling will of course, depend on the type of investigation that is to be undertaken, and hence has to be decided by the researchers.

In this overview, it is proposed to discuss some of the pitfalls and constraints in sampling of biological material for scientific research, taking illustrative examples from the authors own experiences. The work discussed here relates to some nutritional studies on the coconut palm, done during the period 1955 to 1977.

Initially, it has to be noted that the coconut palm being a perennial of long life, responds relatively slowly to external changes, and therefore experimental work needs careful planning, since rectifying any errors can be both time consuming and costly.

During the mid 1950's the widespread occurrence of a yellowing of leaves of coconut palms,led to an extensive study. Although it was presumed to be due to a deficiency of Magnesium, it was realized that such symptoms appeared only when the disorder had reached an advanced state. It was therefore necessary to develop a method to diagnose the disease at the onset of the deficiency. The obvious choice was leaf analysis, and since the general technique of leaf analysis for nutritional studies had already been established by French workers, it was decided to adopt these methods.

French scientists found that the mineral composition of the 14th frond in order of maturity, provided an insight into the nutritional status of the coconut palm. Our efforts to use this method for diagnosing magnesium deficiency resulted in repeated failures. Hence when it was decided to re-examine the sampling technique, it was found that if a younger leaf, especially the 6th frond from the top was used, instead of the mature 14th frond the deficiency could be diagnosed long before the visual symptoms appeared (1). Now this was a case in which an incorrect sampling procedure, led to a delay of several years in developing a quick diagnostic method for a nutritional disorder.

In the early 1970's, we entered the relatively new field of micronutrients. Learning from past experiences, it was decided that before any nutritional studies could commence, the suitability of different materials, techniques for sampling and analytical methods should be carefully studied and evaluated. As was done in the earlier occassion, the mineral composition of the leaf was examined in detail during the initial phase. It was known by this time that the nutrient contents of the coconut frond varied

not only from the proximal to the distal end, but also from the rachis to the tips of leaflets. More surprising were the differences and gradients in the mineral composition of laminae in the same leaflet. Hence a careful re-examination had to be done on leaf composition in order to develop a reliable method for sampling.

In one such investigation. triangular pieces cut outfrom the laminae at regular intervals were compared with whole leaf samples(See Fig.1). Differences were observed between these two leaf sampling techniques, provided sufficiently representative composite samples were prepared (2).

MICRONUTRIENTS IN THE NUTRITION OF COCONUT - I

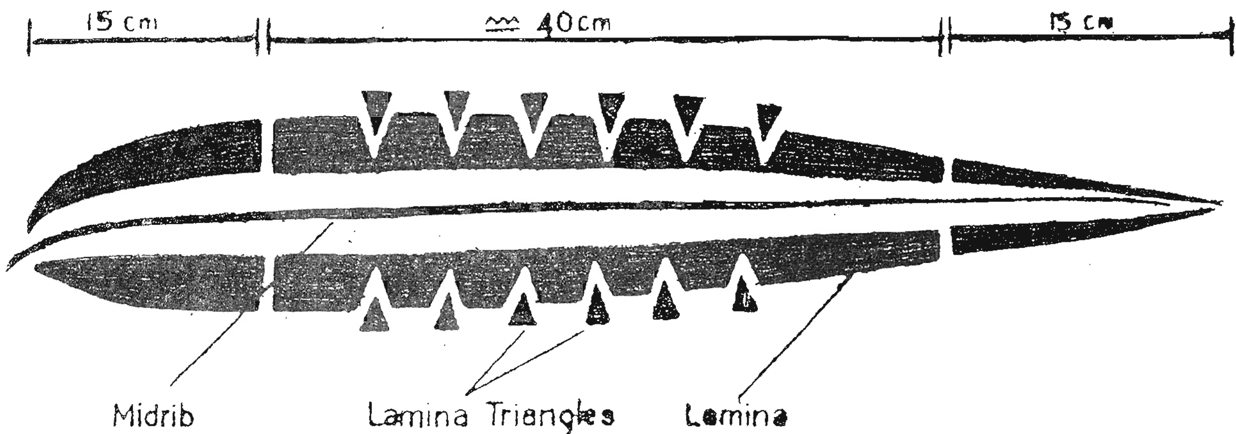


Figure 1 - Diagramatic representation of the sampling procedure for Lamina "triangle"

Although it was known that diurnal changes occured in the mineral composition of leaves, this error factor was eleminated by determining a specific time interval for collection of samples.

In the next phase of the study, the effect of leaf maturity on the mineral composition was studied (2). It was observed that the distribution of each of the micronutrients under investigation behaved

differently with respect to leaf maturity (2). Obviously it meant that different sampling methods would have to be formulated for each of the nutrients under investigation (Fig 2).

In the third phase of this study it was decided to examine other componants of the coconut palm for nutritional studies. One such material was the sap (coconut toddy), which had proved useful for certain other investigations. The data summarized in figures 3 and 4 show the diurnal and seasonal changes in the micronutrient composition sap over a period of six months. The sharp fluctuations observed clearly showed that this material was not suitable for nutritional studies.

Sampling of biological material for nutritional studies can be even more complicated when the plants are being subjected to differential treatments, as for example in a factorial experiment. Here one has to be careful in identifying the role of other elements, in determining a sampling technique for one element. Such a problem was witnessed when a sampling technique was being devised to study the sulphur nutrition of the coconut palm. The data presented in table 1 shows the "t" values for significance of regression for effects of applied sulphur on the sulphur concentration of leaves of different degrees of maturity (3).

	Leaf position				
Factor	1st	6th	10th	14th	17th
B	-2.02	- 3.13*	0.24	-1.29	-0.25
Zn	-0.23	0.27	-2.74*	0.63	-0.83
S	-0.86	- 3.64*	-1.65	0.37	0.25
B^2	-0.05	4.46**	8.31***	3.20*	-2.65*
Zn^2	-2.95*	3.46*	7.34***	0.73	-4.01*
S^2	1.58	13.22***	6.93***	9.14**	4.87**
BxZn	0.51	- 0.57	-2.88*	-1.07	-0.67
BxS	0.33	2.82*	0.98	0.61	0.22
ZnxS	-0.33	0.44	0.55	-1.26	0.68
Multiple correlation coefficient	0.7518	0.9455	0.9045	0.8615	0.7607

*, **, *** significant at P = 0.05, 0.01 and 0.001

TABLE 1 : t values for significance of regression for effects of applied S on the S concentration of different leaves *

* After M.A.T. de Silva etal (3)

Here the dimensions of the multiple correlation coefficients were used to evaluate the sensitivity of treatments involving sulphur, and by this means it was possible to identify the leaf (6th in order of maturity), most suitable for nutritional studies on sulphur.

The examples discussed in this presentation clearly show the complexities involved in devising good sampling procedures for biological materials. It is obvious that in any investigation, if reliable data are to be collected, a carefully devised sampling procedure must be evolved for the material that is to be studied.

REFERENCES

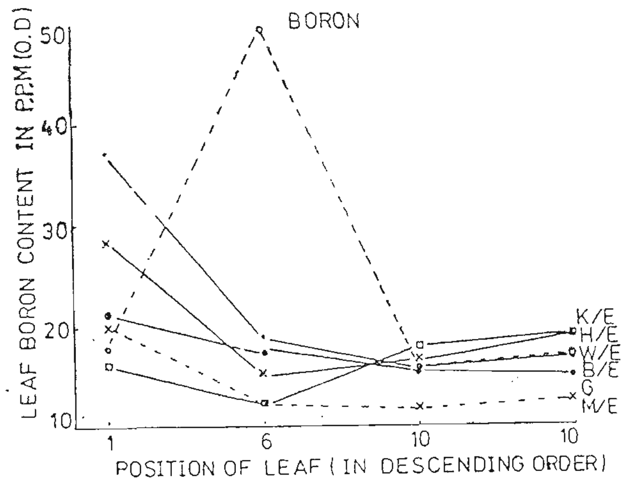
1. Nethsinghe, D.A., De Silva, M.A.T. and Nedimala, D.G. (1962). Diagnosis and correction of Magnesium Deficiency in Coconut Palms. Proc. Ceylon Associ. Adv. Sci (1962), Part I, 16-17.

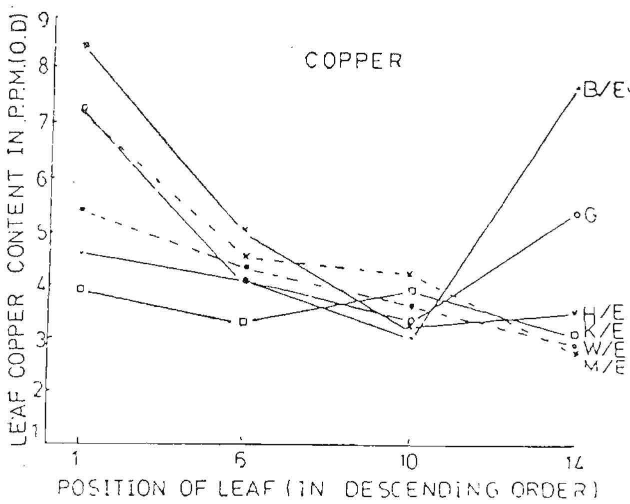
2. De Silva, M.A.T. (1974). Micronutrients in the Nutrition of Coconut. I Methods and Preliminary Investigation. Ceylon Cocon. Q., 24, 116 - 127.

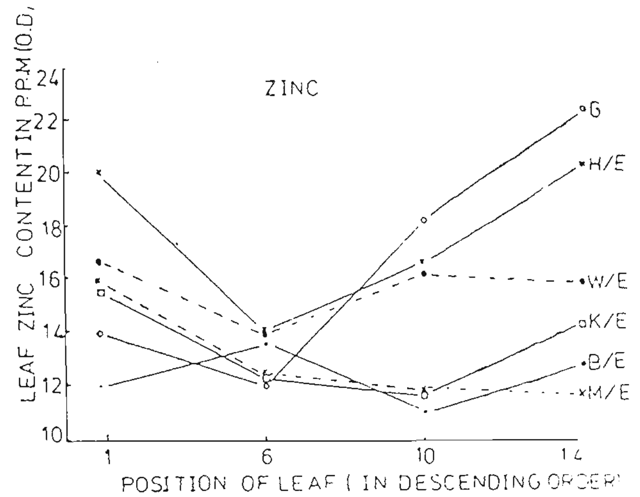
3. De Silva, M.A.T., Anthonypillai, Greta M. and Mathes, D.T. (1977). The Sulphur Nutrition of Coconut Expl. Agric. 13, 265 - 271.

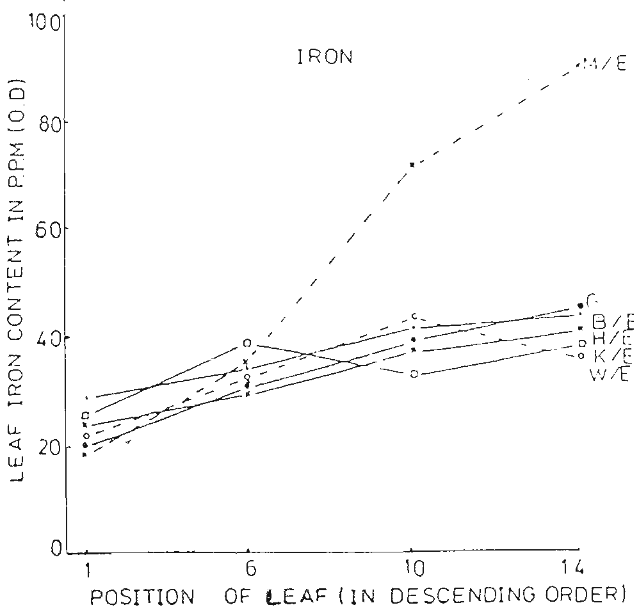
CONTENTS OF Fe, Mn, Cu, AND B, IN COCONUT LEAF SAMPLES WITH RESPECT TO LEAF POSITION

LOCATION	SOIL TYPE	SOIL pH
—— BANDIRIPPUWA ESTATE LUNUWILA	SANDY LOAM	5.60
×——× HOREKELLE ESTATE KUDAWEWA	CINNAMON SAND	5.79
•——• BELIGAMA ESTATE GALEWELA	LIMESTONE DERIVED BROWN SOILS	6.08
o——o KIRIMETIYANA ESTATE LUNUWILA	CLAY LOAM	5.28
×- - -× MARANDAWILA ESTATE BINGIRIYA	SANDY LOAM	5.70
•- - -• WALAHAPITIYA ESTATE HATTANDIYA	LATERITIC LOAM (GRAVEL)	5.50









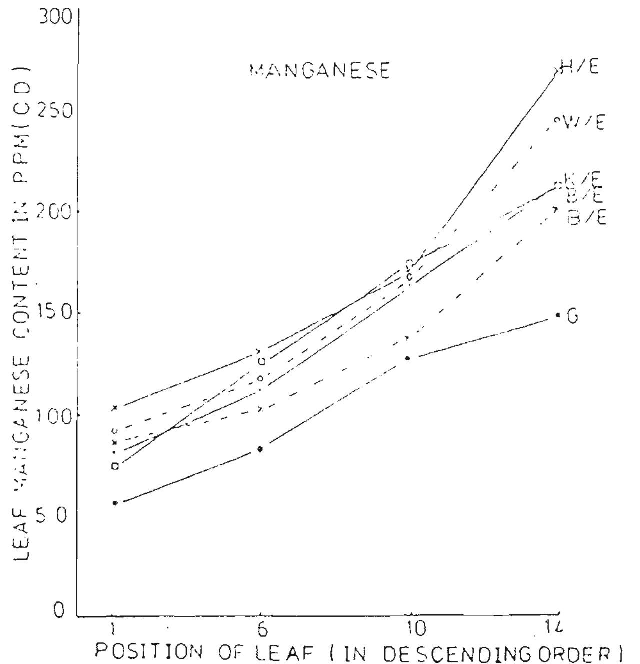


Fig. 2. Contents of Fe, Mn, Cu, Zn and B in coconut leaf samples with respect to leaf position.

(After M.A.T. de Silva et al (3))

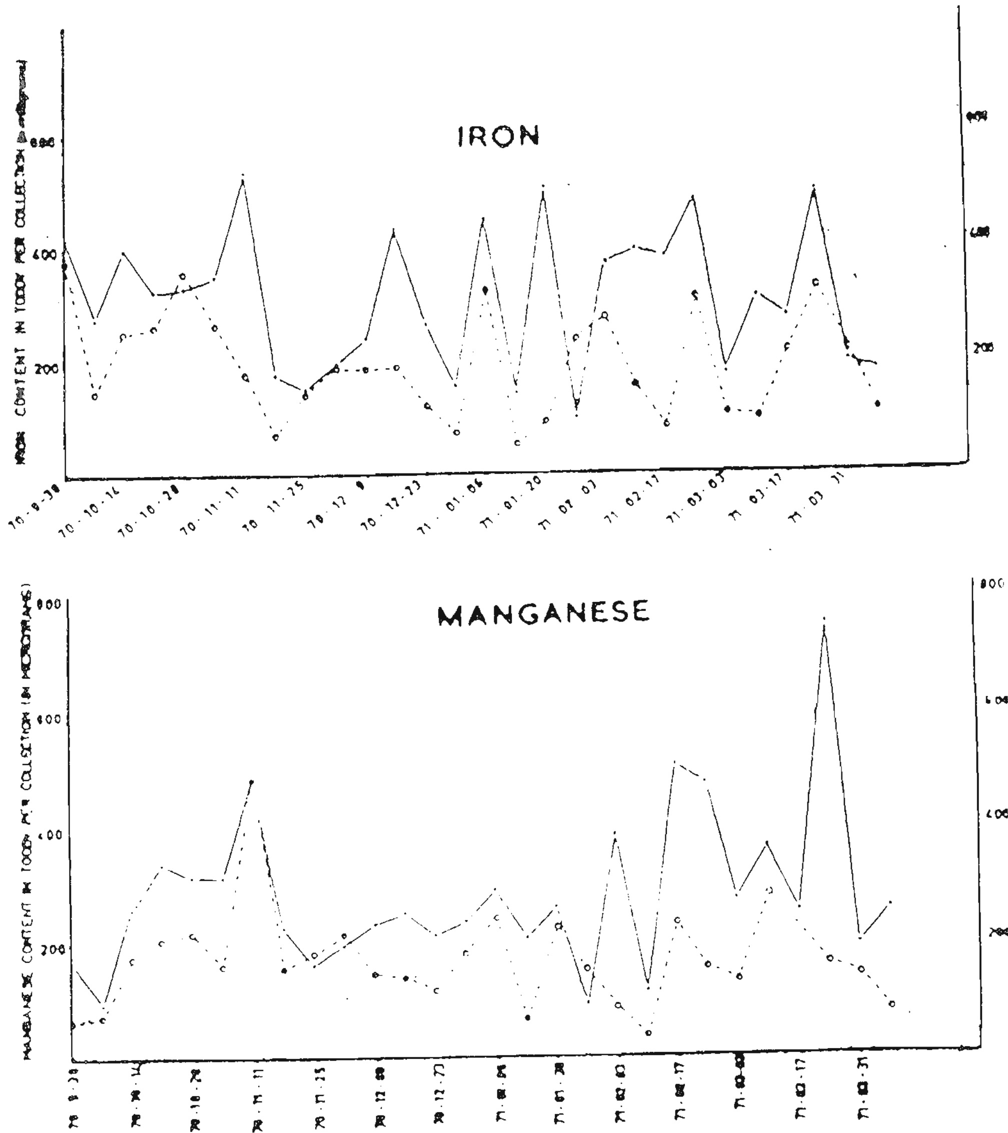


Fig. 3 Weekly variation of Fe and Mn in the sap (toddy) of two palms during a period of six months.

(After M.A.T. de Silva et al (3))

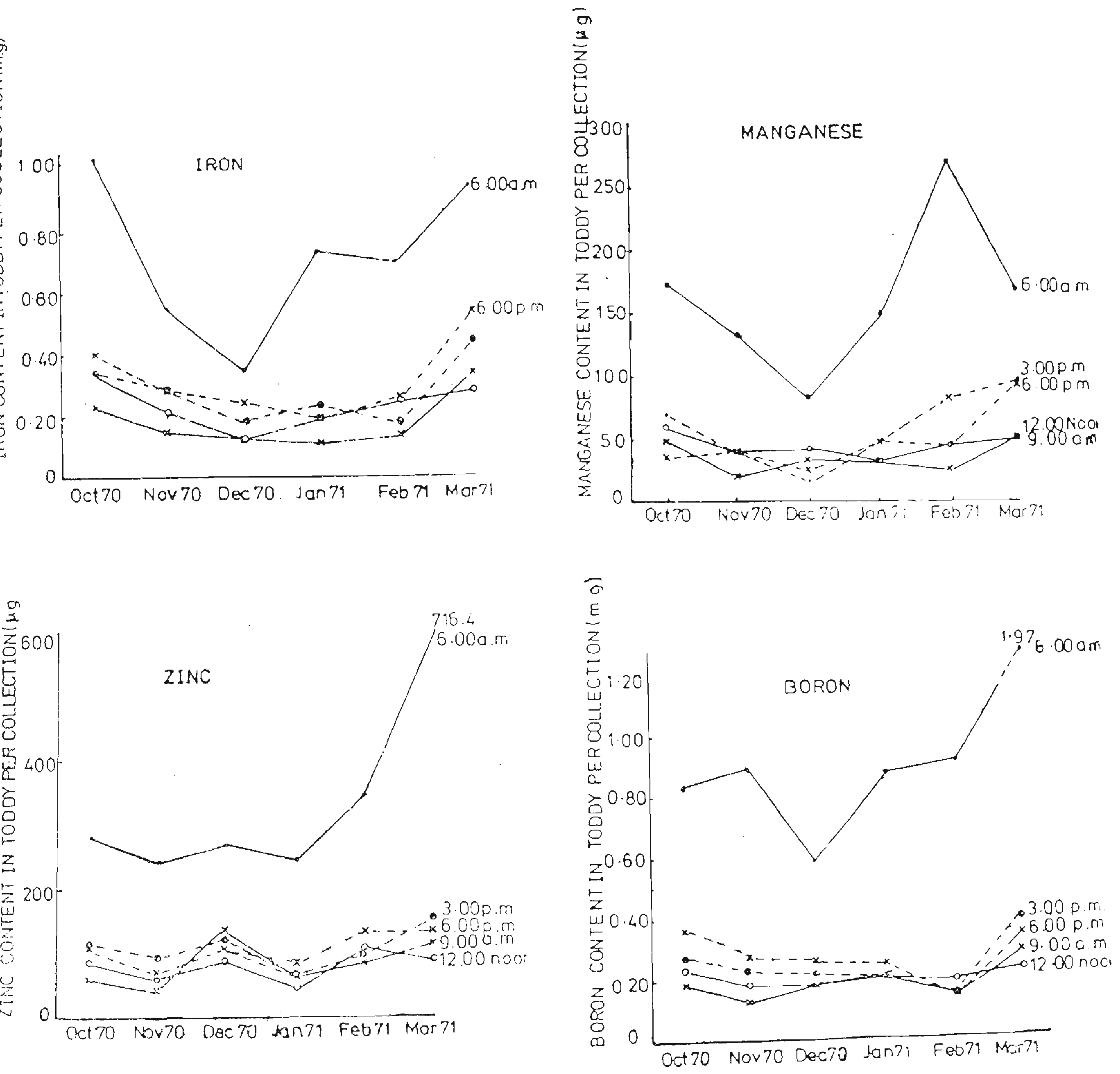


Fig. 4. Diurnal variation of Fe, Mn, Zn and B in the sap (toddy) of coconut palms during a period of six months. (The plots represent the means of 6 palms, and the 6.00 a.m. curve in each graph represents the contents in the collection over a 12-hour period commencing at 6.00 p.m. the previous day).

(After M.A.T. de Silva et al (3))

NA-59-3

STATISTICAL CONSIDERATIONS IN SAMPLING FOR QUANTITATIVE ANALYSIS

V. Abeywardena,

Agroskills Ltd.,

Introduction

Certain definitions of statistics not commonly heard of are that "statistics is the mathematics of the random sample" or that "statistics is the mathematical approach from the sample to the population enabling one to generalise from limited experiences". Such definitions are of immediate relevance to the topic under discussion in this seminar. Given that these definitions are true and meaningful, it is only appropriate that some time has been allocated during this seminar to discuss the role of statistics in investigations based on samples.

Sampling Method

A sampling method is a method of selecting a fraction of the population (or the whole) in a way that the selected sample represents the population. Thereby whatever measurement or count that is made on the sampling units will eventually lead to the calculation of a statistic such as the mean or a percentage which would be an unbiased estimate of the corresponding population parameter.

If it is a measurement X_i on n sampling units, then

$$\bar{X} = \sum_{i=1}^{n} X_i / n \longrightarrow \mu \text{ (the population mean)}$$

$$\text{as } n \longrightarrow \infty$$

Or if it is a count (n_k) of units possessing a specified characteristic out of the total of n sampling units, then

$$n_k / n \longrightarrow P \quad \text{(the population proportion)}$$

$$\text{as } n \longrightarrow \infty$$

By way of generalisation, it can be stated that provided the sample is drawn according to certain rules, the sample statistic would tend to the population parameter as the sample size tends to the complete enumeration.

This sampling approach to understanding the population is as old as society itself. It has been almost instinctive for a person to examine a few articles in order to understand the whole. With the passage of time, scientists began to identify their sources of variation and simultaneously they became more intelligent to making their samples more and more closely representative of the populations they are confronted with. With such a rapid development of the sampling techniques, meaning the physical aspects of sampling, scientists worked quite confidently on a tacit assumption that their sample estimates approximate very closely to the population values. However, with the development of statistical methods and sampling theory during the first half of this century, scientists began to think in terms of a measurable sampling error. This development virtually revolutionized the sampling approach to research.

Sampling error

What is meant by the sampling error? The sampling error does not signify any flaw in the method of sampling, it is something inherent in samples. Immediately a sample is drawn, there is an error tag attached to it. Whatever the method adopted to select the sample, however closely representative of the population the sample be, the sample estimate will inevitably differ from the one that would be obtained from enumerating the whole population with equal care. Even estimates from two samples of identical size drawn from the same population will differ one from the other. The amount by which a sample statistic will

on the average differ from the population parameter is termed the sampling error.

The measurement of sampling error

In spite of the successive refinements to the physical aspects of sampling introduced by scientists over the years, sample estimates could not command unequivocal scientific acceptance. The fact that there was no quantification of the reliability of such estimates or our inability to give a quantified assessment of the closeness of the sample estimate to the population parameter, militated against such acceptance.

Suppose a soil scientist wants to assess the potassium status of the soil in a coconut estate. He would, in the first instance, think of every conceivable domain of variation of the soil potassium content such as (say) different soil depths, different distances from the base of the palms, different elevations etc. Having identified the domains of variation, he would obtain soil samples to represent each such domain of variation - the number of samples to represent each such domain being made roughly proportional to the area under each domain. These soil samples are mixed well and sub-samples removed for laboratory analyses. These analyses would invariably reveal that each sub-sample yields differing values. Now in order to get over this discomforting situation, he would average out the values of each sub-sample and then conclude that this overall average is the best estimate of the population value.

We may concede that this is a good estimate. But we are not in a position to make a quantitative assessment as to how good the estimate is. With the development of the sampling theory however, we are enabled to calculate the sampling error of the estimate and thereby maintain some control over the capricious sample. Even though the population value itself is not known in absolute terms, one could calculate the range within which the population value will lie for any specified level of probability. For instance weare now in a position to make statements such as "I am 95% certain that the population mean lies within this interval" or "I am 90% certain that the error of the estimate will not exceed 10%. With this ability to make such firm statements, we know how reliable our sample estimates are.

Mathematics of the sampling error

The mathematics of the sampling error will herein be developed only for the case of the simplest form of sampling termed "Simple Random Sampling".

CHART 1

Let x_1, x_2 .... (x_i) .... x_{n-1}, x_n represent some measurement on the n sampling units in a sample of size n.

(1) Est. of the population mean is given by :

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

(2) Est. of the population variance given by :

$$\sigma^2 = \left[\sum x_i^2 - \left(\sum x_i\right)^2 / n \right] \div (n-1)$$

(3) Variance of the mean is given by :

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

(4) Standard error of the mean is given by :

$$\sigma_{\bar{x}} = \frac{\rho}{\sqrt{n}}$$

(5) Sampling error is given by :

$$t. \frac{\sigma}{\sqrt{n}}$$

(6) Error percentage is given by :

$$\left(t. \frac{\sigma}{\sqrt{n}} \div \bar{x}\right) \times 100$$

(7) Confidence interval is given by :

$$\bar{x} \pm t . \frac{\sigma}{\sqrt{n}}$$

note : Indices 5,6 and 7 always have a probability tag attached to them. Therefore the "t" value should represent the desired probability level and n-1 degrees of freedom.

It is now seen that with the aid of the standard error calculated from the sample value, one can set limits within which the population value will lie to any specified degree of probability or calculate the percentage error in the estimate at any specified level of probability.

Other uses of the standard error

Apart from this, there is yet a more important use of the standard error. Using the standard error one could determine what sample size should be employed in future comparable situations in order to obtain estimates to any specified degree of precision. A question very often posed to a statistician by scientists in other disciplines is about sample size. What should be my sample size? To this question most statisticians may give an evasive reply such as "the larger the sample size-the more reliable the estimate". This answer of course stems from his knowledge that the standard error varies inversely as the square root of the sample size. This answer however, does not take the researcher anywhere and the statistician could not have done better either. Anyway this question should have been posed differently such as "If I want to keep my error margin (say) within 10% with a probability of 95%, what should be my sample size?" Such a question would be more to the point. Anyway this question cannot be answered unless an estimate of the standard error (from a previous study) in respect of the variable in question or a related variable is available. In the absence of this background information and if one is still keen on obtaining an estimate with the desired precision, it would be necessary to conduct a small

pilot trial to get an estimate of the standard error.

Control of sampling error

(a) Adjustment of sample size

It is already noted that the sampling error can be controlled by varying the sample size. If the sample size is increased the sampling error is reduced and vice versa if the sample size is reduced the sampling error is increased. However controlling the sampling error by the adjustment of sample size is a very costly approach. To reduce the sampling error by half, one would have to increase the sample size fourfold and to reduce the sampling error to a third, one would have to increase the sample size ninefold.

(b) Stratified sampling

A more efficient method of reducing the sampling error is through the method of "stratified sampling".

In this method, the area to be sampled or the population to be sampled is divided into a number of strata such that the character under investigation (or some other character closely related to it) can be expected to be homogeneous within each stratum. In short the stratification should aim at maximising "intra stratum homogeneity and inter stratum heterogeneity".

Securing intra startum homogeneity can be effected in two ways. One can either identify naturally occuring homogeneous strata or one can induce homogeneity artificially by grouping the sampling units on the basis of some criterion related to the character, one is researching on.

The following are a few illustrations of the first method; that is, the identification of naturally occuring homogeneous strata.

i. If the study is conducted in a hilly terrain and if the character under investigation is related to soil fertility or atmospheric temperature or humidity, (such as, for instance, soil nutrients, foliar nutrients or soil micro flora and micro fauna), stratification along the contours should help.

ii. If one is studying some matter or micro organism in suspension in a flowing river, one could think of sampling from a series of imaginary strata (stretching parallel to the flow) from the river bank to midstream.

iii. In irrigated sugarcane, water is let out to the plantation through furrows running almost at right angles to the direction of the water channel. Thereby a soil moisture gradient is created from the proximal ends to the distal ends of the furrows. In such a context, if one is studying some character related to soil moisture (eg. leaf nutrients or sugar content) the strata should be strips parallel to the channel and across the furrows.

As examples of forming strata artificially, one could consider the following :

i. Suppose one is interested in estimating the oil content in rubber seeds, the heap of seeds can be divided into different groups on the basis of size by using a series of trays with a base of wire netting of progressively increasing sizes. Each such group would constitute a stratum.

ii. If one is studying the blood haemoglobin content in a herd of cattle fed on a newly introduced pasture and if one feels that the age of cattle has a bearing on the Hb content, the strata can be formed by grouping the animals according to age groups.

Having stratified thus, one could take random samples from each stratum to make up the whole sample. For instance if one has planned for a total sample size of (say) 30, and if the number of strata happens to be 6, then from each stratum one has to select 5 samples, assuming the strata are of equal size. In the event of the strata being of unequal size, the total sample can be allocated proportional to the size of the strata.

It may not be an exaggeration to say that through this stratified approach one can improve the precision of the estimate more than fourfold

while keeping the sample size unchanged or one can obtain a very precise estimate with a smaller sample.

(c) Use of concomitant variables

Yet another efficient approach to reduce sampling error is by regressing on some related concomitant variable. If, for instance, one is interested in assessing the N content in Hevea leaves, one should think of some other measurable variable which could be related to leaf N. Perhaps such a variable may be tree vigour which can be indicated by tree girth. Thus in each leaf sample in a given tree, while we obtain the N content (y) through chemical analysis, we also get a measure of the tree girth (x) to serve as the concomitant variable. Then by a statistical technique termed "covariance analysis" wherein one adjusts y for x, the sampling error of the mean leaf N (ie-$\bar{Y}$) can be reduced appreciably. This method can be extended to include more than one concomitant variable and multiple covariance techniques can be adopted for the control of the sampling error. For instance in the problem of leaf N estimation one could use in addition to tree girth, a variable such as soil N near the base of the tree.

A fascinating feature in this approach is that there is provision in the statistical analysis to verify whether the concomitant variables one selects on hypothetical grounds, are actually related to the variable under investigation. Adjustments are effected only if the relationship is significant.

Multi-stage sampling

(d) Methodology

Multi-stage sampling has been resorted to by biologists for quite some time now. But apart from calculating a mean value and a few instances of calculating a sampling error, researchers have not put such studies to optimum use for the benefit of future research progress. It is my considered opinion that this exposition will not be complete unless the uses of multi-stage sampling are included.

Suppose a research officer needs to study the foliar nutrient status of some plant crop. He knows that there are differences between trees, so he takes a number of trees at random; he knows that there are differences between leaves of the same tree, so he takes a number of leaves at random from each selected tree; he knows that there are differences in repeat determinations on the same leaf, so he does two or more determinations on each leaf. The fact that there is an attempt to cover every domain of variation and thereby obtain a good overall estimate is very clear. But does he know how good his estimate is? Or does he know the variability between trees, or between leaves within a tree or between determinations within a leaf. Obviously he would not have the haziest idea. Therefore faced with data of this type we should not rest content with an average calculated from the data as is the common practice. We should analyse the data further, with a view to (1) getting some measure of the probable error of this average and (2) planning optimum sampling schemes for future use on similar problems. In essence the first investigation, while it is helpful *per se* should also serve as a pilot trial for future investigations.

Suppose a sampling scheme for the estimation of some leaf nutrient level for Hevea consisted of a = 4 trees, b = 3 leaves from each tree and n = 2 determinations on each leaf (constituting in all 24 determinations). The following values may have been obtained.

Plant a = 4	Leaf b = 3	Determination n = 2		Leaf Total	Plant Total	Grand Total (GT)
		1	2			
1	1	4.2	4.1	8.3		
	2	3.9	4.0	7.9		
	3	3.4	3.2	6.6	22.8	
2	1	2.0	2.2	4.2		
	2	1.7	1.9	3.6		
	3	2.6	2.4	5.0	12.8	
3	1	5.0	4.7	9.7		
	2	4.8	4.4	7.9		
	3	3.9	4.0	7.9	26.8	
4	1	3.6	3.5	7.1		
	2	3.1	3.3	6.4		
	3	3.9	3.7	7.6	21.1	83.5

(e) Statistical analysis

This type of sampling involving subsampling at different stages, gives rise to a hierachial classification and the analysis of variance takes a different form from that applicable to orthogonal classifications. The details of the analysis of variance are not shown here but the final Anova is reproduced below.

Analysis of Variance

Source of Variation	D.F	S.S	M.S	F
Plants	3	17.34	5.7800	18.35*
Leaves wn plant	8	2.52	0.3150	14.52*
Determinations wn Leaf	12	0.26	0.0217	

From the above it can be concluded that there are significant differences between trees and between leaves within trees.

(f) Sample estimate

(1) The mean leaf nutrient content of Hevea leaf is given by :

$$\bar{x} = 83.5/24 = 3.50$$

(2) The variance of the mean is given by :

$$S^2_{\bar{x}} = \frac{5.78}{24} = 0.2408$$

(3) The standard error of the mean is given by :

$$S_{\bar{x}} = \sqrt{0.2408} = 0.49$$

(4) 95% probable error is given by :

$$t.S_{\bar{x}} = 2.069 \times 0.49 = 1.01$$

(5) Percentage sampling error (95% probability) is given by :

$$(t.S_{\bar{x}} \div \bar{x}) \times 100 = \frac{1.01}{3.50} \times 100 = 29\%$$

(6) 95% confidence interval is given by :

$$\bar{x} \pm t.S_{\bar{x}} = 3.50 \pm 1.01 = 2.49 < \mu < 4.51$$

(g) Variance components

Through the above statistical analysis, we have for all intents and purposes fulfilled the objectives of the study. We have an estimate of the foliar nutrient content, we know the probable error of the estimate and we even know the limits within which the population value will most probably lie.

But there is a lot more information to be tapped from the data collected. At the very outset of this investigation, the researcher had sufficient ingenuity to identify three sources of variation that the overall sample would be subject to. He expected variation between trees; he expected variation between leaves within a tree and he expected variation between determinations within a leaf. Therefore it makes sense that this data be made use of in the first instance to estimate the dimensions of these three sources of variation so that if a particular source of variation is negligible, he can ignore it in future studies or if a particular source of variation is very large he can give more weightage to it.

The dimensions of these sources of variation can be estimated through certain statistics termed "Variance Components".

Suppose the number of trees is "a", the number of leaves within a tree is "b" and the number of determinations within a leaf is "n". The analysis of variance and the structure of the "mean square" in terms of the variance components are shown below.

Source of Variation	D.F	M.S.	Variance Components
Plants	a-1	M_p =	$\sigma_d^2 + n\sigma_e^2 + nb\sigma^2$
Leaves wn Plant	a(b-1)	M_e =	$\sigma_d^2 + n\sigma_e^2$
Determination wn leaf	ab(n-1)	M_d =	σ_d^2

Thus the variance components for the example given would be as follows:

Source of Variation	Variance component
Bn determinations	$\sigma_d^2 = M_d = 0.0217$
Bn leaves wn tree	$\sigma_e^2 = (M_e - M_d)/n = 0.1466$
Bn plants	$\sigma_p^2 = (M_p - M_e)/nb = 0.9108$

It is seen that between plants variation is very large and between determinations variation is negligible. This information can be made use of in future investigations.

(h) Planning future sampling schemes

One can make use of the variance components calculated from this study to develop efficient sampling schemes for the future, assuming of course that situations are fairly comparable.

The variance of the overall mean is given by :

$$\frac{M_p}{abn}$$

Substituting the variance components structure for M_p, the variance of the overall mean would be :

$$\frac{M_p}{abn} = \frac{\sigma_d^2}{abn} + \frac{n\sigma_1^2}{abn} + \frac{nb\sigma_p^2}{abn}$$

$$= \frac{\sigma_d^2}{abn} + \frac{\sigma_1^2}{ab} + \frac{\sigma_p^2}{a}$$

Thus the variance of the mean, which we calculated earlier as the ratio of the "between plant mean square" to the total sample size, can also be expressed as a function of the variance components as shown above.

Therefore making use of these variance components, we can estimate what the probable error would be for different sampling schemes to be adopted in the future and thereby plan to obtain estimates to any desired level of precision.

The standard errors applicable to some tentative sampling scheme have been calculated below :

Sampling Scheme	Total No. of Units	Variance of Estimate $\frac{\sigma_d^2}{abn} + \frac{\sigma_e^2}{ab} + \frac{\sigma_p^2}{a}$	$S^2_{\bar{x}}$	$SE_{\bar{x}}$
a=4; b=3; n=2	24	$\frac{0.0217}{24} + \frac{0.1466}{12} + \frac{0.9108}{4}$	0.2408	0.4907
a=10; b=2; n=1	20	$\frac{0.0217}{20} + \frac{0.1466}{20} + \frac{0.9108}{10}$	0.0995	0.3154
a=20; b=1; n=1	20	$\frac{0.0217}{20} + \frac{0.1466}{20} + \frac{0.9108}{20}$	0.0540	0.2323

The components of variance calculated earlier show that the components for "between plants" is relatively very large. Thus any attempt to reduce the error of the estimate should be by taking more units from that source namely plants and less so on determinations which has the lowest variance component.

In the worked out example above, while when we took 4 plants, 3 leaves from each plant and two determinations from each leaf, constituting an overall sample of 24 determinations, the standard error was as much as 0.4907, by doing only 20 determinations consisting of 10 plants, 2 leaves from each plants and one determination from each leaf, the standard error was lowered to 0.3154. Then again by taking 20 plants, one leaf from each plant and also one determination from each leaf, the standard error was further reduced to 0.2323. It is seen that while earlier with a sample size of 24, the standard error was 0.4907 giving rise to a sampling error of 29%, now with a sample of only 20, by a mere reallocation of the total sample, the standard error has been reduced to 0.2323 resulting in a sample error of only 14%.

Conclusion

Statistical methods and statistical sampling procedures are helpful not merely in the measurement and control of the error inherent in sampling but also in the identification and evaluation of the sources of variation, which latter would enable researchers to plan for sample estimates carrying pre-determined precision levels.

NA - 59 - 4

PRACTICAL DIFFICULTIES IN SELECTING A REPRESENTATIVE SAMPLE

M.D. Paranavitana

Sri Lanka Standardards Institution

Objectives of Sampling

A clear idea of the objective of sampling should be known before samples are collected. Objectives could be broadly categorised into three areas.

(a) For analysis, research and decision making.

(b) For control purposes (process control)

(c) For inspection (Acceptance Sampling)

It has often been observed that those engaged in testing and analysis is less interested in sampling and preparation of samples, though they are very enthusiastic of the results and the action they hope to initiate based on these results.

Those involved in manufacturing technologies are keen on improving such technologies, but they are apt to draw incorrect conclusions if they are not aware of the sampling errors involved.

Hence adequate seriousness should be given to the collection of samples.

Selection of Sample

The main activity in sampling is to obtain a representative sample from a lot or batch. When selecting a representative sample the following problems will have to be envisaged.

1. **Homogenity of the lot**

   This is of crucial importance especially in acceptance sampling. The lot should be homogeneous with respect to the property that is assessed. eg. In a PVC pipe manufacturing plant which has many extruders which could produce the same size of pipe with

the same batch of raw material, all the pipes will be homogeneous with respect to the toxicity levels since the same raw materials are used but will not be homogeneous with respect to size (diameter) as different extruders have been used although the same size has been manufactured. Hence if a lot is truly homogeneous even one sample is enough for decision making.. But invariably this is not so.

In case of industrial products, a lot may consist of a number of batches manufactured in different shifts, numbers of machines etc., In case of Agricultural products a lot may consist of products from different growing areas harvested at varying conditions and probably dried too under different conditions.

Because of the above reasons, a decision cannot be made only by taking one sample and hence the sample size has to be increased. Especially in acceptance sampling, if the lot is not homogeneous a manufacturer's risk of rejection is increased.

2. **Sample size**

A larger sample is more representative than a smaller sample. But the following aspects have to be carefully studied in selecting a larger sample.

a. Cost of Sample

If the product is to be destructively tested the cost of the product has to be considered as no manufacturer will like to bear this cost or else this will be passed on to the consumer.

b. Cost of testing

In testing where "life" of a product is involved, invariably the cost of testing is high.

c. Time for Testing

In testing products which are imported or exported time taken for testing plays an important part. If a test report is required for clearance of goods from the harbour, testing has to be quick as otherwise demurrage has to be paid and this cost will be passed onto the consumer as the importer will not like to bear it due to the delay on the part of the inspection agency.

Similarly for products to be exported, unless the test report is available quickly, there will be a delay in processing the export documents and thus the exporter will not be able to ship on schedule.

d. Increase of non - sampling cross

These are errors due to human involvement; chief of which are human fatigue and testing fatigue. If samples are to be taken in adverse condition such as in cold rooms where the temperature is - 20°C or in warehouses where there is hardly any ventilation these contribute to human fatigue. If a large number of samples are to be tested by one officer, it will contribute to testing fatigue.

3. Sample Selection and Preparation

a. Selection

In selecting samples it is of vital and paramount importance that biasness should be avoided and randomness should be ensured. It is common practice that when production staff is involved in selection they tend to select "good" products while if buyers select samples they tend to select "bad" products, if visually there could be a distinction.

Randomness could be ensured by the use of random number tables. For this it is necessary that conditions for selecting random samples should be created. In warehouses, package should be packed in a manner which could ensure easy retrieval. However, the most easy and sure way of selecting a random sample is selecting while the lot is "moving", i.e. while loading or unloading. If the unit product is packed in cartons, cartons could be selected at random and from each carton the required number of products could be selected.

Whenever possible and practical, sampling instruments should be used to select samples. If samples are to be selected from a bulk, depending on the quantity the number of selecting points could be decided.

b. Preparation

In the preparation of samples all the instructions specified should be carefully followed because invariably only a small quantity will be tested and this should truly represent the lot.

i. In Fertilizer, from a consignment of 15,000 MT the amount ultimately tested for nutrient content is as little as 5g.

ii. In some products due to damage in handling the outer layers are discarded e.g.

1. outer yarn in a textile cone
2. outer layers in a paper ream

iii. In certain products the test pieces are cut from specified areas e.g.

1. in leather skin and hides, certain areas are specified for certain tests
2. In technically specified rubber (TSR), pieces are cut from diagnolly opposite sides in a block.

iv. In certain rubber products it is not possible to obtain from the product itself, the required test pieces. In such instances a test piece is made from the same raw material, vulcanized under same conditions of pressure, temperature and time. Here one draw back may be that in the laboratory when the test piece is made it may have ideal conditions, while in commercial production it may not be so.

v. Products where a composite sample is made from primary samples, it should be ensured that proper coning and quartering is done to ensure that a representative sample is obtained.

Decision Criteria

The decision should be made on the basis of the sample tested. The decision on the acceptability of the lot is based on the sampling plan used. However it should be borne in mind that the figure obtained by the test is the sum total of the Mean of the lot + sampling error + reduction error + analytical error. These together with an idea of processing of the product helps in the final decision.

Conclusion

To obtain a representative sample a great deal is dependent on the Inspector, his honesty and his ability to obtain a representative sample.

In lot acceptance, it is worthwhile to develop a rapport with the supplier and to ensure that he has a sound quality control system thus ensuring that products are manufactured to acceptable standard as this to a great extent diminishes the need for inspection and thus eliminating the problems encountered in selecting a representative sample.

NA-59-5

SAMPLING IN BIOLOGY AND MEDICINE

S. Mendis

Faculty of Medicine, University of Peradeniya

In its broadest sense a sample refers to any specific collection of observations drawn from a parent population. We may take a sample of patients, a sample of blood pressure readings and so on. At one extreme the sample may include all the units in the parent population. At the other extreme a sample may consist of only one unit selected from the population. In a narrower sense a sample is a collection of observations which have been selected from the population in a particular way.

I. Objectives of Sampling

There are two main objectives of sampling. They are

a. estimation
b. testing of hypotheses

In general the main objective in selecting a sample is to obtain a representative selection of observations from the parent population. A study of such a sample allows inferences to be made about the population from which it is drawn. Information on disease rates and distribution of disease in a sample is an indirect estimation of the amount of disease and distribution of disease within a population. A study of a 'healthy' sample of people drawn from the population may provide basic descriptions of human biology or may allow 'Normal' values to be estimated.

Formulation and testing of hypotheses is one of the most important and at the same time the most complex aims of sampling. For example the observations of increased rates of a disease in certain groups of people in a study sample may suggest personal or environmental factors of possible etiological importance. This etiological hypothesis may then be tested by further studies.

Sampling is done in such a way so as to investigate whether the frequency of disease increases in proportion to the level of exposure to the factor in question.

2. Methods of sampling

In clinical studies it is often satisfactory to gather a series of cases without regard to the source from which they were obtained; the need is simply to collect cases as quickly as possible. However, if on the basis of examination of a sample, inferences are to be made about the population from which the sample is drawn a representative sample will have to be selected from the parent population. Various methods of sampling are available but only the principles and a few techniques will be described here.

2.1 Preparing a nominal roll

The first requirement is a nominal roll or sampling frame enumerating each individual or each unit of the population from which the sample is drawn. Sometimes existing official lists may be used. For eg., population census lists, electoral registers, school registers and list of employees could provide a satisfactory sampling frame. Where no satisfactory census exists, it may be necessary to take one specially for the study. This is generally a costly, time consuming and a technically difficult undertaking and is likely to be possible only in a very restricted area or on a cluster-sample basis.

2.2 Sampling techniques

The sampling techniques that are commonly used in biology and medicine are

2.2.1 simple random sampling
2.2.2 stratified and weighted sampling
2.2.3 systematic sampling
2.2.4 cluster sampling
2.2.5 two stage sampling
2.2.6 control sampling

2.2.1 Simple Random Samples

A sample may be selected in such a way that each unit or observation in the population has an equal chance of selection. This type of sample is known as a simple random sample. This is the principle used in the selection of winning tickets in a lottery. It is unwise to select the first individuals who meet the necessary criteria. For example the order in which names appear on the census may have some relation to the risk of disease even if this seems improbable.

Two sample techniques for drawing a random sample is available. In one technique each person on the census is assigned a number. Thus using a table of random numbers, individual numbers are picked until a sample of the required size has been obtained. In the other technique if n = the required sample size, and N = the number of persons in the enumerated population, then the selection of every (n/N) th person on the list will provide an effectively random sample, provided that the starting point is randomly selected. The first technique is preferred because of possible effects of order on the census list as mentioned above.

Tables of random numbers are tables of the digits 0,1,2,...9, each digit having an equal chance of selection at any draw. Among the larger tables are those published by the Rand Corporation (1955) -1 million digits - and by Kendall and Smith (1938) - 100,000 digits. Numerous tables are available in standard statistical texts.

2.2.2 Stratified and weighted samples

A refinement of the simple random sample is the stratified random sample. If the population is unevenly distributed with regard to some relevant known variable such as age or sex, it may be useful to draw such a sample. The population is first divided into classes according to a known variable and a random sample selected from each class. This sampling ratio may then be adjusted so as to provide samples of

appropriate size from within each class.

The important feature of stratified random sampling is that the sample proportions are predetermined. The sampling error in a stratified random sample is usually less than the sampling error in a simple random sample of the same size.

2.2.3 Systematic sampling

If the N units in the population are numbered 1-N in some order, to select a sample of n units, we take a unit at random from the first k units and every K^{th} unit thereafter. For instance if K is 15 and if the first unit drawn is number 13 the subsequent units are numbers 28, 43, 58 and so on. This type is called an every k^{th} systematic sample.

There are several advantages of this method over simple random sampling. It is easier to draw a systematic sample and often easier to execute without mistakes specially in field studies.

The systematic sample is spread more evenly over the population than a stratified sample and therefore is considerably more precise than stratifed random sampling.

2.2.4 Cluster sampling

If the sampling units consists of a group of smaller units, sampling is known as cluster sampling. There are two main reasons for the widespread application of cluster campling. In many field surveys it is found that no reliable lists of the sub-units of population are available and that it would be probably expensive to construct such lists. From maps of the region however the area can be divided into units such as blocks in the cities.

Cluster sampling introduces some additional sampling error and the method of calculating sampling errors is different (see Snedecor and Cochran 1967).

In countries without any adequate national census the best system is generally one of sub-sampling in clusters. A complete list of administrative areas is first obtained. A random (or stratified) sample of clusters is drawn; within each selected area a list of residents must then be obtained. Examination is performed on a random sample or on all persons in the selected clusters. Economic considerations may point to the choice of a larger cluster unit. For example greater field costs are incurred in locating 600 house holds randomly selected in a town than in locating 20 blocks and visiting all households in these blocks.

2.2.5 Two-stage sampling

For some special purposes it may be useful to examine the whole population or a sample of it in a simple way and then to use this initial data as a basis for drawing appropriately weighted subsamples for more detailed study. For example if the objective of a study is to measure disease rates in smokers and non-smokers in a popultion in which the proportion of non-smokers is low, a very large random sample would have to be examined in order to provide a sufficient number of non-smokers. It might therefore be decided to send out a smoking questionnaire to the whole enumerated population and thus to call for examination randomly chosen samples of equal numbers of smokers and non-smokers.

Whenever any process involves chemical, physical or biological tests that can be performed on a small amount of material it is drawn as a sub-sample from a large amount that is itself a sample. This technique is called sub-sampling or two-stage sampling because the sample is taken in two steps. The first is to select a sample of units often called the primary units and the second is to select a sample of second stage units or sub-units from each chosen primary unit.

2.2.6. Control samples

In medical research studies the use of what are called control samples is a common and important technique. If risk factors of coronary heart disease such as smoking, hypertension and hypercholesterolaemia are studied in a group of patients with coronary heart disease and if one wishes to determine whether or not any of these risk factors are significant in relation to coronary heart disease, it will be necessary to compare these factors with another group who do not suffer from coronary heart disease. The group who do not suffer from coronary heart disease but who in other respects may be assumed to be comparable with the group of patients is known as the control group. A commonly used type of control group is the 'matched sample'. The control group and the group of patients are identical with respect to certain characteristics such as age, sex, occupation and so on. When a patient enters the study he is matched with a control individual with the same characteristics.

In principle the control group is randomly selected from the population and is taken to be representative of the population. In practice the controls may have to be used which are not in fact randomly selected from the population.

3. Size of Sample

3.1 Fixed size and sequential samples

What was described upto now refers to "fixed size" sampling procedures. Determination of sample size is one of the most important matters to be decided in designing an experiment. The experimentor can decide how large the sample size should be by using certain well known techniques. A sample of a predetermined minimum size is selected and the experiment is carried on until all the sample results become available.

In medical research it is sometimes inconvenient and undesirable on ethical grounds to conduct an experiment using a fixed-size sample. Another type of sampling procedure which is very useful in medical research is known as sequential sampling. In a sequential sampling scheme sample items are collected and analysis of the data continues throughout the period of the experiment instead of being dependant on the completion of a 'fixed size' sample. The experiment is terminated as soon as sufficient results are available to accept or reject the hypothesis of the experiment.

Suppose a new drug is being tested for treatment of a particular disease. The new drug is given to a sample of patients and for comparative purposes suitably matched control group is treated with a drug already being used for treatment of the disease or a placebo. The trial is terminated as soon as sufficient results have been obtained to make a decision on the efficiency of the new drug. The statistical techniques involved in sequential sampling are different from those used for 'fixed size' sample and will not be discussed here.

3.2 Statistical consideration of sample size

In the case of quantitative variables any random measurement error naturally affects the accuracy of results and may necessitate an increase in the sample size. The choice of sample size is also affected by many practical considerations besides the statistical estimates. Too large a study is time consuming and expensive and the overstretching of resources may impair the quality of results. However before commencing a study the best possible calculations of sample size should be made in order to arrive at reasonable answers to study questions.

3.2.1 Prevalence and incidence studies

If the aim of the study is to estimate the prevalence of a condition, then in order to determine an appropriate sample size, the investigator needs to know of

(a) the approximate order of frequency of the condition

(b) the degree of precision required

Table 1 gives exact 95% confidence limits for various rates and sizes of random samples based on the effects of sampling variability. the error is inversely proportional to the square root of sample size.

An appropriate formula for the calculation is given as follows :

$$S.E. = \sqrt{pq/n} \quad \text{when} \quad S.E. = \text{Standard error}$$

p = proportion affected, q = (1-p) and n=number in sample

In comparison between two studies the significance of the differences between estimates is given by the formula.

$$\frac{P_1 - P_2}{S.E.\ \text{difference}}$$

When P_1 = proportion affected in first sample,

P_2 = proportion affected in second sample. The S.E. difference is given by the expression

$$S.E.\ \text{difference} = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

In order to calculate the sample sizes (n_1, n_2) a second type of error must be considered; namely, that of failing to demonstrate a difference that really exists. Some of the required sample sizes which gives an 85% chance of recognizing a real difference between two populations is given in Table 2.

3.2.2. Studies of quantitative variables

The error in estimating the mean value of quantitative variables depend on the sample size as well as the amount of variation of individual values between subjects.

$$\text{Standard error} = \sqrt{S.D^2./n}$$

When S.D. = Standard deviation of individual values for each subject, n = number of subjects.

If this variability is known approximately from a pilot study then the sample size that will yield a specified confidence range can be estimated.

The variance of individual values between subjects has three components ;

(a) true between - subject variation
(b) biological variation within subjects
(c) measurement error

The effects of the last two may be decreased by proper technique and by replicate measurements. This will reduce the required sample size.

In comparison between two studies the significance of a difference observed between two estimates of mean values is given by

$$\frac{X_1 - X_2}{\text{S.E. difference}}$$

where X_1 and X_2 are the two mean values, and

$$\text{S.E. difference} = \sqrt{\text{S.E.}_1^2 + \text{S.E.}_2^2}$$

In planning a comparative study, if the following requirements are known, the necessary sample size can be calculated. These are

(a) the magnitude of difference that is desired to detect

(b) the significance level at which any difference can be shown

(c) the risk the investigator is prepared to take of failing to demonstrate a difference that exists

(d) estimation of the variability of individual values

Table 3 illustrates the sample sizes that will be needed to meet various specifications which should serve only as a rough guide.

Detailed discussion of the statistical principles on sample size determination may be found in a standard statistical references (e.g. Moser & Kalton 1971, Schwartz et al 1980, Armitage 1971).

3.2.3. Response Rate

It is important to obtain a good response rate so that serious bias will not occur even if the non respondents are unrepresentative. In practice this cannot always be achieved, and one must then try to assess the bias resulting from the omission of non-respondents on the basis of such information as is available for the whole sample eg. age, sex, residence.

With a high prevalence condition a poor response rate is less likely to be serious. Suppose a particular symptom such as headache is present in 50% of an elderly population. If 20% of the sample failed to attend for examination and 80% of these persons had headache the estimated prevalence among the respondents would only be reduced 42%, a relatively small error.

TABLE 1

95% confidence limits for various rates and samples sizes

Estimated percentage	95% confidence limits for samples of the following sizes	
	500	1000
2	1.0 - 3.7	1.2 - 3.1
4	2.5 - 6.1	2.9 - 5.4
6	4.1 - 8.5	4.6 - 7.7
8	5.8 - 10.7	6.4 - 9.9

T A B L E 2

Required sample sizes for giving an 85% chance of recognizing a specified difference in rates between two populations

Estimated true rates (%)		Required sample sizes
P_1	P_2	$n_2 = n_1$
5	2	670
15	5	160
20	5	80

T A B L E 3

Required sample sizes for giving an 85% chance of recognizing a specified difference in mean values between two populations

Standard deviation of individual values	Required sample size for demonstrating the following differences in mean values			
	1	5	25	50
1	20	10	10	10
5	450	20	10	10
25	1100	450	110	10

S.E. difference $\sqrt{S.E_1^2 + S.E_2^2}$

REFERENCES

Snedecor, E.W. and Cohran, W.G. (1967)
Statistical Methods, 6th ed., Iowa City, Iowa State
University Press

Random Corporation (1955)
A Million Random Digits
Free Pres Glencoe III

Kendall, M.G. and Smith, B.B. (1938)
Randomness and random sampling numbers
Journal Roy. Stat. Soc. 101, 147-166

Armitage, P. (1971)
Statistical Methods in medical research
New York, Wiley

Moser, C.A. & Kalton, G. (1971)
Survey methods in social investigation, London,
Heinmann

Schwartz, D. et al (1980). Clinical trials, London
Academic Press.

NA-59-6

SAMPLING AIR FOR PARTICULATES

M. Ponnambalam,

Division of Occupational Hygiene, Dept. of Labour.

The nature and effects of polluted air are complex and one is not surprised, therefore, that errors of fact and in interpretation of facts are commonly to be found in current accounts. There is a real danger that the propagation of some errors, especially those which lead to overstatement of the dangers and to prophecies of doom, may eventually weaken the obviously strong case for the prevention and abatement of pollution. (Lawther, 1970). The particulates of significance to hygienists include all particles, solid or liquid, which are suspended in air and may be inhaled. The particles may be of all sizes from molecular dimensions upto about 100 micrometers in diameter. Sampling is performed by drawing a measured volume of air through, a filter, cyclone or other instruments for collection of particulates. The size of the particles is determined by separating out the different sizes of the collected particulates in the laboratory, using a microscope or liquid settling. When a particle is released from rest and falls in air, the particle falls with a steady velocity known as its terminal velocity. Many particulates are not uniform spheres, in fact, they may be irregular aggregates. The preferred 'diameter' or size of a particle is it's 'equivalent' or 'aerodynamic' diameter. This is equal to the diameter of spherical particles of unit density which have the same falling velocity in air as the particle in question. For some types of particles with extreme shape, other parameters are sometimes used. Thus, asbestos fibres, which are very long in relation to their diameter, are characterized by their length.

Amongst particles which are inhaled, those with an equivalent diameter greater than 20 micrometers are deposited by impingement in the nose and upper respiratory tract. Smaller ones down to 0.5 micrometers in diameter are carried into the smaller airways and alveoli, and

are deposited there under gravity. Particulates with an aerodynamic diameter of 7 μm and above will not normally reach the alveoli, but particulates below that size will and are therefore termed 'respirable'. Most particulates are harmless but in sufficient concentrations they can cause discomfort and unpleasantness. At such levels they are termed nuisance particulates. Particulates, most dangerous for the lungs are normally between 0.2 and 7 micrometers in size. These harmful particulates may cause chronic lung diseases, asthma, cancer etc. Particulates from graphite, sand and most hard rocks are harmful at respirable size but are normally cleared from the lung if larger; whereas pollens, spores and mists are larger and can cause problems in the upper respiratory passages. Fumes from molten metal are very small, below 1 micrometer, and can give rise to metal fume fever and more serious disorders. When determining the concentration of airborne dust it is important to understand what size of dust is to be measured as this influences the method of sampling. Site of deposition of particulates in the respiratory tract is presented in figure 1.

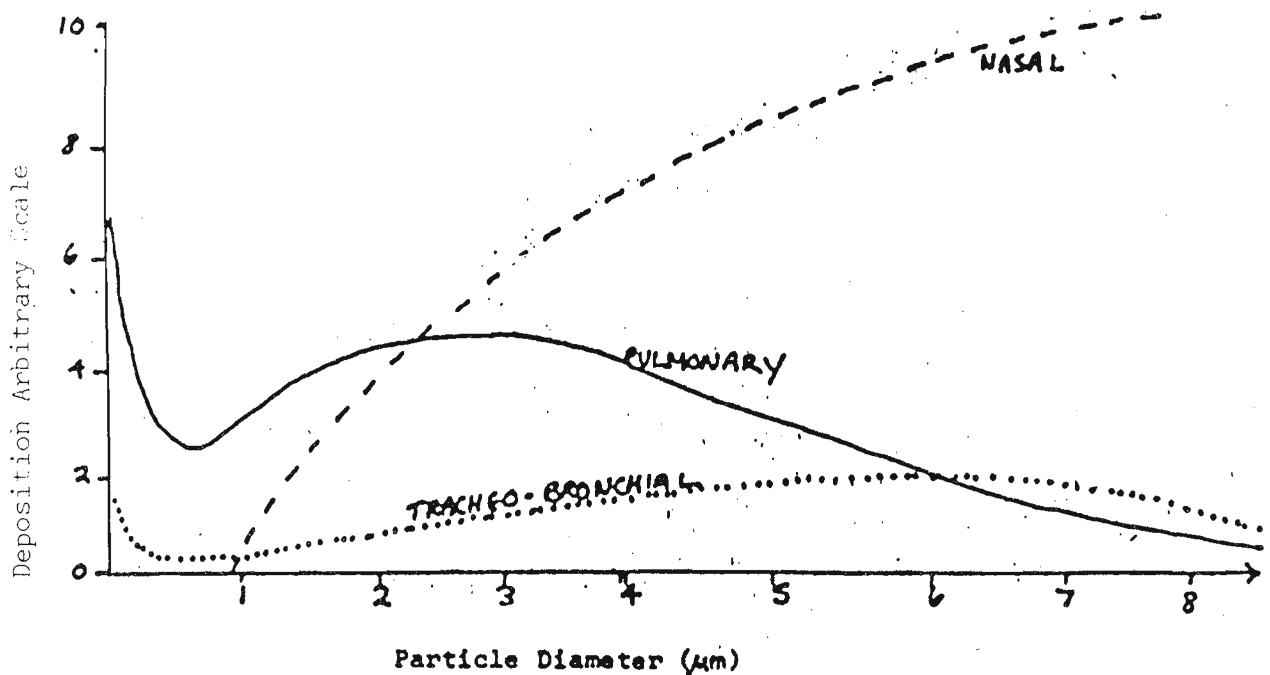


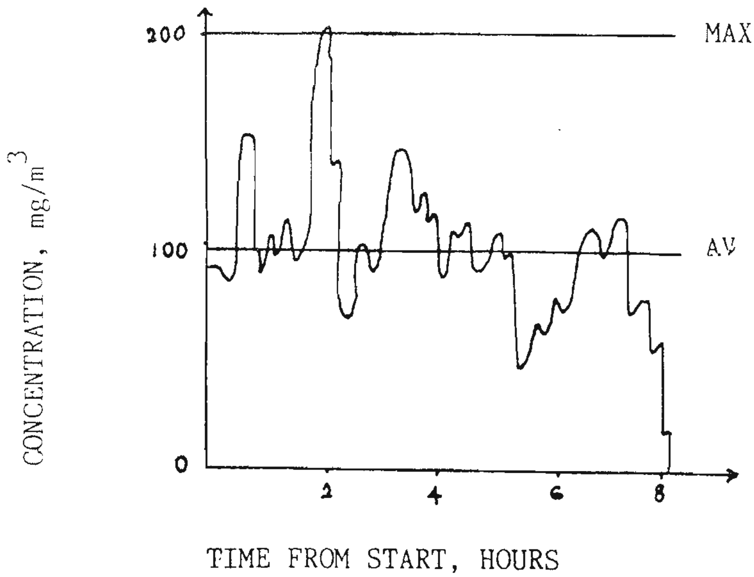
Fig. 1: SITE OF DEPOSITION OF AEROSOLS IN RESPIRATORY TRACT

The main reason for sampling for atmospheric particulates in the working environment is to estimate the concentration in the air which is inhaled by the employees. A determination may be made of the concentration of all the particulates or just those which have particular sizes or shapes. This is done in order to assess whether there is a risk to the health of workers exposed to the environment. This judgement is made by comparing the results against hygiene standards. The result obtained by atmospheric sampling depends very much on the time and place where the samples are taken and the type of instrument used. Those concerned with setting hygiene standards usually take account of the wide variation in the results that may be obtained.

Concentrations of particulates may vary around the average value from zero up to 2½ times the average, even in conditions where the work appears to be done at a steady rate. An illustration of the variation in concentration of contamination during a work-shift, is given in figure 2. The airborne particulate concentrations in the work environment vary from place to place. They not only vary from place to place but vary from time to time in the work place. An example of the variation in dust concentration experienced by a coal miner during a working shift is shown in the figure 3. If the measurements are repeated on different days, the daily average will also be found to vary; an example is shown in figure 4. The frequency distribution of dust concentration as measured in mines is shown in figure 5. Daily lead in air, concentrations in a lead smelting works is given in table 1.

Table 1 : Daily Lead in Air Concentration

Test Day	Sampling Duration Minutes	Pb in Air mg/m^3
1	480	0.74
2	480	0.68
3	480	0.56
4	480	1.22
5	480	0.60
6.	480	0.74
7	480	0.47



TIME FROM START, HOURS

Fig. 2; VARIATION IN CONCENTRATION CONTAMINATION DURING A WORK -SHIFT

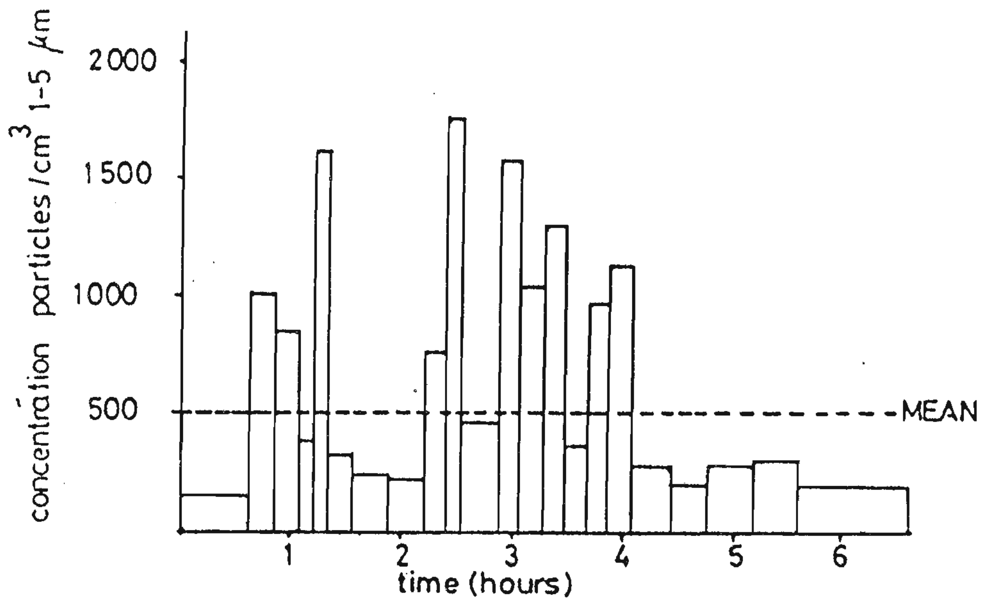


Fig. 3: VARIABILITY OF DUST CONCENTRATION DURING A WORKING SHIFT (WALTON, 1967)

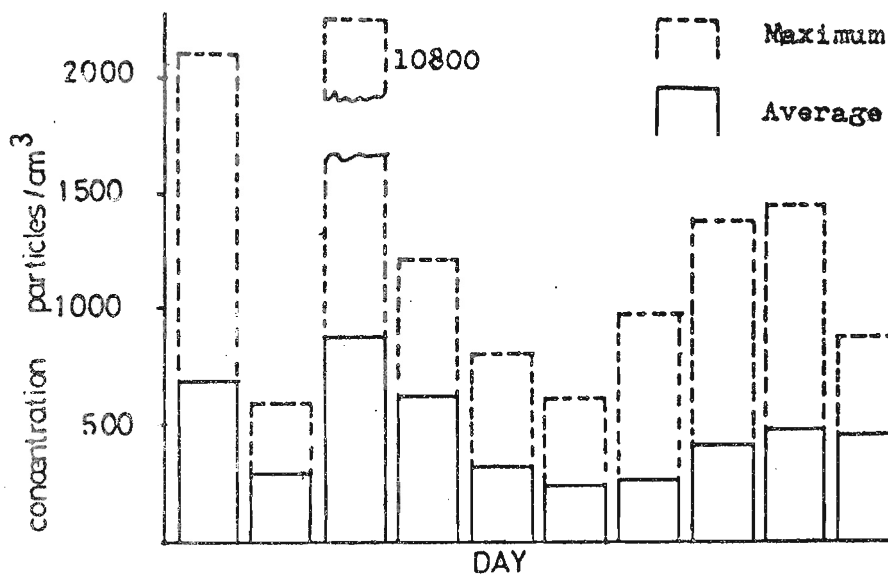


Fig. 4 : VARIABILITY OF DUST CONCENTRATION AT SAME WORKING SHIFT (WALTON, 1967)

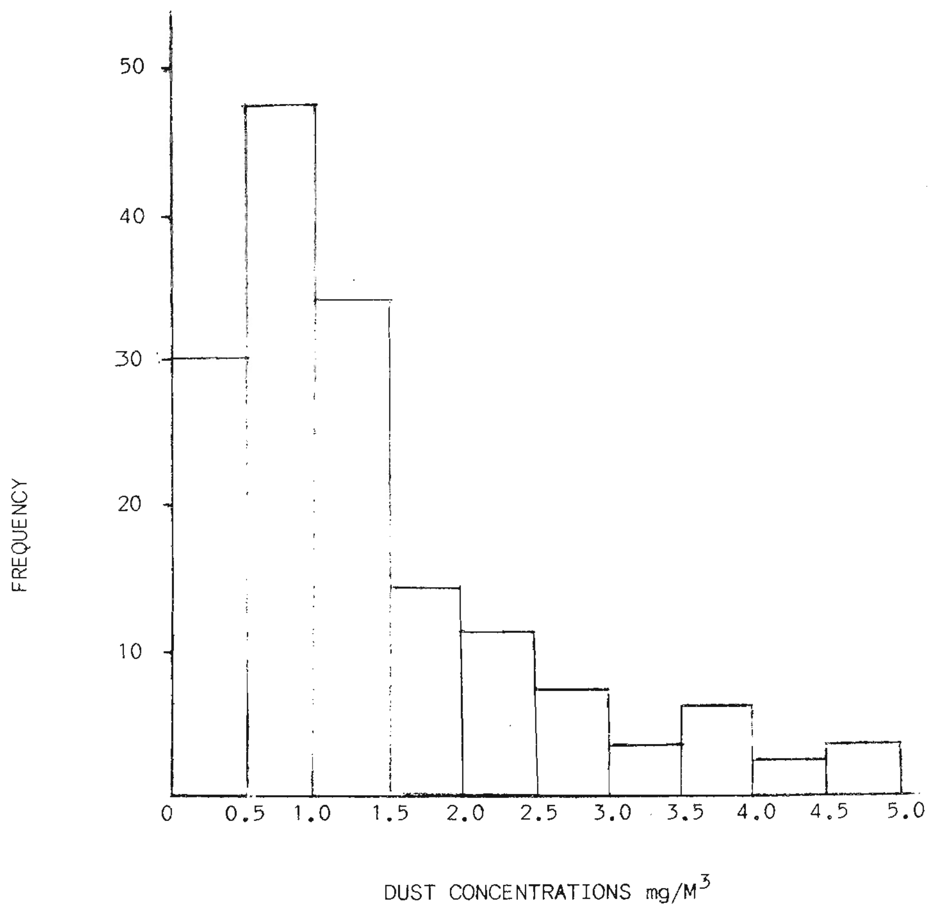


Fig. 5 : **DUST CONCENTRATION IN MINE AIR**

Thus, it is found in real situations the distribution of concentration of particulates in the environment, the size of the particulates and their shapes are all variable. They are dispersed in the air which is a moving system and hence dynamic. Further, no two different types of airborne particulate samplers yield the same result. It takes the most painstaking tests to yield similar results from two identical instruments.

Roach (1966) has developed a reasoned and consistant system of sampling based on the assurence that a critical body burden is not exceeded. It is impossible to prove from a set of sampling results that a particular concentration was never exceeded. It is possible only to show that the likelihood of its having been exceeded is remote. This means that to come to a favourable decision from a limited number of samples, a degree of risk has to be accepted. This may be made as small as one wishes, but judgement must be excercised on what is an acceptable risk.

It has been shown that, provided the duration of the individual samples is no more than one-tenth the half-time, the coefficient of variation of a man's body burden will be no more than one-fifth the coefficient of variation of the air concentration. Consequently, by maintaining the average air concentration one standard deviation below the threshold eight-hour steady air concentration, the average body burden is then maintained five standard deviations below the threshold body burden. No matter what the nature of the size-frequency distribution of the air concentration, the fluctuations of the body burden from time to time tend to have a normal distribution, as the time scale of the fluctuations becomes small in comparison with the half-time. This means that the risk of exceeding the threshold body burden at any time would be less than one in a million.

The number of samples taken during a given test period is usually small in statistical terms, and most hygienists are satisfied with the simplest of statistical tests. The simplest measure of dispersion is the range, that is, the difference between the highest and lowest value in the set of results. A safe approximation to the standard deviation from small samples is given by:

$$\text{Range}/(n-1)^{\frac{1}{2}}$$

Where n is the number of samples.

Also, random analytical errors have the effect of increasing the variability of the results, whose standard deviation is therefore higher than the standard deviation of the concentration.

A safe procedure that emerges from this is as follows. First, a lower limit for the biological half-time must be estimated. Given a lower limit to the half-time, the necessary duration of each sample is one-tenth of this value. Finally, it remains to establish with such samples that the average concentration is at least one standard deviation lower than the threshold limit value. To do this, the following expression has been worked out by Roach (1966).

$$\text{Average} + \text{Range}/(n-1)^{\frac{1}{2}}$$

When this is less than the threshold limit value, a favourable report regarding the particulates in the environment may be given.

The most common method of sampling Airborne particulates is to draw a known volume of Air through a filter medium by means of a pump. A steady flow-rate sampling pump capable of supplying air at the flow rate of the rotameter is employed. Soap bubble calibraters are used for caliberation.

Filters are mainly 'fibrous' in structure, made from glass, paper, polystyrene or from a 'membrane' of cellulose derivatives, PVC and polycarbonate. There is also a sintered silver filter available. The correct filter must be chosen to suit the airborne contaminant to be sampled and the subsequent analysis to be undertaken : for example, some can be dissolved in chemicals for further analysis of the collected dust, some can be made transparent for optical microscopic examination of the material, whilst others allow the collected dust to remain on the surface for scanning electron microscopic examination. Some filters are more sensitive to atmospheric moisture content than others and need to be pre-conditioned before weighing.

Table 2. Details of filters to be used for different particulates.

Type of particulates	Method of analysis	Filter required	Filter holder required
Asbestos fibres	Optical microscopy	Cellulose ester*	Open face
	Scanning electron microscopy	Nucleopore	Open face
Man-made fibres	X-ray diffraction	Silver membrance	Open face
	Optical microscopy	Cellulose ester*	Open face
	Gravimetric	Glass fibre	modified UKAEA
Silica	X-ray diffraction	Silver membrance	Cyclone
	Infra-red	Polyvinyl-chloride	Cyclone
Lead, heavy metals, their oxides and salts	Atomic absorption spectroscopy	Cellulose ester or Glass fibre	UKAEA with 4 mm hole
Nuisance and general	Gravimetric	Glass fibre	Open face
	Optical microscopy	Cellulose ester	Open face
Unknown particles	X-ray diffraction	Silver membrance	Open face
Graphite	Gravimetric	Glass fibre	Cyclone or MRE 113
Graphite with rock	Infra-red	Polyvinyl chloride	Cyclone or MRE 113
	X-ray diffraction	Silver membrance	Cyclone or MRE 113
Oil mists	Gravimetric	Glass fibre	Open face
	Fluorescent spectroscopy	Cellulose ester*	Open face
Welding fume	Gravimetric	Cellulose ester*	Open face
	Atomic absorption spectroscopy	Cellulose ester*	Open face

* Pore size should be 0.8 μm

Pore sizes of the filters vary between 0.1 micrometers to 10 micrometers, but it should be noted that the pore size does not limit the size of the dust to be collected, that is a 5 micrometer pore size filter is capable of capturing dusts smaller than 5 micrometers by virtue of the inertial and electrostatic forces that occur within the filter medium. In fact, it is often desirable to use pore sizes in the range 5 - 10 micrometers even for respirable dusts. When sampling for specific particulate it is important to select the correct filter depending on the method of analysis to be used later.

Filters recommended for different particulates for specific methods of analysis are given in table 2,(Gill & Ashton, 1982)

References

1. Gill, Frank S. and Ashton, Indira, (1982). Monitoring for health hazards at work; Grant Mcintyre Ltd. London WCIB 3 PY, UK.

2. Lawther, P.L., (1970); Proc. Roy. Instn. Gt. Br; 44, No. 203.

3. Roach S.A., (1966), Am. Ind. Hyg. Assoc.J. 27 1-12

4. Walton, W.H. (1967), W.H.O. Pneumoconiosis: report on a symposium, Copenhagen.

NA -59-7

SAMPLING OF WATER FOR ANALYSIS OF CATIONS AND ANIONS

H.D. Gunawardhana

Dept. of Chemistry, Univ. of Colombo

Water can exert an important influence on most chemical reactions in the earth's crust. It is like a liquid extractor plant in which crushed rock debris is continuously leached by water. In this process cations and anions are leached into water. The ions are hydrated and kept apart by water molecules such that the electrical neutrality is maintained. Natural weathering process adds cations and anions to natural water systems. This addition is enhanced by man's activities such as farming, deforestration etc.

Water has always been a vital natural resource for man's existence. It can be said that fresh water is the most important natural resource. It's use for drinking, cooking, agriculture, transport industry etc. immediately shows the extent to which it is an integral part of our life. No matter what the purpose for which water is required, it has long been recognised that its suitability for that purpose can be affected by dissolved anions and cations as well as suspended matter (precipitated anions and cations) in water. When the quality of water is affected in this manner, the water is said to be "polluted". The water pollution can be defined as follows: "Through man's exploitation of his water resources, the dynamic balance in aquatic ecosystems is frequently disturbed which may result in such dramatic responses as foaming, taste, odour, etc. These ecological responses are described in terms of "pollution" since they interfere with man's use of the water resources".

Once the water resources are affected, it is necessary to analyse them, to check the suitability for drinking, cooking, industry, agriculture etc. Hundred years ago, the water analysis was confined to checking the suitability for human consumption. The last two decades or so have seen a large and steady increase in the importance attaching to many aspects of water quality for all of the above uses to which it

is put. Today, the number of parameters of concern is continually increasing and smaller and smaller concentrations are becoming of interest. The number of laboratories making routine analysis of water is continually increasing as is the number of analysts concerned with water. In addition, a much wider range of analytical techniques and more sophisticated instrumentation and methods than hitherto are being employed to meet requirements of growing difficulty.

In order to assess the suitability of the water body for any purposes mentioned, it is necessary to analyse a sample from the water body. Prior to sampling, it is necessary to locate the sampling site(s).

The analytical chemist should ensure that the locations at which samples are taken are representative of the water of interest.

Time of sampling	:	may be related to discharge of industrial effluent into the water body concerned.
Frequency of Sampling	:	The ideal approach might often be to obtain a continuous record of the concentrations of anions and cations. This is, at present, generally impracticable. A large sampling frequency is generally advisable.

Basic aims of sampling may be summarized as follows :

(a) to obtain a sample whose concentrations of determinands (parameters i.e. anions and cations in this case) are identical to those in the water of interest at the time of sampling.
i.e. a <u>representative</u> sample

(b) to ensure that the concentrations of the determinands do not change between sampling and analysis.

How do we obtain a representative sample ?

Sampling Location

If we want to measure the efficiency of a chemical plant for purifying water, sampling locations will be required before and after the plant. Similarly, when the effect of an effluent discharge on the water-quality

of the receiving river is of interest, sample will be required from locations upstream and downstream, of the discharge. More complex programmes involve the quality surveillance in an entire river basin or within the potable-water distribution system of a large area.

Sampling Positions

Even when the general locations have been chosen, the exact positions from which to collect samples will often need to be chosen carefully if representative samples are to be obtained. Sampling positions seem often to be chosen as the basis of accessibility and convenience than their suitability as sources of representative samples. This may sometimes be unavoidable but it is always worth seeking an alternative position if the one that is accessible, is likely to provide information of uncertain value. Sampling from the boundaries of water systems should generally be avoided except when those regions are of direct interest. Undesirable sampling positions may be cited as the banks, surface or bottoms of rivers etc. Sometimes it may be necessary to carry out an experimental investigation on the proposed sampling positions. On the first day, several samples from several sampling positions could be analysed to find out the convenient representative sampling position. A sample of water taken from a pump or tap initially will differ from one taken after the water has been allowed to flow a set amount of time.

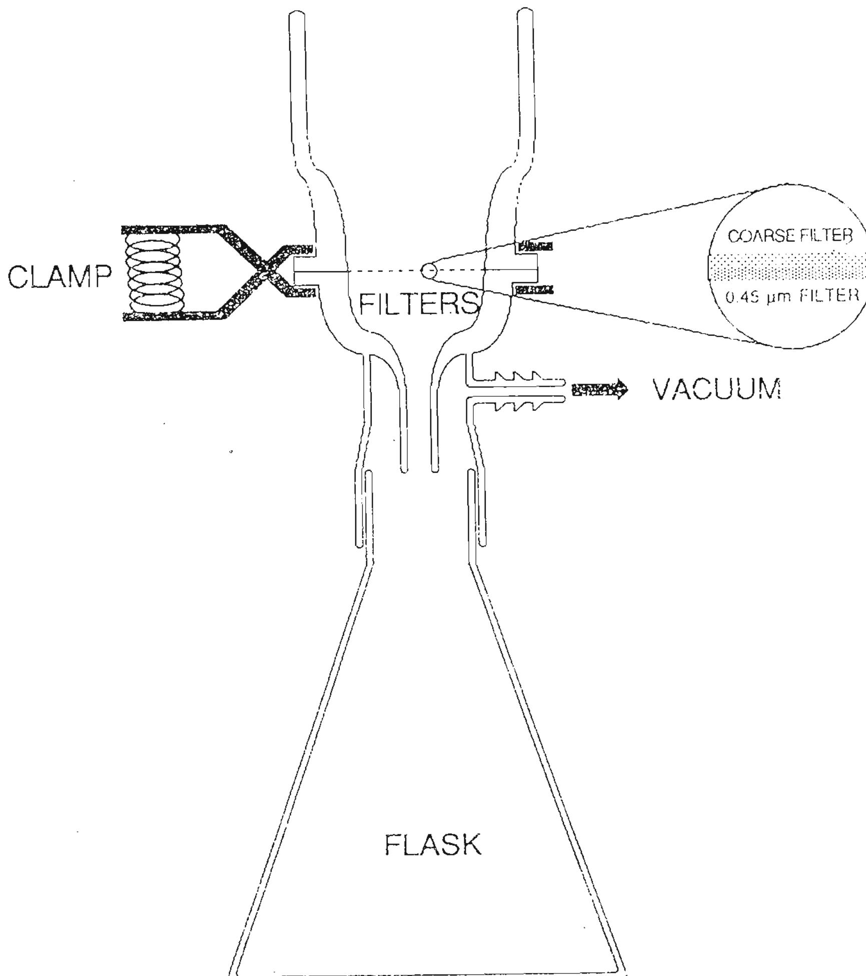
Problems in selecting suitable sampling positions arise whenever the determinands are not homogeneously distributed throughout the water of interest. Anions and cations may occur as solids leading to heterogeneity. eg. $Ca_3 (PO_4)_2$. These inorganic sediments in rivers and other natural waters tend to increase in concentrations with increasing depth because the particles tend to settle. Sometimes these insoluble materials are homogeneously distributed. eg. Pipe, Plant with sufficient turbulance.

If the water body is heterogeneous, more than one sampling position should be chosen from each sampling location. Having decided the sampling positions, the extent of suspended matter should be checked experimentally. These samples should be taken from the same depth. Turbidimetry may be used to check the extent of suspended matter. Depending on the turbidimetric measurements the number of sampling

positions can be reduced. Next, at each position samples should be taken at different depths. These also should be subject to turbidimetric measurements.

The sample should be filtered through 0.45 µm filter immediately after collection. A filtering device in Fig. 1 may be used for this purpose. The vaccum in the field can be achieved using a bicycle pump. The filtration achieves the separation of the suspended matter from the true solution.

Fig 1 - WATER FILTRATION



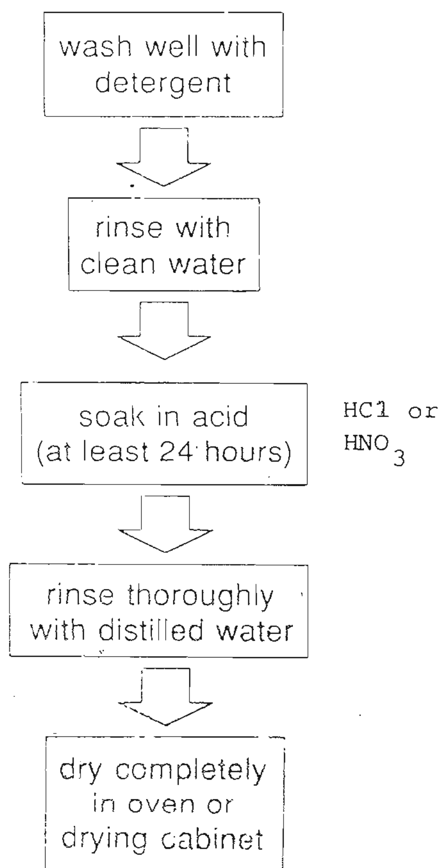
Inhomogeneity also arises due to incompletely mixed waters, eg.

i. the thermal stratification of lakes and reservoirs;
ii. the salt and fresh water estuaries;
iii. the different composition of the water at different points within stream/water drum boilers;
iv. the discharge of effluents into rivers or streams

In all the above cases, several sampling positions are necessary; the number depending on the size of the water system and the degree of heterogeneity. The degree of heterogeneity can vary with time. Frequent sampling may be required initially to study this effect.

To reduce the amount of work involved in analysing samples from a number of positions, it is sometimes useful to combine the samples before analysis. When this is done, the proportions in which the individual samples are mixed must be carefully chosen so that the composite sample is adequately representative.

Cleaning procedure for containers :



HCl or HNO_3

Method of sampling

The cleaned container is now used for sampling. For many applications in which natural waters are of interest, no specific sampling method is required. It is often sufficient to immerse the container in the water of interest. The mouth of the container should face away from the direction of the water flow. This procedure is quite adequate if the determinands are the dissolved anions and cations. The container need not be dry. It is recommended to rinse the container two or three times with the sample before finally filling the container. When it is required to sample from depths, special containers are available for lowering into the water and obtaining a sealed sample collected from a chosen depth.

When undissolved cations and anions (eg. Calcium phosphate) are of interest, a different method of sampling should be used. Then an adequately clean, dry (free from water) container can be used without rinsing. The sample is collected directly into the container, and the mouth of the container should face in to the water flow. Ideally, the velocity of the sample stream should be equal to the velocity of the water being sampled. This is the iso-kinetic sampling. (Fig. 2)

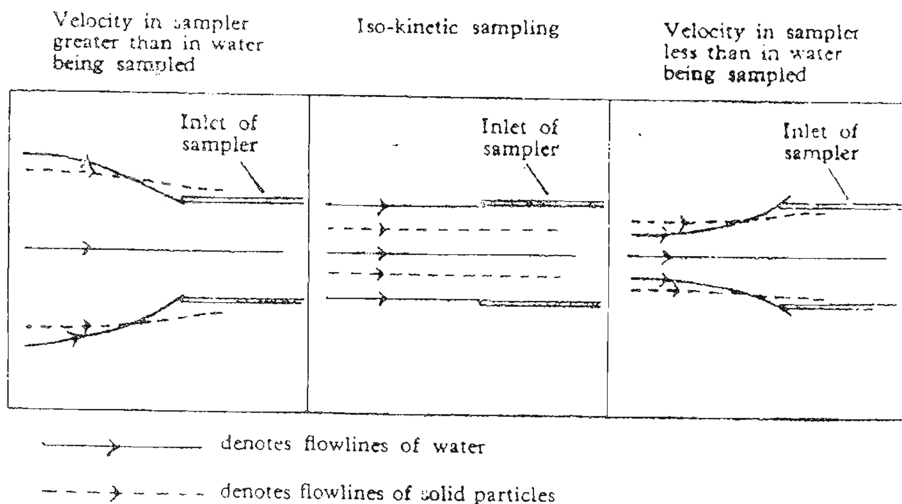


FIG. 2 *ISO-KINETIC SAMPLING*

If the iso-kinetic sampling is not achieved the sample contains either more undissolved matter (or lesser undissolved matter) than the water in the stream. Immediate filtration (Fig 1) is recommended after sampling.

Sample stability

To ensure the concentrations of determinand do not change between sampling and analysis, it is sometimes necessary to preserve the samples. Depending on the nature of determinands, the sample stability can sometimes be achieved by the addition of preservatives.

Main factors affecting the stability of the samples are :

i. the nature of the samples
ii. the sample container
iii. the addition of preserving reagents to the sample

(i) Depending on the nature of the sample Chemical and Biological reactions occur between sampling and analysis.

Chemical Reactions

Precipitation of calcium carbonate. This leads to changes in pH, alkalinity and hardness. Precipitation of iron as $Fe(OH)_3$ on exposure to air.

Biological Activity

Bacteria and algae can consume partially or completely, a number of substances required for their growth. eg. nitrogen, phosphorus and silicon compounds.

The biological activity is prevented or reduced by storing the sample at 4°C. We can preserve many anions including nitrate, phosphate, & silicate by reducing the temperature to 4°C immediately after sampling.

Algal and bacterial activity can also be reduced sufficiently by filtering the sample during or immediately after collection. Filtering sample through 0.45μm glass fibre filter and storing the filtrate at 4°C will prevent or mitigate the biological activity.

(ii) Sample containers may have important effect on sample stability. There are recommendations on the type of container suitable for each determinand. Many publications on this subject do not always agree. Polyethylene (plastic, polypropylene) or glass

(preferably boro-silicate) containers are widely used. Glass containers can be more efficiently cleaned.

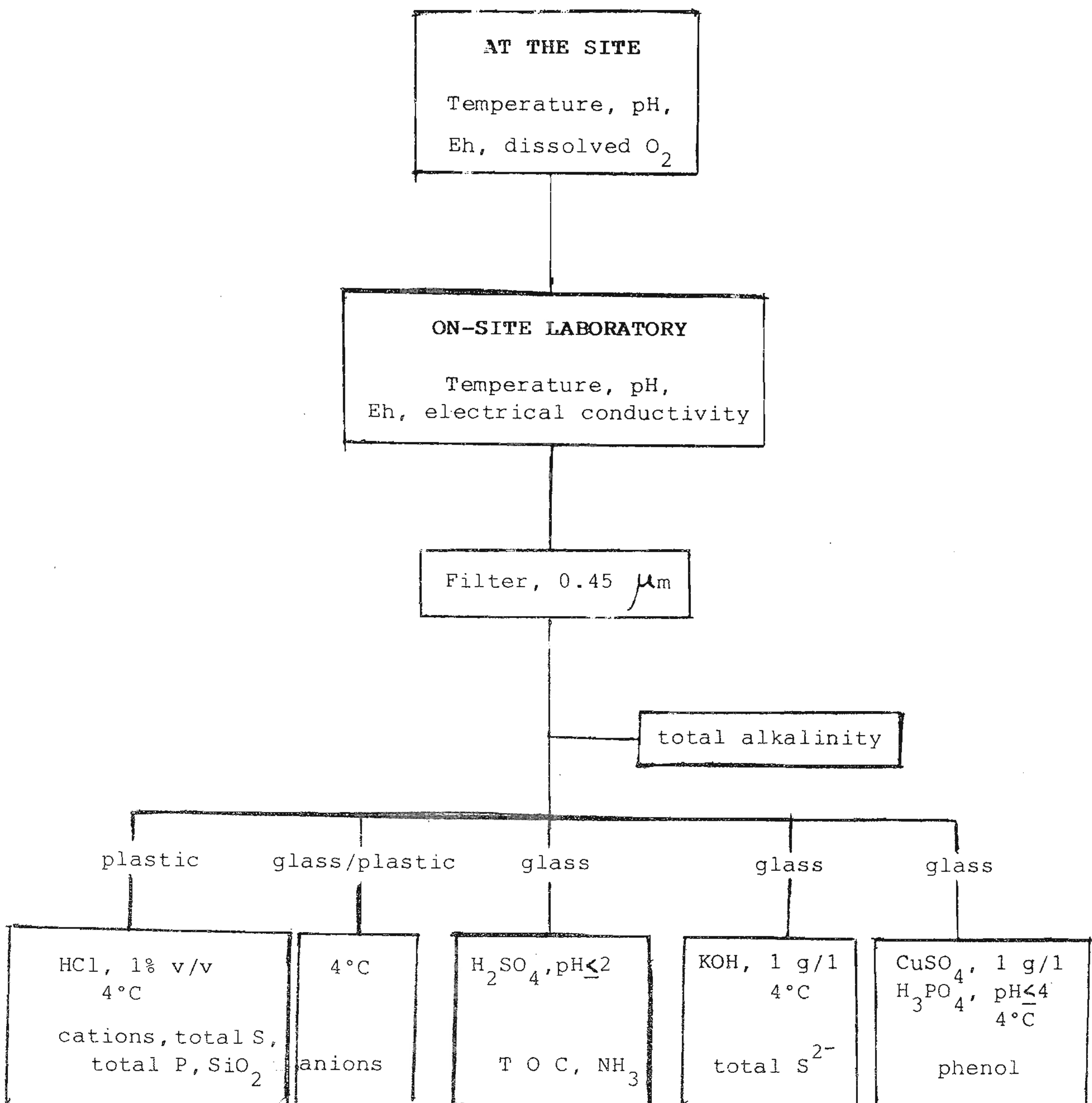
Polyethylene containers should be used when traces of silicon or sodium are to be determined. These are also used when small concentrations of cations are of interest. Lithium and copper contamination from polyethylene containers has been reported. Glass containers are preferred when Lithium, Copper and organic compounds are of interest. Iron, manganese, zinc and lead may be leached from glass. Therefore, glass containers are unsuitable if cations except lithium, copper and ammonium are of interest. In general, it is best to check experimentally for one's own particular conditions (available plastic and glass containers) whether there is a contamination from the container (or loss to the container).

(iii) Satisfactory sample stability may often be achieved by the addition of a reagant to the sample directly after collection or to the empty sample container before collection. Many different reagents have been used. The conclusions reached by different workers on the relative suitabilities of these reagents do not always agree.

There is a general agreement that acidification of samples is necessary when trace metals are to be determined. The addition of a mineral acid (nitric, sulphuric, or hydrochloric) to bring down the pH of the sample to 1-2 is recommended. This will prevent the chemical reactions leading to the formation of hydroxides which subsequently will be adsorbed on to the wall of the container. The acidification to pH 1 ensures stability of elements such as iron, copper, nickel, aluminium and zinc for at least several weeks. The addition of 0.8 cm^3 of concentrated sulphuric acid is also recommended to preserve ammonia and nitrate.

Biocidal reagents such as mercuric chloride, toluene etc. are used to reduce the biological activity in phosphorus and nitrogen compounds. 5 cm^3 of chloroform per dm^3 of the sample are recommended to preserve

phosphate. If the determination of total alkalinity and calcium, magnesium, chloride and sulphate is involved, no preservatives should be added. General scheme for sample containers and preservation is given below :



Sometimes the preserving reagants may interfere in the analytical procedure. In our recent exercise on interlaboratory comparisons of analytical methods, we found that the standard colorimetric procedure for nitrate using phenol 2.4 disulphonic acid reagent suffers interference from sulphuric acid which is used as a preserving reagent for nitrate. But the copperized-cadmium reduction method could not be affected by sulphuric acid. Therefore, depending on the analytical method used suitable preserving reagents should be carefully selected.

References :

1. Wilson, A.L. (1974)
"The Chemical Analysis of Water"
The society for Analytical Chemistry, (Royal Society of Chemistry)
London

2. Kolthoff, Eving and Stross (Editors)
"Treatise on Analytical Chemistry"
Part III Volume 2
Wiley - Interscience, Newyork

3. Gunawardhana, H. Dasaratha (co-ordinator), (1986)
Proc. Sri Lanka Assoc. Advt. Science 42, 176

4. idem (1987)
ibid, 43, 206.

NA-59-8

SAMPLING FOR ANALYSIS OF SOME SELECTED FOOD MATERIALS

N.M. Pieris

Analytical Chemistry Section/CISIR

Introduction

In this short time that has been assigned to me I will consider two specific examples that have been encountered by us during the work that we have carried out.

(a) Sampling of Manioc (*Manihot utilissima* Crantz) tubers for determination of Cyanogenic glucoside (total cyanide) content.

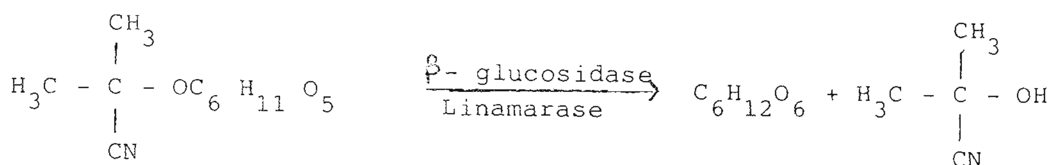
(b) Sampling of fruits and vegetables for determination of vitamin C.

Before getting on to sampling techniques proper, I will give you some information on Cyanogenic glucosides and vitamin C so that a better understanding of the reasons for certain precautions that have to be taken during sampling could be obtained.

Sampling of Manioc tubers for determination of Cyanogenic glucoside content

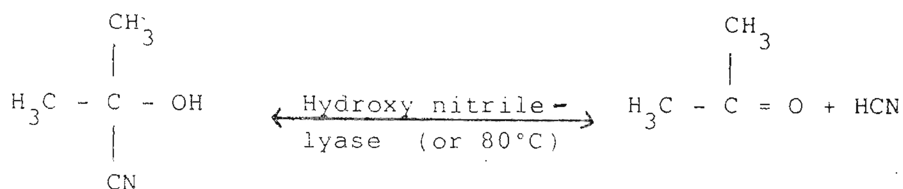
Cyanogenic glucoside

Manioc contains a Cyanogenic glucoside called Linamarin which on hydrolysis yields hydrogen cyanide. The hydrogen cyanide can be estimated and hence the total cyanide content determined.



Linamarin

(β - glucoside)



The glucoside occurs together with an enzyme, Linamarase. The glucoside and the enzyme are usually separated by intracellular localisation and when the cells are damaged come into contact liberating hydrogen cyanide. Therefore when manioc tubers are sampled for determination of Cyanogenic glucoside content any sample preparation that has to be done has to be carried out fast and samples not kept standing before the determination is carried out to obtain a true value for total cyanide content.

All varieties of the plant and all parts of the plant contain the Cyanogenic glucoside.

The concentration of the glucoside depends on (a) variety (b) tissue (c) stage of development and (d) environment.

Distribution of Cyanogenic glucoside in tubers

In the edible part of the root (tuber) the proximal end contains the highest concentration of the glucoside and there is an increase in glucoside concentration from centre outwards, that is the peel contains higher levels than the central part. No two tubers of the same plant contain the same amount of glucoside.

Therefore sampling of tubers become a problem as bulky tubers (500g to 2Kg) with a concentration gradient in both horizontal and radial directions are encountered.

Sampling of tubers

In the determination of total cyanide content a method of selection of representative samples and the ideal sample size had to be known before carrying out the quantitative analysis.

In our work several treatments had to be compared. When the Cyanogenic glucoside content of raw manioc had to be compared with boiled manioc or fried manioc to determine losses on boiling or frying two samples had to be compared. When raw manioc was converted to chips which in turn was converted to flour for use in preparation of processed products such as bread, roti or pittu or when chips were converted to starch several samples had to be compared. When the total cyanide content on ageing of tubers had to be determined the cyanide content from tuber to tuber had to be compared. Sampling procedures had to be worked out for each of the above.

For comparison of two samples the sampling method that was adopted was that of quartering the tuber and pooling opposite quarters. The samples were blended until completely homogeneous before and after treatment. The results obtained as given in Table 1 show that this sampling procedure was adequate.

Table 1

Sampling of whole manioc

Expt. No.	Total Cyanide (mg/kg, fresh weight)	
	Sample A	Sample B
1	122	124
2	51	50
3	50	50

When a comparison of more than two samples had to be carried out such as in the conversion of raw manioc into chips and drying chips in a forced draft oven or still oven at various temperatures or on sun drying, the sampling procedure that was adopted was a sampling of chips at random so that each sample contained a total of 50 to 125 chips weighing 100 to 250g on dry weight or 200 to 500g on wet weight. The results obtained by this sampling procedure are as given in Table 2.

Table 2

Sampling of Manioc chips

Batch No.	Total Cyanide (mg/kg, dry weight)
1	209 ± 11
2	64 ± 4
3	138 ± 7
4	101 ± 6
5	135 ± 13

Number of samples in each batch was 6.

For sampling of Manioc tubers for ageing experiments whole manioc tubers were sampled at random with a sample size of at least 2.5Kg for each sampling point. The results obtained are as given in Table 3.

Table 3

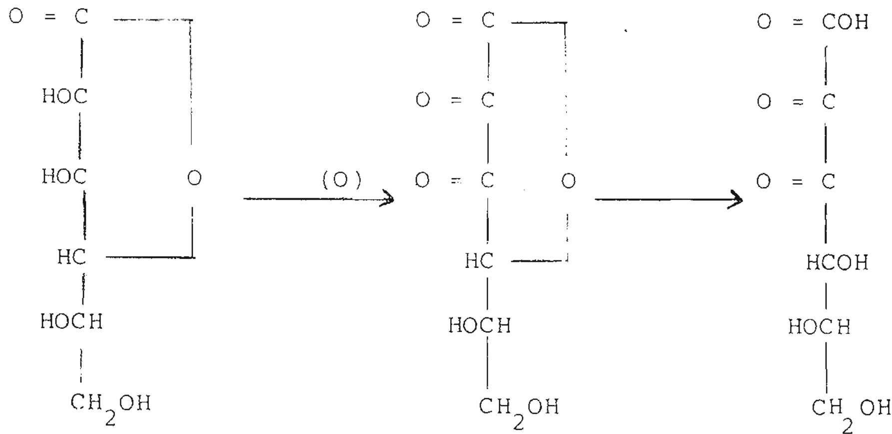
Sampling of Manioc tubers

Batch No.	Total Cyanide (mg/kg, dry weight)	
	Sample A	Sample B
1	102	106
2	147	104
3	106	129
4	136	151
5	137	157

Sampling of Fruits and Vegetables for determination of Vitamin C content

Vitamin C

Ascorbic acid is the main source of Vitamin C.



L-ascorbic acid L-dehydro ascorbic acid 2,3 diketo gulonic acid

L-ascorbic acid slowly undergoes oxidation to dehydro ascorbic acid which is then converted to 2,3 diketo gulonic acid and finally to smaller break down products. The rate of oxidation depends on temperature, light levels of oxidising enzymes or of certain metallic ions which act as catalysts to the breakdown process.

Ascorbic acid and dehydro ascorbic acid have similar activity but diketo gulonic acid is inactive.

Obtaining a representative sample is a very important part of any analysis but is particularly so for vitamin C. Liquid samples such as fruit juices, cordials etc. or solid samples such as artificial drink powders or processed products are fairly straight forward and only require mixing before sampling.

The real problem arises when fruits and vegetable products have to be sampled.

The vitamin C content in most materials varies with (a) variety (b) maturity (c) age after harvesting (d) environment and (e) season.

Sampling

The distribution of vitamin C is anything but even in a single product. Many vegetables eg. Cabbage (Brassica oleracea) have considerably higher concentrations of the vitamin in the outer layers compared with the middle. (Table 4) Similarly the distribution round the head is not even. Hence it is necessary to take segments such as in an orange and make extracts of these. It is necessary to take far more samples than is required for the analysis in order to ensure that the final extract is representative of the sample.

TABLE 4

Distribution of Vitamin C

Common Name	Botanical Name	Vitamin C content in edible, mg/100g
Cabbage (outer leaves)	Brassica oleracea	72 - 76
Cabbage (inner leaves)		21 - 28

In most instances a certain weight of the sample is macerated with the extracting solution and made up to a large volume and aliquots taken for analysis. This type of procedure was satisfactory for samples containing a small percentage of insoluble solids such as juices. However this method was unsatisfactory for samples containing substantial levels of insoluble solids such as in pineapple (Ananas sativas), Wood apple (Feronia limonia) or Beli (Aegle marmelos). In such cases the insoluble matter has to be filtered and reextracted with solvent. In some cases it was necessary to repeat this process and the two or more extracts bulked.

Precautions to be taken

In general samples for vitamin C analysis should not be stored. Fresh food should be analysed within a day or so at the most and the food should not be broken or cut unless the analysis can be completed. Cooked food should be frozen rapidly, preferably adding extracting solution and kept frozen until the analysis could be made.

Distribution of Vitamin C

The variation of vitamin C content with maturity (Table 5), variety (Table 6), and part of plant (Table 7), is given below illustrating that specifying the above is very important in expression of results.

Table 5

Variation of Vitamin C with maturity

Common Name	Botanical Name	Vitamin C content in edible, mg/100g
Passion fruit leaves (tender)	_Passiflora edulis_	500 - 515
Passion fruit leaves (mature)		300 - 317
Beheth Nelli (immature)	_Phylanthus emblica_	100 - 120
Beheth Nelli (mature)		300 - 330
Guava (immature)	_Psidium gujava_	16 - 18
Guava (mature)		195 - 205

Table 6

Variation of Vitamin C with variety

Common Name	Botanical Name	Vitamin C content in edible, mg/100g
Green Chillies	Capsicum annum	125 - 135
Capsicum (malu miris)	Capsicum grossum	90 - 95
Beheth Nelli	Phylanthus emblica	300 - 330
Rata Nelli	Phylanthus disticnus	7 - 11
Guava	Psidium gujava	195 - 205
Guava (cheenapera)	Psidium cattleianum	2 - 4

Table 7

Variation of Vitamin C with part of plant

Common Name	Botanical Name	Part	Vitamin C content in edible, mg/100g
Passion fruit	Passiflora	leaves	300 - 317
	edulis	fruits	12 - 15
Drumstick	Moringa	leaves	125 - 135
	pterigosperma	fruits	130 - 135
Katurumurunga	Sesbania grandi-	leaves	97 - 160
	flora	flowers	22 - 25

Conclusion

In conclusion a word of advice to the young researcher - (1) Study the parameter to be determined thoroughly so that the necessary precautions could be taken whilst sampling. (2) Determine distribution in the sample, this would help not only in devising methods of sampling but also in determining sample size. (3) The history of the sample must be known (maturity, age after harvesting etc.) if not the results will not mean much. As far as possible obtain samples for analysis directly from the field specially in determining real values eg. in the determination of vitamin C in fruits or vegetables. (4) Analyse at least six samples of the same type so that a range could be obtained. (5) As far as possible the variety of the plant material the environment and the season should be known so that specific values could be obtained.