AN OIL SPILL SAMPLING STRATEGY

by
WOOLLCOTT SMITH

Reprinted from

SAMPLING BIOLOGICAL POPULATIONS
R. M. Cormack, G. P. Patil, and D. S. Robson (eds.)
International Co-operative Publishing House
Fairland, Maryland

pp. 355-363

A publication from the
Satellite Program in Statistical Ecology
International Statistical Ecology Program

Statistical Ecology Series
Volume 5

International Co-operative Publishing House
P.O. Box 245
Burtonsville, Maryland 20730

Statistical Ecology Series

General Editor: G. P. Patil

Volume 1*	SPATIAL PATTERNS AND STATISTICAL DISTRIBUTIONS edited by G. P. Patil, E. C. Pielou, and W. E. Waters (1971)
Volume 2*	SAMPLING AND MODELING BIOLOGICAL POPULATIONS AND POPULATION DYNAMICS edited by G. P. Patil, E. C. Pielou, and W. E. Waters (1971)
Volume 3*	MANY SPECIES POPULATIONS, ECOSYSTEMS, AND SYSTEMS ANALYSIS edited by G. P. Patil, E. C. Pielou, and W. E. Waters (1971)
Volume 4	STATISTICAL DISTRIBUTIONS IN ECOLOGICAL WORK edited by J. K. Ord, G. P. Patil, and C. Taillie (1979)
Volume 5	SAMPLING BIOLOGICAL POPULATIONS edited by R. M. Cormack, G. P. Patil, and D. S. Robson (1979)
Volume 6	ECOLOGICAL DIVERSITY IN THEORY AND PRACTICE edited by J. F. Grassle, G. P. Patil, W. K. Smith, and C. Taillie (1979)
Volume 7	MULTIVARIATE METHODS IN ECOLOGICAL WORK edited by L. Orloci, C. R. Rao, and W. M. Stiteler (1979)
Volume 8	SPATIAL AND TEMPORAL ANALYSIS IN ECOLOGY edited by R. M. Cormack and J. K. Ord (1979)
Volume 9	SYSTEMS ANALYSIS OF ECOSYSTEMS edited by G. S. Innis and R. V. O'Neill (1979)
Volume 10	COMPARTMENTAL ANALYSIS OF ECOSYSTEM MODELS edited by J. H. Matis, B. C. Patten, and G. C. White (1979)
Volume 11	ENVIRONMENTAL BIOMONITORING, ASSESSMENT, PREDICTION, AND MANAGEMENT — CERTAIN CASE STUDIES AND RELATED QUANTITATIVE ISSUES edited by J. Cairns, Jr., G. P. Patil, and W. E. Waters (1979)
Volume 12	CONTEMPORARY QUANTITATIVE ECOLOGY AND RELATED ECOMETRICS edited by G. P. Patil and M. L. Rosenzweig (1979)
Volume 13	QUANTITATIVE POPULATION DYNAMICS edited by D. G. Chapman and V. Gallucci (1979)

*For these first three volumes, contact:
The Pennsylvania State University Press
University Park, PA 16802 USA

For all of the remaining volumes, contact:
International Co-operative Publishing House
P.O. Box 245
Burtonsville, MD 20730 USA

R. M. Cormack, G. P. Patil, and D. S. Robson, (eds.),
Sampling Biological Populations, pp. 355-363.

AN OIL SPILL SAMPLING STRATEGY

WOOLLCOTT SMITH

Woods Hole Oceanographic Institute
Woods Hole, Massachusetts 02543 USA

SUMMARY. A strategy for oil spill surveys is outlined. It depends on over-sampling in the first survey, and recording easy-to-measure variables associated with each sample. In the second stage, the samples are sub-sampled for the more expensive and time-consuming biological and chemical analysis. This procedure is an extension of double sampling. This kind of survey is easy to implement on short notice and, because it is based on classical sampling theory, the survey's results should be relatively easy to defend in administrative and legal proceedings.

KEY WORDS. double sampling, survey design, oil spill, environmental surveys.

1. INTRODUCTION

In most survey design problems it is assumed that the investigator has a good deal of knowledge about the processes he is investigating. The goals of the survey are stated and the survey constraints are well defined. Thus, before designing a survey, the investigator has a clear quantitative definition of the survey design problem. An oil spill is an unplanned event. It may occur in an area of the world where there is little knowledge about the biology, sediment type, or water circulation patterns. Even if this information were available, it would be difficult on short notice to assemble the baseline data to develop an adequate sampling model.

At the time an oil spill survey program is planned there may be no consensus about the important questions that data from a survey should answer. These questions may be developed in legal proceedings long after the relevant data should have been collected. Finally, at the initiation of the spill study, the constraints, such as the total cost of the spill study, cannot be specified. Even under these circumstances, classical statistical theory can give at least some qualitative rules for survey design.

Environmental studies, and oil spill studies in particular, differ in many fundamental respects from the agricultural experiments that were the impetus for much of the statistical work in experimental design. The first and most important difference is that the basic experiment, the oil spill, is not repeatable. To see what this means, suppose we have an oil spill of 50,000 barrels of No. 2 fuel oil in Xanadu Bay on April 1, 1984. The basic environmental and legal question would be: what is the difference in the biology and chemistry of the bay due to the oil spill? That is, we wish to contrast the properties of the bay if there had not been a spill with the properties of the bay with an oil spill.

An idealized experimental design would have four replicate Xanadu Bays on April 1, 1984. Spill 50,000 gallons of No. 2 oil in two bays and leave the others untreated. This would be the most efficient, if impractical, experimental design to answer the basic environmental questions. If we could implement this design there is still the further question of how to sample efficiently the four replicate Xanadu Bays. Of course, the problem of non-repeatability is partially alleviated if there is environmental base line data for the bay covering several years prior to the spill. Most oil spills now occur in areas with inadequate base line data. It is this situation that must be considered carefully. If we have exactly one experimental unit, we must evaluate the effect of the oil spill within that unit. To the extent that this is true, our approach will follow closely classical survey design, rather than experimental design.

The second more practical but less fundamental problem is that oil spill studies must be implemented at short notice. Sampling of the water column and benthos must begin shortly after the oil spill event. The question that classical sampling design addresses is how to take samples to achieve a fixed level of accuracy or how best to design a sampling program given that the total cost of the program has been fixed. In most oil spill studies the sampling program must begin before the total expense of the sampling program has been fixed or the desired precision of the sampling program fully defined. Any proposed sampling program must deal with this annoying problem realistically.

In evaluating the effect of an oil spill it is important to determine the rate of recovery from the spill. To do this samples must be taken at several times. The number and distribution of the time points will greatly affect both the cost and the accuracy of the sampling program. Some literature on sampling time-varying systems (Smith, 1978; Scott and Smith, 1974) is generally applicable to the oil spill and other environmental survey problems. This problem is beyond the scope of the present note. However, it is an area where modern statistical methods would make a useful contribution.

2. THE INITIAL OIL SPILL SURVEY

In this section we set forth the general procedures for designing an oil spill survey. The basic constraints were discussed in the Introduction. They are: 1) There is only one experimental unit. The effects of the oil spill must be judged in terms of differences between relatively unaffected areas and those more severely affected. 2) The total cost and effort to be allocated to the entire survey is unknown at the start of the survey program. 3) Base line data and detailed knowledge of the sediment structure is not available. Against this formidable array of difficulties the oil spill investigator has one important practical advantage. The cost of collecting and preserving oil spill samples is small in comparison with cost of doing the chemical and biological analysis of the samples. Thus, our sampling procedure will be based on our ability to oversample the area, then to refine our procedure for subsampling the original samples to obtain the biological and chemical data.

The kind of sampling program we have in mind is a variation of the double sampling described in Cochran (1962, p. 327-354). For example, suppose we are interested in estimating the effect of the oil spill on the benthic infauna. A simplified sampling model would assume that the density of the infauna is controlled by three effects: 1) sediment type; 2) depth of sample; and 3) presence of oil at the station. These effects are assumed to be fixed. In addition to these fixed effects, there are three kinds of random variations associated with the sampling model: 1) random error due to measurement and lab techniques; 2) variation due to small scale heterogeneity in the sediment (patchiness); and 3) variation due to biological, physical, and chemical processes not accounted for by the fixed effects in the sampling model. Our problem then is to find the set of samples which will most precisely estimate the effect of the oil spill on, for example, the benthic infauna.

The optimum sampling strategy depends on the sampling model proposed and also on the parameters of the model for which accurate estimates are needed. The reader should bear in mind that if different questions are asked, the sampling program will need to be adjusted. The following discussion defines what we believe to be the basic environmental question, and shows how a double sampling program can be used efficiently to answer that question.

The first step in defining our sampling model is to divide the area of the spill into subareas or strata. It is hoped that the subareas are relatively homogeneous with respect to biological and chemical properties we wish to measure. For example, in the benthic infauna study we might define four strata by dividing the area into intertidal and subtidal and by a rough division of the sediment into areas which are greater than 50 per cent silt-clay and those which are less than 50 per cent silt-clay. We let $W_i^{(o)}$ denote the proportion of the oil spill area in strata i and let $W_i^{(c)}$ denote the proportion of the control area in strata i. Now let $\overline{Y}_i^{(c)}$ and $\overline{Y}_i^{(o)}$ represent the mean value of an environmental variable in the control and affected strata, respectively. We might display these values in the following tabular form:

affected area

	clay-silt $\geq$ 50%	clay-silt $\leq$ 50%
subtidal	$W_1^{(o)}$	$W_2^{(o)}$
intertidal	$W_3^{(o)}$	$W_4^{(o)}$

control area

	clay-silt $\geq$ 50%	clay-silt $\leq$ 50%
subtidal	$W_1^{(c)}$	$W_2^{(c)}$
intertidal	$W_3^{(c)}$	$W_4^{(c)}$

We can now state more precisely the parameters of the population that we wish to estimate. The original problem is to estimate the difference between the mean value of the variable within the affected area and the mean value of that variable if the area had not been affected. That is, we wish to estimate the difference between

$$\overline{Y}^{(o)} = \Sigma W_i^{(o)} \overline{Y}_i^{(o)} ,$$

the mean for the affected area, and

$$\overline{Y}^* = \sum_i W_i^{(o)} \overline{Y}_i^{(c)} ,$$

which is the mean for the control areas weighted by the proportion of strata that are found in the affected area. Combining these equations we have a weighted mean of the difference between affected and control strata

$$\overline{\Delta} = \sum_{i=1}^{n} W_i^{(o)} \left(Y_i^{(o)} - Y_i^{(c)} \right) .$$

Thus the parameter $\overline{\Delta}$ contrasts the affected area with a hypothetical control area which has identical stratification.

The contrast is defined in terms of unknown strata weights W_i and unknown strata means $\overline{Y}_i$. The sampling design problem is to find a sampling program that will yield good estimates of $\overline{\Delta}$ as well as of the strata means and strata proportions. To do this we propose a double sampling scheme that takes advantage of the relative ease in determining the stratum to which a sample belongs to.

Each station sampled has a set of properties associated with it. Some of these properties are easy and inexpensive to measure; they include the x,y position of the station, depth of the station, whether the station is inside or outside the oil slick, and sediment composition. These easily obtained data will be used to estimate the strata proportions. The original samples can then be subsampled to determine the biological differences between the control and affected area.

The notation and algebra for the full sampling model is cumbersome and follows from the minor extensions of the double sampling results in Cochran (1963). These are detailed in Smith (1975). However, the method is neither difficult to understand nor to apply.

We consider the hypothetical oil spill and sampling program pictured in Figure 1. The sampling program consists of 115 original samples. Distribution into the four strata are given in Table 1. Now suppose that after the original sampling program is completed it is determined that we can only analyze the chemical and biological properties of 24 of these samples. This limit might be due to cost, limited time, or the number of personnel available to do the more demanding analysis.

TABLE 1: Grid sampling program. Upper row: original samples (N=115); lower row: subsamples for biological analysis (n=24).

affected area

	clay-silt ≥ 50%	clay-silt ≤ 50%
subtidal	8	18
intertidal	0	9

control area

	clay-silt ≥ 50%	clay-silt ≤ 50%
subtidal	14	53
intertidal	0	13

affected area

	clay-silt ≥ 50%	clay-silt ≤ 50%
subtidal	3	6
intertidal	0	3

control area

	clay-silt ≥ 50%	clay-silt ≤ 50%
subtidal	3	6
intertidal	0	3

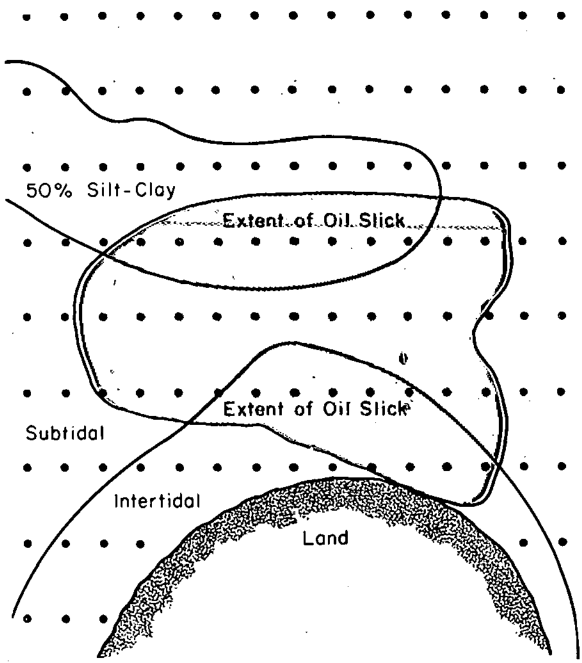

FIG. 1: Hypothetical oil spill and grid sampling.

Let $N_i^{(o)}$ and $N_i^{(c)}$ denote the number of samples that fall in stratum i of the affected and control areas respectively, and let $n_i^{(o)}$ and $n_i^{(c)}$ denote the number of subsamples of the original sample from each stratum. An unbiased estimate of $\overline{\Delta}$ is

$$\hat{\overline{\Delta}} = \Sigma w_i^{(o)} (\bar{y}_i^{(o)} - \bar{y}_i^{(c)}) ,$$

where $w_i^{(o)}$ is the estimate of $W_i^{(o)}$

$$w_i^{(o)} = \frac{N_i^o}{N^o}$$

and $\bar{y}_i^{(c)}$ and $\bar{y}_i^{(o)}$ are the estimated means for the control and affected areas, respectively.

The notation and algebra for the variance of the estimator (1) are cumbersome and will not be given here. However, it is clear that an efficient subsampling program would subsample strata in proportion to their size in the affected area; see Table 1.

3. CONCLUSIONS

We have tried to set the problem of estimating the effect of an oil spill from a sample survey in the context of formal sampling theory. Although the notation and algebra can be complex, the basic principles of sampling an oil spill for a field biologist are straightforward. We will summarize them here.

A. *Sampling program to estimate the total effect of an oil spill.*

1) Samples should be taken in a grid pattern covering the entire area.

2) All easily obtainable ancillary data should be collected for each sample, including position, depth, whether sample is in slick, etc.

3) Ancillary information can be used to select subsamples for biological and chemical analysis that will yield the most information about effects of the spill.

4) Size of the grid surveyed should be determined by the cost and manpower available to collect samples and not by the cost of analyzing the samples.

5) The control area should include sediment and water depths similar to the affected area.

6) Deciding the number of samples used in the biological and chemical analysis is a complex function of cost, information gained, and usefulness of the information in assessing the effect of the oil spill. This decision should be made after evaluation of the preliminary data and with the consultation of biologist, chemist, lawyer, and statistician.

B. Repeated sampling to investigate response over time to an oil spill.

1) A small number of benthic stations can be used to evaluate the effect of the oil over time.

2) Both control and affected stations should be carefully matched for sediment type and depth.

3) Returning to the same station is important to increase the accuracy of the sampling program.

The legal and environmental questions that are raised may differ from one spill to the next. We have outlined a sampling plan that is relatively flexible and allows a more exact statement of the problem after the initial survey is made.

A sampling plan like this has two practical advantages to the oil spill investigator. The first is that the field part of the sampling program is easy to implement at short notice. Second, the difficult part of the plan, subsampling the original samples, can be postponed until the goals of the particular study are better defined. The sampling plan outlined is based on well-known classical sampling procedures and would be relatively easy to defend in an administrative or legal proceeding. Katz (1975) discusses the problems faced by a statistician in presenting survey results in a legal setting.

Finally, the conclusions that can be drawn from an oil spill survey are necessarily weaker than those that can be drawn from a carefully designed experiment. We can evaluate the differences in the properties of the control and affected areas selected. However, unlike a carefully designed experiment, we cannot, from the survey data alone, assign a direct or indirect causal relationship between the oil spill and differences observed in the survey. One must turn to scientific results from controlled experiments of the effects of oil to show the probable relationship between the impact of oil on the survey area and results of the survey analysis. Both the scientific evidence of the effect of oil in the environment and the survey results for a particular oil spill are necessary components in establishing the impact of an oil spill.

ACKNOWLEDGEMENT

This work was supported by NOAA Sea Grant 04-8-M01-149 and the Ocean Industries Program of the Woods Hole Oceanographic Institution.

REFERENCES

Cochran, W. G. (1963). *Sampling Techniques.* 2nd ed. Wiley, New York.

Katz, L. (1975). Presentation of a confidence interval estimate as evidence in a legal proceeding. *American Statistician*, 29, 138-142.

Scott, A. J. and Smith, T. M. F. (1974). Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association*, 69, 674-678.

Smith, W. K. (1978). Environmental design: a time series approach. *Estuarine and Coastal Marine Science*, 6, 217-224.

Smith, W. K. (1975). An oil spill sampling strategy. Unpublished manuscript.

[*Received July* 1977. *Revised March* 1979]

INTERNATIONAL STATISTICAL ECOLOGY PROGRAM

The International Statistical Ecology Program (ISEP) consists of the activities of the Statistical Ecology Section of the International Association for Ecology and of the Liaison Committee on Statistical Ecology of the International Statistical Institute, the Biometric Society, and the International Association for Ecology. The ISEP is a non-profit program formulated to serve the needs of interdisciplinary research and training in the newly emerging fields of Statistical Ecology and Ecological Statistics.

SATELLITE PROGRAM IN STATISTICAL ECOLOGY

The Second International Congress of Ecology was held in Jerusalem during September 1978. In this connection, ISEP organized a Satellite Program in Statistical Ecology during 1977 and 1978. The emphasis was on research, review, and exposition concerned with the interface between quantitative ecology and relevant quantitative methods. Both theory and application of ecology and ecometrics received attention. The Satellite Program consisted of instructional coursework, seminar series, thematic research conferences, and collaborative research workshops.

Research papers and research-review-expositions were specially prepared for the program by concerned experts and expositors. These materials have been refereed and revised, and are now available in a series of ten edited volumes listed on page ii of this volume.

The Satellite Program takes as its theme the better melding of fundamental ecological concepts with rigorous empirical quantification. The overall result should be progress toward a stronger body of general ecologic and ecometric theory and practice.

FUTURE DIRECTIONS

The satellite-like-programs help create and sustain enthusiasm, inward strength, and working efficiency of those who desire to meet a contemporary social need in the form of some interdisciplinary work. It should be only proper and rewarding for everyone involved that such programs are planned from time to time.

Plans are being made for a satellite program in conjunction with the next Biennial Conference of the International Statistical Institute and the next International Congress of Ecology. Care should be exercised that the next program not become a mere replica of the present one, however successful it has been. Instead, the next program should be organized so that it helps further the evolution of statistical ecology as a productive field.

The next program is being discussed in terms of subject area groups. Each subject group is to have a coordinator assisted by small committees, such as a program committee, a research committee, an annual review committee, a journal committee, and an education committee. This approach is expected to respond to the need for a journal on statistical ecology, and also to the need of bringing out well planned annual review volumes. The education committee would formulate plans for timely modules and monographs. Interested readers may feel free to communicate their ideas and interests to those involved in planning the next program. The mailing address is: International Statistical Ecology Program, P. O. Box 218, State College, PA 16801, USA.

International Statistical Ecology Program
Satellite Program in Statistical Ecology

CONTENTS OF EDITED VOLUMES

BACKGROUND

Close to 350 lectures and discussions were held during 50 days in the middle of the summer seasons of 1977 and 1978. The program was endowed with stimulating local setting as well as with atmosphere conducive for scholarship. Participants represented broad background and expertise. Advice was sought from a number of special advisors. The advice received was immensely helpful in guiding selection of the best experts in the field to achieve as representative and balanced a coverage as possible.

The editors together with the referees took a rather critical and constructive attitude from initial to final stages of preparation of papers by offering specific suggestions concerning the suitability, and also the structure, content, and size. These efforts of coordination and revision were intensified through editorial sessions at the program itself as a necessary step for the benefit of both the readers and the participants. Everyone went by scientific interests often at the expense of personal preferences. The program atmosphere became truly creative and friendly, and this remarkable development contributed to the maximal cohesion of the program and its proceedings within the limited time period available.

The edited research papers and research-review-expositions prepared for the program are organized in ten volumes. Altogether, they consist of an estimated 4,000 pages of research, review, and exposition in addition to a common foreword in each, followed by individual volume introductions. Subject and author indexes are also prepared for each volume. May this ten volume set serve as a useful stimulus to further research and training in statistical ecology for the nineteen eighties.

CONTENTS OF THE VOLUME

corresponding to this reprint