INSECT POPULATION SAMPLING*

K. GRAHAM† AND R. W. STARK‡

GENERAL CONSIDERATIONS

Purpose of Sampling

Sampling has four main functions:—

1. The *qualitative* determination of what is present within an area. This includes insects of taxonomic interest, noxious insects, and biotic agents of control, such as parasites and predators.

2. The *quantitative* determination of the status of a population within the area. It is desired to know the population levels of a particular insect, its parasites and predators. These determinations may be required to estimate the efficacy of artificial control. When done empirically, this is limited in scope and application to occasions requiring an immediate decision for treatment purposes or where interest or funds are not sufficient to make more detailed studies. Generally speaking, these are inadequate for determining success or failure of a control operation.

3. Usually, and particularly with those insects of economic importance, it is not sufficient to know only the present status of a population. It is also necessary to know something of the *dynamics* of the population; that is, what the population trends are when successive generations are studied.

4. In widely distributed insect populations it is often desirable and necessary to investigate the *ecology* of the species. This might include zones of abundance according to the locality, merely as a fact, or the abundance related to specific conditions such as physical environment or biotic community.

These four functions cover, in a broad sense, all of the purposes of sampling. The fields of sampling of the greatest economic importance and subject to the most error and criticism are the quantitative methods. Certainly, those methods dealing with the dynamics of insect populations are of the greatest use. From these stem reliable estimates of present and future insect damage, decisions as to the necessity for control, and information important to the application of biological or artificial control.

It seems timely to review some of the commoner sampling problems encountered in various fields of entomology, and the premises and techniques on which reliable sampling is based. Without suitable sampling we cannot use the tools of statistical analysis to describe the character of variability, to prove the reliability of estimates, or to show the significance of apparent differences. Therefore, it is the purpose of this discussion to review briefly the above considerations and illustrate them with samples from workers in the various fields.

Collection of Data

Everyone who studies insects has some need to sample even if it is in the empirical sense of the collector. Even he has to arrive at some decision as to the adequacy of a variable series. For example, most insects are not so constant morphologically that single specimens are truly representative of the species.

When it is a question of determining a population within a given natural universe and determining the mortality factors within it, the problems become much more complex. The most fundamental requirement is that *the sampling of a particular insect population must be resolved about the distribution and life-cycle of the insect involved.* There is no “ universal ” sampling method.

The environments of insects achieving economic importance comprise a tremendous variety, from onion-crops to mature

---

* Contribution No. 143, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada. Report of a discussion at the Fifty-third Annual Meeting of the Entomological Society of British Columbia, March 13th, 1954.

† Department of Zoology, University of British Columbia.

‡ Agricultural Research Officer, Forest Biology Division, Calgary, Alta.

forest stands, each of which has individual variations depending on age, density, site, shape, and other factors. In selecting the sampling technique to be followed, there are several factors to consider. At what stage in the life-cycle will it be easiest to sample, considering the eventual treatment of the data as well as more practical considerations such as economy. For example, nearly every insect has a stage that is more or less static. Samples taken during this stage lend themselves well to sound, and standard, statistical procedures in the treatment and interpretation of data. However, sampling of an immobile stage, although easier, may not be of sufficient value nor suited to the purpose of the investigation. This condition arises in artificial control programmes and where the expected trend from single sampling is desired.

A common problem is that populations, regardless of the stage sampled, are never distributed so uniformly throughout a sampling universe that a few observations can give a reliable indication of the average density. Obviously, neither mortality nor parasitism can be judged from single specimens. It is clear that there must be a compromise between the proportion of the sampling universe covered and the number of individuals examined (Oakland 1953). The spatial pattern of area coverage also affects the reliability of the estimate.

Accepting the fact that only an approximation of the truth must suffice, no matter how statistically sound, several problems are presented in seeking the nearest approximation to that truth with the greatest economy of facilities at hand. These are not so much in actual techniques as in the questions—where? how much? and how many?

Treatment and Interpretation of Data

To understand fully the requirements of sampling, the patterns of variability within a sampling universe must be considered. Much emphasis has been given to the concept of random sampling because it eliminates conscious or unconscious bias. There are many circumstances in nature, however, which make unrestricted random sampling illogical, especially where major subdivisions in the environment can be defined. In ecological studies it is often most logical to retain some homogeneity in terms of common features of environment. Thus one may wish to relate populations to restricted forest stands, field crops, individual orchards, and physical environmental factors such as altitude and aspect. This is often referred to as *stratification of sampling;* that is, sampling with respect to known environmental variates. Sampling in that manner is purposive and it has been demonstrated that within the confines of the stratum, sampling randomization can be attained to eliminate bias.

In summary, within any sampling universe there are two conditions of sampling that must be satisfied. These have been stated by Yates (1949), who writes:—

" 1. If bias is to be avoided, the selection of samples must be determined by some process uninfluenced by the qualities of the objects sampled and free from any element of choice on the part of the observer.

" 2. If a valid estimate of sampling error is to be available each batch of material must be so sampled that two or more sampling units are obtained from it. These sampling units must be a random selection from the batch of material, and all the sampling units in the aggregate must be approximately the same size and pattern and must together comprise the whole of the batch of material."

Personal selection, such as a tendency to over- or under-estimate and a tendency to select for particular characteristics, biases the estimation of the mean, and the standard errors derived for these estimates are meaningless.

A word of caution is advisable. The form of analysis of the data depends upon the sampling method used and upon the proportion of the population sampled. This subject is too detailed to include here but is simply and adequately covered in the elementary text by Quenouille (1950).

It has been shown by various authors and is fully discussed in Quenouille that the arithmetic mean of a sample gives a more accurate value for the mean of the population than any other measure, and that the standard deviation will usually give a more accurate measurement of the scatter in a population estimate than any other value. Moreover, most of the properties of a sample can be found from its mean and standard deviation. However, it must be borne in mind, when presenting data, that other expressions of central tendency and scatter, such as the mode, quartiles, deciles, and range, may illustrate a point better.

There is a further point to remember when sampling. Because the objective is to establish the character of the population in general, an attempt is made to estimate the frequency distribution of that population. In biological work it is often possible to estimate the general form of the distribution from field observations prior to sampling and hence to find the type of distribution with a smaller sample.

A distribution which often occurs, but, alas, not so frequently in biological work, is the *normal distribution*. When this applies it is indeed fortunate, for the mean and variance can be used to determine *all* about the population. This is the perfect case, making it simple to present the proportions of observations between any two values in summary form. However, as was stated, this does not obtain frequently in biological populations. Much work has been done in the field of biometrics, and it is not intended here to present a capsule digest, since this would be impossible. All that can be hoped is to stimulate interest and provide references to satisfy that interest.

In biological work, particularly entomology, there have been several population distributions described to suit particular cases. These include Poisson, binomial, negative binomial, Fisher's logarithmic, Neymann's, Cole's, the Thomas double Poisson, and Polya's. In regard to these, attention is called to a recent article which is, in our opinion, the most valuable single contribution in this field in recent years. This is an article by C. I. Bliss and R. A. Fisher (1953). In addition to demonstrating with examples the fitting of this distribution, they illustrate its advantages by comparison with the major types of distributions in use. Furthermore, they supply the reference sources for them. The article is well written and easily readable by anyone with a basic grasp of statistics, and it indicates that the problem of non-random biological distributions is being attacked from the view-point of the man who is faced with the problem; viz., the biologist.

The importance of distributions in quantitative sampling work has arisen from the fact that the statistical tests for significance of data, and, more important, the comparison and interpretation of data such as the variance-ratio test, are largely based upon the normal distribution. To be able to assess and accurately compare the characteristics of a population that is not normally distributed, the distribution must be related in some way to the normal distribution. This brings up the use of *transformations* to fulfil the above requirement. This cannot be elaborated upon here, and again reference is made to more qualified sources. Quenouille contains an easily understandable discussion on the subject, and one other reference of value is the article by Churchill Eisenhart (1947).

Selected References

Bliss, C. I., and R. A. Fisher. Fitting the negative binomial distribution to biological data and a note on the efficient fitting of the negative binomial. Biometrics 9:176–199. 1953.

Eisenhart, Churchill. Assumptions underlying the analysis of variance. Biometrics 3 (1). 1947.

Oakland, G. B. Determining sample size. Canadian Ent. 85:108–113. 1953.

Quenouille, M. H. Introductory statistics. London, Butterworth-Springer Ltd. 1950.

Yates, F. Sampling methods for censuses and surveys. New York. 1949.

FOREST DEFOLIATORS*

Forest-insect population studies during the past decade have shown increasing recognition of the fact that valid conclusions regarding population trends require measurements made within certain limits of accuracy. The value of population estimates and trends is obviously enhanced if these can be related to actual forest damage. Not only the populations of the destructive insects, but also those of their parasites and predators are important. Indeed, all the mortality factors, such as climatic extremes and disease must be accounted for in the most valuable population studies.

There are many examples in the literature of population sampling. De Gryse (1934) described the most important techniques extant in 1934, but there have been many important advances since that time. For illustrative purposes only, three examples are presented which will serve to demonstrate the methods involved and practical applications.

European Spruce Sawfly

This example is drawn from the work of M. L. Prebble (1943), who determined populations of the European spruce sawfly in the Gaspé Peninsula. This exhaustive and detailed work discussed the general requirements of sampling forest insects, demonstrated the difficulties encountered in this particular study, and showed how they were surmounted. The conclusions, based on eight years of field data, pertain solely to the sampling methods applicable to the cocoon stage.

It has been mentioned that choosing the stage of insect to be sampled is of great importance. The most desirable objective is the sampling of all stages related to a common denominator. In the spruce-sawfly work, larval population studies were carried out, but the amount of work required seriously limited their applicability. However, they are of value for long-term ecological studies, such as will be mentioned in a later example. It was determined that the cocoon stage, spent in the moss and debris of the forest floor, was best suited for sampling. It gave the maximum information on sawfly populations as well as the various control factors operating against the cocoon stage. Furthermore, the forest floor constituted a sampling universe that could be statistically regulated.

Three techniques were tested. The first consisted of counts of cocoons in 2- by 2-foot quadrats spaced mechanically and uniformly within representative forest types, regardless of suitability of location of individual quadrats. The second, using the same size of quadrat, required their location in a restricted universe, which was defined as " the area of suitable ground cover lying under dominant and codominant spruce trees in the forest types selected for study." The quadrats had to lie within a ring-shaped band whose inner margin was a circle around the tree 1 foot from the trunk and whose outer margin was bounded by the outer edge of the vertical projection of the crown. The third method used the same restricted sampling universe but divided the quadrats into four equal subsamples. This was tantamount to quadrupling the number of samples in the statistical analysis, although it introduces certain disadvantages. The manner of devising the latter two methods indicated that an *a priori* knowledge of the distribution in the forest floor assisted in reducing the required number of samples.

The first technique was abandoned, as it was found from analysis that the variability was so high that the number of samples necessary to give an accurate estimate was beyond practical limits. Both the second and third methods proved usable, but disadvantages in statistical interpretation were found in each. Without elaborating on these, it was concluded that where successive samples were to be taken in the same plot, the second method was the better; i.e., using the tree as a sampling unit. This would be usual prac-

---

* This section written by R. W. Stark.

tice when we are interested in population fluctuations.

An interesting side investigation reported in this paper was a comparison of the variability between the efficiency of workers. Its purpose was to determine the advisability of applying a correction factor to account for missing cocoons. No correction was needed, but checks were necessary to bring the counts to a satisfactory degree of accuracy.

An additional reference of interest is the discussion by Butcher (1951) of forest-insect sampling problems with special reference to the larch sawfly, whose habit of overwintering in cocoons in the soil renders the sampling problem similar to that of the spruce sawfly.

The Lodgepole Needle Miner

Intensive study of this insect began in 1948, and it was 1951 before adequate sampling methods were derived. Graham, in 1951 (1952), presented to this society the ecological aspects of the problem; therefore, only methodology will be discussed here.

Statistical analysis in 1951 demonstrated that early coverage of the outbreak in the National parks was almost empirical. However, this analysis did provide the *a priori* knowledge, so often useful, for more efficient work. It demonstrated the practicability of using the branch tip as a sampling unit and indicated how many samples would be required to give a sampling accuracy of 10 per cent. Because a leaf-mining insect was involved, choice of the stage to be sampled was not the difficult problem that it is with most open leaf-feeders.

In the summer of 1951 four branch tips from each of 105 trees were sampled to establish the sampling method. This number of trees was predetermined only approximately; sampling was continued until the error was below 10 per cent. The sampling was done with respect to altitude, and it was established, population-wise, that we may consider three sampling universes with respect to altitude and two with respect to pine-trees. This was established from an analysis of variance between crown levels (arbitrarily dividing the crown into upper and lower, the division at the midpoint) and three altitude levels—valley-bottom (approximately 5,000 feet), valley-bottom plus 750 feet, and plus 1,500 feet. In the Rocky Mountains the latter elevation is usually close to timber-line. From the analysis of these samples it was determined that the counts of live larvæ in the sampling unit of a five-year branch tip were acceptable, and that, for practical usage, four branch tips from approximately thirty trees would give absolute population estimates within 10 per cent accuracy (Stark, 1952(*a*)). This was applicable to all stages found in the needle, from first instar to pupæ. However, to compensate for extremely high values in the low infestation counts, a transformation of the data was required. The transformation was supplied by G. B. Oakland, of the Biometrics Unit, Ottawa, as $\sqrt{x+0.5}$, where x is the individual sample. Analysis of the transformed data corroborated the assumption that the distribution was, in effect, normal. More important, use of the transformation made possible a new application of the sequential sampling technique (Stark, 1952(*b*)).

The sequential formulæ were supplied by Oakland for the normal distribution. This technique is proving extremely useful in lodgepole-needle-miner surveys. Briefly, it is sampling with no fixed sample size. Limits of infestation classes and tolerated error are set, and the appropriate formulæ are applied. What is arrived at is graphical (or tabular) limits composed of upper and lower acceptance levels. Sampling is continued until the cumulative sum of samples is smaller or larger than the acceptance levels, thereby falling into an infestation " class," i.e., light, medium, etc.

The first system set up for the needle miner was arbitrary. Classes were called light, medium, and heavy, and were set with respect to existing populations. It has since been possible to relate infestation classes to the amount of defoliation, and new limits for needle-miner sampling by this method are being established. This will greatly increase the value of this method as a survey tool.

Spruce Budworm

Because of its complex life-history and habits, development of sampling methods for this forest defoliator has proved to be a most difficult problem. However, after years of research it has been largely resolved and is being dealt with fully in other papers now in the process of publication. Only a brief résumé of the problems involved and a few of the results will be given here.

It was found that the egg-masses of the budworm constitute a sampling unit that is satisfactory from points of view of ease of collection, examination, and statistical analysis. This was not true of the first two instars, particularly the first. In both instars there is considerable wandering and wind dispersal, and the first-instar larvæ hibernate in tiny well-concealed hibernacula. These factors do not permit direct field sampling. When counting was required, samples had to be brought into the laboratory, and the larvæ were forced to emerge. The third to sixth instars are open feeders, reasonably limited in movement. As these may be found together in the field, they are considered as one sampling stage. Similar direct sampling of pupal cases gives information on pupal mortality factors and, indirectly, on moth populations and sex ratio, since sex of the moth may be determined from empty, as well as sound, pupal cases.

The physical problems of sampling for budworm were also great. In large trees, when felling was resorted to, the loss of larvæ and particularly of pupal cases was often large enough to affect sampling accuracy. To surmount this the investigators resorted to sectional aluminium pole pruners and an extensible aluminium ladder. The use of these in sampling was described by Morris (1950).

One notable achievement is the relation of all stages of the insect to a common denominator, the mean number of individuals per 10 square feet of foliage surface. This permits direct comparisons of all stages, important in ecological studies, and also allows conversion of populations at any stage to estimates of absolute populations.

The sample branches are whole primaries, the width and length are measured, and the area estimated allowing for branch shape. Special sampling by vertical and horizontal crown sections was used to establish the distribution and variance of the population in the tree (Morris, 1949).

It is understood that a successful application of the sequential technique has been applied to the spruce budworm, based on a different distribution than that described for the needle miner.

These examples serve to illustrate the advancements made in sampling forest defoliators within the past decade. The pattern of scientific investigation has been similar in all three examples and gives promise that the problems in sampling which are fundamental to further studies of the causes of insect fluctuations may yet be resolved.

Selected References

Butcher, J. W. The problem of sampling for forest insect populations. Proc. Ent. Soc. Manitoba 7:30–34. 1951.

de Gryse, J. J. Quantitative methods in the study of forest insects. Sci. Agr. 14:477–495. 1934.

Graham, K. A model problem in insect ecology. Proc. Ent. Soc. B.C. (1951) 48:97–99. 1952.

Morris, R. F. Spruce budworm in Green River. Canada Dept. Agr., For. Biol. Div., Bi-monthly Prog. Rept. 5(4):1. 1949.

——— Techniques for population sampling in standing trees. Canada Dept. Agr., For. Biol. Div., Bi-monthly Prog. Rept. 6(6):1–2. 1950.

Prebble, M. L. Sampling methods in population studies of the European spruce sawfly, *Gilpinia hercyniæ* (Hartig), in Eastern Canada. Trans. Royal Soc. of Canada, Ser. III, Sec. V, Vol. 37:93–126. 1943.

Stark, R. W. Analysis of a population sampling method for the lodgepole needle miner in Canadian Rocky Mountain parks. Canadian Ent. 84:316–321. 1952 (*a*).

——— Sequential sampling of lodgepole needle miner. For. Chron. 28:57–60. 1952 (*b*).