INTERNATIONAL LIVESTOCK CENTRE FOR AFRICA

Sampling Procedures and Statistical Methodology used in Low Level Aerial Surveys

An Appraisal

Report to ILCA PO Box 2468

Addis Ababa, Ethiopia.

FHC Marriott and W Wint

Resource Inventory and Management Ltd

PO Box 316, Compendium House

1, Wesley St., St. Hellier

Jersey, Channel Islands, U.K.

November 1985

CONTENTS

1. AIMS AND METHODS

Low level aerial surveys are used to collect a wide variety of data. In many circumstances, they are much faster and cheaper than surveys of similar accuracy carried out on the ground, and on the other hand they can collect more detailed information than is possible with satellite observations. Data may be collected by direct observation by trained observers, by aerial photography, or both, and information may be obtained simultaneously on different features of the area. The survey may be used to estimate total numbers of animals or buildings and total areas under crops. It may also be used to produce maps showing relative densities over the area, and successive surveys may estimate changes in total populations, or changes in distribution such as animal migrations.

These various aims of an aerial survey may be, to some extent in conflict. The ideal procedure for collecting one type of data may be unsuitable for another, and when a survey has several objectives some compromise must be sought, depending on the estimates required and their relative importance. It must not be expected that the same type of survey will be ideal for all purposes.

a) Types of data.

The different types of data collected in an aerial survey present different problems in observation and interpretation.

(i) Crop areas.

Estimation of areas under different crops, or of different types of terrain, is extremely difficult from observer reports, and nearly always requires aerial photography. Observers may be able to identify different types, perhaps more reliably than is possible from a photograph, but accurate estimates of area cannot be made by eye. Such estimates may be based on a complete photographic record of the sample area, or on photographs of those parts where the crop is identified.

Measuring field areas on photographs is difficult and time-consuming, and can be seriously inaccurate. Often more accurate estimates can be based on the proportion of line transects that fall in areas under the crop. This method is particularly suited to aerial survey data. Usually, the sample consists of straight strips across the area, and the line down the middle of the strip can be regarded as a line transect across the area. The proportion of these lines corresponding to a particular crop is then an estimate of the proportion of the total area under the crop.

An alternative to line transects for this purpose is a dot count. The photographed area is sampled by a fixed number of points - usually on a regular grid, though random sampling is also possible - and the proportion of the area under the crop is estimated as the proportion of dots falling on the parts of the photograph identified as corresponding to the crop. This method was used, for example, in the Bauchi State Survey (RIM, 1984).

(ii) Counts of buildings.

Buildings tend to occur in clusters, and the variability of counts between sampling units is correspondingly high. Further, the variability is related to the mean, so that the usual sampling theory, based on the normal distribution, is suspect. The use of the t distribution to give confidence limits assumes a constant variance, and this is unjustified.

For large samples, this is not important; the central limit theorem ensures that total counts in large samples will have an approximately normal distribution. For small samples, or for mapping density, this may not be so, and confidence intervals may be unreliable. Possible modifications to the statistical procedure, especially involving transformations, will be discussed later.

Other problems involve the exact definition of a building or building type. It is important that this should be framed unambiguously, so that there is seldom doubt about whether a particular building should be included. There is also some difficulty in counting a large group of buildings, but this may be alleviated by repeated overflying and by photography. Often, the survey is concerned only with rural habitations; groups of any considerable size are not included.

(iii) Wild animals.

These present the same problems of clustering, often in an exaggerated form. Animals are clustered in herds, often varying greatly in size, and the herds themselves may be clustered. Some species are found in very large herds, and population estimates depend critically on how many of these herds are included in the sample. They are mobile, and may be disturbed by the aircraft. Counting biases are difficult to eliminate. Observer training and photography are important, but cannot compensate for animals that are unobserved because they are under cover or obscured by other animals. The problem of bias in counting will be discussed in more detail later. In general, counts of wild animals present the most difficult problems in conducting and analysing aerial surveys. Furthermore, for this type of data, other methods of sampling involve even more intractible difficulties.

(iv) Domestic animals.

The problems here are the same as for wild animals, but various features make accurate estimates rather easier. More may be known about their distribution and movements, at least if they are attached to permanent settlements rather than associated with nomadic herdsmen. There are no domestic animals where there is no human population, and a knowledge of the distribution and movements of the human population may make it easier to sample the area more efficiently for estimating domestic animal numbers. Finally, it is easier to get accurate estimates for selected areas by "ground-truthing"; ground counts of wild animals are no more reliably accurate, in general, than aerial counts.

Nevertheless, counts of domestic animals still present grave difficulties. Herd size may vary less, but herd clustering is still more of a problem, and large-scale movement may take place seasonally, or over longer periods.

b) Aims

A sample survey may be intended primarily to obtain a single estimate of a population value in the area, such as the total number of cattle or the total area under maize. The estimate should be as accurate as possible, and should include an estimate of its own accuracy. Sampling theory is mainly concerned with this sort of problem; bias should be avoided, the sample should be representative of the population, satisfactory error estimates should be available, and so on. The theory is well understood, and is set out in textbooks, such as Yates, 1981. Here it will be discussed only in the particular context of aerial survey.

A sample over an area naturally lends itself to the construction of maps. These may be general purpose maps, showing different types of terrain or husbandry, or single-feature maps, showing areas of concentration of particular species of animal or types of crop. The construction of contour maps is discussed in some detail later. The main point to come out of the discussion is the advantage of systematic samples over random samples. When a single estimate is required, there are arguments in favour of random, or stratified random, sampling - though even then systematic samples are often to be preferred. For mapping purposes, all the advantages lie with systematic samples.

Where the main purpose of a survey is to estimate changes, special considerations apply. In the first place, biases tend to cancel out. Specifically, if successive surveys each underestimate a population by 10%, the percentage change in the population will be estimated without bias; if they involve the same absolute error, the absolute change will be estimated without bias. This implies that successive surveys should be conducted in exactly the same way; an improvement in technique in the second survey may give a better estimate, but will bias the estimate of the change.

Further, the samples should, if possible, be identical, covering exactly the same sampling units. In practice, the flight-lines and the grid-squares will not coincide exactly because of navigational variation, but the units nominally the same in the two samples will at least be closely related. The estimate can then be based on the changes in those units - the method of paired comparisons - and is likely to be much more accurate as a result. This is particularly important when sampling static objects such as buildings; the advantage is much less for mobile animal populations.

These remarks illustrate the point that the best technique for answering one question may not be suitable for another. Planning a survey, or a series of surveys, must take account of the information wanted and its relative importance. Sometimes it may be possible to enlarge the scope or modify the technique in later surveys without making comparison with earlier work impossible, but care is always needed.

2. BIAS AND OTHER ERRORS.

Errors in estimates based on samples are of two types: random errors and bias. Random errors depend on the exact sampling units selected. If the sampling procedure is repeated with re-randomization, the random errors will tend to average out. In a single sample, however, there are sampling errors, and the sample should be chosen so that they are as small as possible, and so that their magnitude can be inferred from the observations. Random errors of sample estimates can be reduced by stratification, by systematic sampling, and by simply taking larger samples (Yates,1981). There are also methods of adjustment that may increase the accuracy of estimates after the sample has been taken.

An aerial survey may cover country for which accurate maps are available, or country that has also been surveyed by satellite. The information available from these is much less detailed than that from the aerial survey, but may indicate ways in which the sample selected was unrepresentative. Suppose the map makes it possible to divide the area into several different types of country, and that these are well defined so that the proportion of the area occupied by each can be accurately determined. The proportion of the sampling units falling in each type will not be exactly the same, and if the different types are associated with different densities of whatever is being estimated, errors will be introduced. If for technical reasons, such as flying problems, a particular type of habitat is underrepresented, the estimates may be biased.

The procedure for correcting them is straightforward, and exactly analogous to the analysis of stratified samples. Separate estimates are made for each type of country, and are combined by weighting according to the true proportions, as estimated from the map or satellite data. Just as in stratified sampling, the estimates of standard error and confidence intervals are based on the variability between sample units within each type. This leads, in general, to adjusted estimates with lower standard errors.

Thus,large-scale supplementary information can be used to reduce random errors. Usually, the adjustments will not be large; a sample of reasonable size, whether random or systematic, is unlikely to be seriously unrepresentative of the area. Nevertheless, it is important to be able to check the point. It is important that the supplementary information used is accurate and up-to-date; if the proportions used in the adjustment are wrong, a bias will be introduced and errors increased rather than reduced.

Sources of bias.

Bias in aerial survey has been widely studied, and the main sources are well known; for extensive accounts see Norton-Griffiths, 1978; ILCA, 1979. The main errors arise from variations in aircraft position relative to the ground, and from errors in identifying and counting objects on the ground.

Variations in aircraft height imply variation in the width of the strip observed or photographed, leading to errors in estimates. The importance of an accurate continuous check on height is now widely recognized, and provided such a record is kept, adjustment is easy. Banking can also produce errors, but they are usually of small importance when the area is covered in long transects.

Counting errors are the main source of inaccuracy in population estimates by aerial survey. Nearly all errors of this type lead to underestimation. They may be briefly listed:

Counting animals outside the sample area / omitting animals in the area.
Undercounting large groups.
Failure to observe small groups.
Missing animals under cover.
Missing animals obscured by other animals.

Observer training can greatly reduce errors in the first three categories. Further, photographs of large herds can be studied at leisure and counted much more accurately than when animals and aircraft are moving rapidly. It is then possible to make appropriate corrections. The corrections are, of course, specific to the individual observers, and a single observer may give different biases at different times. Probably the best practice, when sufficient counts corresponding to countable photographs are available, is to estimate the correction separately for each observer on each day.

Animals under cover are much more difficult to estimate. The record - whether photographic or of observer's reports - should give an indication of the type of cover, whether it could obscure animals or buildings, and what proportion of the surface is covered. This indicates the possibility of bias, but unless something is known about the probability that animals are in the open or under cover, no reliable correction is possible. It is possible to give some idea of the effect of these errors on estimates, making suitable assumptions about the proportion observed in different types of vegetation. Thus, the Gongola report (RIM,1984) shows the errors that would have arisen if the proportion observed in dense cover had been 0.5, 0.25 or 0.1. In this case, the effects were fairly small, suggesting that bias of this type could not have greatly affected the conclusions.

Again, supplementary information may be used to remove bias; in this case, it is detailed information from expensive, small-scale, ground surveys. The idea is to get an accurate count of animals in certain sample units, estimate the bias, and assume that it applies to other,similar, parts of the sample.

This is not easy. In the first place, accurate ground estimates of numbers of buildings, and probably of domestic animals, are possible, but for wild animals ground estimates may be subject to biases just as great as aerial counts. Secondly, such counts are much more expensive than aerial sampling, and can be used on only a small proportion of the sample. Even if the bias can be accurately estimated for these units, applying the same correction to other parts of the sample may give misleading results.

Where ground surveys are used to estimate and correct counting biases, it is important that they should cover all the types of country in the area studied. The application of a bias correction suitable in one area to another with different characteristics may result in even greater bias.

In conclusion, biases from variation in the height and bank of aircraft can be corrected, provided accurate height records are kept. Biases due to observer miscounting can be reduced by training, and corrected by comparison with photographs. Biases associated with objects invisible from the aircraft are much less tractable, and can be corrected only when reliable ground-truthing is possible.

3. TYPES OF SAMPLE.

Sampling units.

a). Quadrat samples.

A quadrat is a compact area, selected randomly or according to some other system, used as a sampling unit. Quadrat sampling is very widely used in botany and soil science, and in ground sampling of animal populations.

For aerial survey, however, quadrat sampling has serious disadvantages. It is difficult to identify the quadrat boundaries on the ground, and errors may arise if the prescribed area is over- or under-estimated, or if it is displaced. Further, aerial examination of a quadrat usually involves several passes; flying time spent in turning, and in moving from quadrat to quadrat, is wasted. It has occasionally been used, particularly for counting animal populations in wooded country, where the disturbance caused by repeated overflying may make reliable counting easier. Nevertheless, it is unsuitable for general aerial survey work, and will not be further discussed here.

b). Strip samples.

A strip sample is a band of constant width across the whole area sampled, or across the whole stratum in the case of stratified sampling. This type of sampling unit is the one most widely used in aerial surveys. It has obvious advantages from the point of view of efficiency; a minimum of time is spent in unproductive flying, and it is comparatively easy to fly a straight course at constant height, and to observe a strip of known constant width on the ground. Provided a check is kept on height, adjustment can be made for minor variations, as has already been explained.

The direction of flight should ideally be chosen in the light of what is known about the shape and characteristics of the area sampled. In practice, it is usually dictated by practical considerations related to landing places and type of country. If there is obviously more variability in one direction than in another, it is desirable to orient the strips in that direction, so that each strip is fairly representative of the whole area, and the variation between strips is relatively small. A long, thin area is usually better sampled in long strips, so that errors at the ends are less important.

In general, the strips will be of different length. This makes the analysis slightly more complicated than that for equal quadrats, but the appropriate statistical method (ratio sampling, or Jolly's method 2) is well known.

c). Line intercepts.

A line intercept is a line - not a strip of finite width - drawn across the whole area. Line intercept sampling is particularly useful for estimating the proportion of the area covered by different types of crops or cultivation. The proportion of the total length falling in particular

types gives, under very general assumptions, an unbiased estimate of the proportion of the total area corresponding to those types. This avoids the difficulty of trying to measure or estimate irregular areas.

Line intercept sampling can often be associated with strip sampling. If a complete photographic record of the strip is obtained, a line drawn down the middle of the strip constitutes a line intercept, and can be used for estimating coverage proportions. Similarly, if an observer notes each time the type of country or crop changes in the middle of the strip, estimation - though less accurate - is possible.

It is also possible to use line intercepts to estimate numbers of objects, provided the size distribution of the objects is known. This technique is useful in certain applications, but seems to have no advantages in aerial survey work.

d). Grid samples.

A grid sample consists of samples at points on a regular lattice. When an area is sampled in strips, observations are usually recorded on successive fixed lengths of the strip, known as grid cells. The centres of these cells lie on a regular grid - often a square grid, in which the length of the cells is equal to the distance between transects. The grid cells then constitute a systematic sample of the two-dimensional area. Samples of this sort have been widely used in various contexts, and appropriate methods of analysis have been devised. They make it possible to allow for variability along the line of flight and at right angles to it. In fact, in sampling a rectangular area, the lines constituting the first grid cells, the second grid cells, and so on, have the same status in the sample as the lines of flight.

The grid cells of an aerial survey differ from systematic point samples over an area; each cell is a rectangle, touching the neighbouring cells on the flight path, and separated from the corresponding cells on the flight paths on either side. Nevertheless, for purposes of analysis observations on the cell are regarded as representative of conditions at the centre-point, and error estimates and mapping should treat the cells in exactly the same way as a systematic point sample.

This contradicts Norton-Griffiths (1981) who writes:- "Grid cells ... are not sampling units, nor are they used in analysis ...". He does not give reasons for this statement, and it is hard to see any justification for it. Grid cells - as he agrees - do give information about spatial pattern, and this information can be used to obtain better error estimates. In particular, if the line intercepts have, unfortunately, been taken perpendicular to the main direction of variation, the error variance will be grossly overestimated if this information is not used.

In later sections methods for error estimation for systematic samples of this sort will be discussed. They are also used for the construction of maps, again regarding the grid cell as typical of the area around its centre point.

e). Accuracy of dot counts.

Suppose N points are taken in an area, and r of them fall in a certain type of country - a particular crop, for example. Then, obviously, r/N is an estimate of the proportion of the area occupied by that type. The problem is to choose N so that the proportion is estimated with known accuracy.

If the true proportion is P, and p=r/N, then the variance of p is P(1-P)/N, and this may be used to construct a confidence interval for P. There are two important cases that may be considered separately, when P is near $\frac{1}{2}$, and when P is very small (or very close to 1).

When $P=\frac{1}{2}$, the variance of p is maximum, and the appropriate 95% confidence interval is p +/- 1/ N. Thus, if N=100, the confidence interval extends 0.1 on either side of the observed proportion p. Choosing N=100 ensures that the confidence interval will be no longer than this; if P is not $\frac{1}{2}$, the interval will be rather shorter, but in fact the accuracy does not vary much. If P=0.8 or P=0.2, the interval is 0.8/ N on either side of the observed value.

In calculating the confidence interval, P is not known, and the variance is usually estimated as p(1-p)/N, replacing P by the estimate p=r/N. This approximation is quite satisfactory if P is not very small or very close to 1, and if N is reasonably large.

The variance given above is exact whatever the value of P, but when P is very small the normal approximation is invalid, and the use of p in place of P in the variance may give absurd results, unless N is very large. The estimate of P is more accurate in absolute terms, but much larger relatively - a confidence interval of length 0.2 is not very useful when the estimate of P is, say, 0.04.

In this situation, r, the number of points of the type of interest, may be regarded as a Poisson variable. Confidence intervals can then be found for the expected number, NP. These may most easily be read from a table.

r	m_1	m_2
0	-	3.69
1	0.24	5.57
2	0.62	7.22
3	1.09	8.77
4	1.62	10.24
5	2.20	11.67
6	2.81	13.06
7	3.45	14.42
8	4.12	15.76
9	4.80	17.08

95% confidence intervals for the mean of a Poisson distribution, where r is the sample count.

Notice that these intervals are not symmetrically on either side of r. In a sample of 100, a count of 1 suggests that P lies between 0.0024 and 0.0557, while the best estimate of P is 0.01.

The inverse problem, of deciding how large a sample is needed for given accuracy, is now clearer. For a count of 7, the confidence interval extends from roughly half to double the estimate. This is the sort of accuracy that can be expected if N=100 and P=0.07, or if N=1000 and P=0.007. Clearly, for small fractions reasonable accuracy demands very large samples - and that may introduce further complications because adjacent points can no longer be regarded as independent.

A related question is the probability of missing an uncommon feature altogether. This, again, can be calculated from the Poisson distribution. If the true proportion is P, the probability that it is unobserved is exp(-NP). This is 0.05 when NP=3, and 0.025 when NP=3.69. This corresponds to the first line of the table above.

All these calculations assume that the sample points are placed at random. In practice, they will generally lie on a regular grid. This will give rather greater accuracy when the areas of interest are clustered.

Sample structure.

a). Types of sample.

A naive biologist, asked whether he would prefer to estimate population properties from a random or systematic sample, would almost certainly prefer the latter. With certain reservations, he would be right. A systematic sample from a "patchy" population ensures that all parts are represented nearly in the right proportions, whereas a random sample may, by chance, fall in areas of high or low density more often than the proportion of such areas in the population would suggest. Almost certainly the systematic sample will give the more accurate estimates.

This is true of total population estimates, for example of animals or buildings, and for estimates of the total area under cultivation or under certain crops. The advantages of systematic sampling are even more marked when the chief aim of the survey is to produce a map. In this case, systematic sampling ensures that no large areas are completely missed, and that the precision of the local estimates is more or less the same throughout the area.

There is one important caveat. When there are regular periodic structures in the population, systematic sampling can be misleading. Fig 1 represents an area drained by valleys, roughly parallel and equally spaced. A systematic sample of transects parallel to these valleys could fall entirely along the ridges, or entirely in the valleys close to the streams. Such a sample could give grossly biased estimates of, for example, the proportion of forest, the prevalence of water, or the numbers of any animals tending to concentrate in different types of terrain. Further, no examination of the sample values would give any suggestion of a bias.

Once this situation is recognized, it should cause only minor problems. If maps suggest variation of this type, transects should be taken across the lines (N-S in Fig 1), or, if that is impossible, transects should be arranged so that all types of terrain are fairly represented. If it is recognized only during the survey, extra sampling may be needed. The chances of such a regular structure coinciding with the sampling pattern, and being unnoticed, are remote.

The advantages of systematic sampling may be appreciated from a simplified example (Fig 2). Consider a square area, with the Eastern half having very high density and the Western half very low density of the species to be sampled. Suppose the sample is to consist of ten transects, and consider four sampling schemes:-

a) Random transects E-W
b) Systematic transects E-W
c) Random transects N-S
d) Systematic transects N-S

Here, a) and b) may be expected to give satisfactory results of similar accuracy. c) will be wildly inaccurate, with some high and some low counts. The mean count will be critically dependent on the number of transects falling in the high-density half, and the high variability will be reflected in a high variance. d), on the other hand, will always have five transects in each half, and will be just as accurate as a) or b). If, however, the variance is estimated as if it were a random sample, the

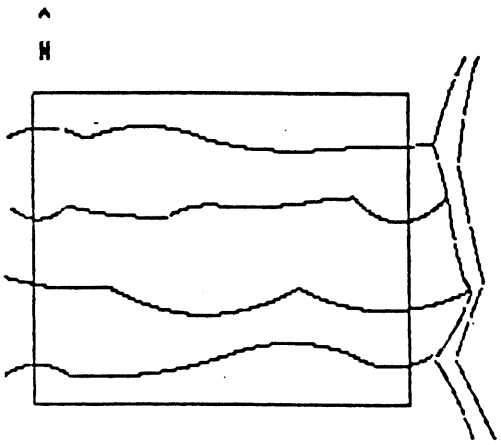

Fig. 3.1. Systematic sampling by E-W flight lines may be grossly biased.

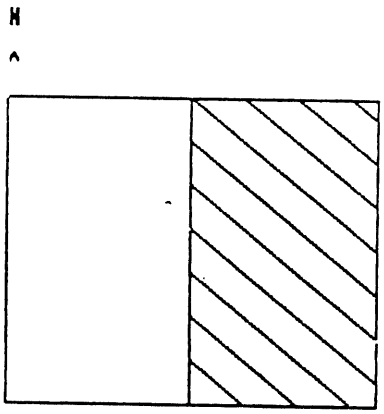

Fig 3.2. For E-W transects random and systematic samples give similar accuracy. For N-S transects, systematic sampling gives much less variable estimates.

apparent error will be just as high as c) - in fact, on average, higher.

Of course, if the situation is recognized beforehand, E-W transects would be chosen. The example, however, illustrates important properties of systematic samples:-

(i) In sampling spatially heterogeneous material, systematic samples often give much more accurate estimates.

(ii) The variability of the observations does not reflect this improvement in accuracy. In fact, precisely when systematic sampling has most advantages, the variability among observations tends to be greater than that of random samples. We have, therefore, to find a method of estimating accuracy, giving standard errors or confidence intervals, that reflects reasonably well the true reliability of the results of systematic sampling.

b) Precision and accuracy.

A distinction is often made between the precision and the accuracy of an estimate. Precision refers to the repeatibility of the result; if repeated samples, using the same technique, give closely similar results, then the precision of the method is high. Accuracy refers to the variability of the estimate about the true value. If an aerial survey counts only 10% of animals of a particular species, but counts that proportion quite consistently, it may give estimates of high precision, but obviously of very low accuracy. High accuracy implies high precision, but a biased estimate may have high precision but low accuracy.

In practice, the standard errors attached to estimates refer to the precision rather than the accuracy. Of course, biases are avoided as far as possible, or corrected by an appropriate adjustment, but the final value assumes that bias has been eliminated. A confidence interval, by definition, has a certain probability of including the true value, but the calculation cannot take account of an unrecognized bias.

The distinction is important in some contexts, but not in all. A biased estimate of total population size, or of total area under a crop, is misleading, and it must be realised that the error attached to the estimate assumes that bias has been effectively removed. Biased estimates of high precision can, however give estimates of the changes between successive samples, and maps based on such estimates can give good estimates of relative densities of animals or crops. In such cases, the bias at least tends to cancel out.

c) Random samples.

The main advantage of random samples is the simplicity of estimating the error variance. For random samples of equal area, the usual variance is unbiased.

Thus, if a sample of observations y(1) ...y(n) are taken, covering an area a of a total area A,the total of y in A is estimated as

$$Y = (A/a)\sum v.$$

Here, the values of y may represent, for example, counts of animals, or areas under a particular crop.

The variance of Y is then estimated as

$$\text{var}(Y) = (A^2/a^2)(1-a/A)ns^2$$

and the standard error S.E.(Y) is the square root of this variance.

Here, 1-a/A is the "finite population correction"; it is usually close to 1 and can often be ignored. In any case, it is not strictly applicable to estimates based on counts of moving animals.
s^2 is the estimated variance of the random variation in y.

Finally, a confidence interval may be calculated for Y.

$$Y \pm 1.96 \text{ S.E.}(Y)$$

represents a 95% confidence interval. For small samples, 1.96 is replaced by the appropriate percentage point of t.

Notice that s^2 represents the variance of the <u>random</u> part of y; that is, the deviation of the value observed from the true value of the area represented by a point chosen by the sampling rule. For random samples, this is estimated as

$$\sum (y-\bar{y})^2/(n-1),$$

but this estimate is quite inappropriate for the points of a systematic sample.

When samples are of unequal size, as for example when random transects are taken across an irregular area, the estimate and its variance are those obtained from the ratio method (Jolly's method 2). Once again, this must be modified to give reasonable estimates for systematic transects.

d) Systematic samples.

A systematic sample of an area consists of equally spaced transects across the area, or of samples centred at the points of a regular grid. Usually, the location of the sample is chosen at random; that is, the position of one of the transects, or grid points, is random, and the rest of the sample is based on that position. This means that estimates based on the sample are technically unbiased. Repeating the sample with newly randomized starting points would give unbiased estimates and an unbiased error estimate.

That, however, is an academic point. In practice, estimates and errors are based on a single systematic sample. Provided that the sort of periodic effects discussed above do not occur, estimates based on the systematic sample will be more reliable than those based on a random sample. The problem is to derive reliable error estimates from a single systematic sample.

Consider first the simplest case of systematic transects of equal length. The model underlying the sample structure is

$$y(i) = m(i) + e(i)$$

Where m(i) is the unknown true value of y(i) averaged over the part of the area represented by sample i, and e(i) is the random error. The usual formulae for random sampling then apply, if we can find an estimate of var(e).

No exact unbiased estimate is available, but it is easy to find approximate estimates by making assumptions about the form of m(i). Further, such estimates will always be <u>conservative</u> - that is, will give overestimates of variance and confidence limits that are rather too wide - provided that there are no periodic effects.

The simplest assumption is one of local linearity

$$m(i) = [\, m(i-1) + m(i+1) \,]/2$$

If this were strictly true,

$$s^2 = \Sigma[2y(i) - y(i-1) - y(i+1)]^2 \; /6n^*$$

would be an unbiased estimate of the variance of e, where $n^* = n-2$. Of course, this is a very restrictive assumption, but it may well be approximately true for most of the transects. If this is so, the error estimate will be only slightly biased upwards.

The corresponding estimate for grid samples is:-

$$s^2 = \Sigma[4y(i,j) - y(i-1,j) - y(i+1,j) - y(i,j-1) - y(i,j+1)]^2 \; /20n^*$$

where n^* is the number of points of the grid with four nearest neighbours (which depends on the size and shape of the area sampled). The variance is based on the difference between the sample value at (i,j) and the average of its four neighbours (provided they all fall in the sample). The sum of squares of these differences, with an appropriate divisor,

gives an estimate that will be unbiased if the assumption of local linearity holds, and is only slightly biased upwards in many practical situations.

This is the simplest error estimate for grid samples, based on sets of five points, and for aerial survey work, particularly when dealing with counts of high variability, it is probably the best. More elaborate estimates may be used. For squares of nine points, numbered y_1 ... y_9 by rows, one may calculate the differences:-

$$12y_5-2(y_2+y_4+y_6+y_8)-(y_1+y_3+y_7+y_9)$$

and the sum of squares of these differences, divided by $164n^*$, is an estimate of variance. Here, the assumption of local linearity is the same, but the weighting given to the four nearest neighbours and to the diagonal neighbours is arbitrary.

Yates (1981) suggests a scheme of "balanced differences" based on 16 points - a 4x4 square. The scheme of weights is shown in Fig. 3. Unlike the other arrangements discussed, this is not centred on a single square; that is not important, but the centred schemes have the advantage that a single outlying observation shows up more clearly.

All these methods lose some information because they are restricted to points with neighbours - the value of n^* may be much smaller than n if the total number of grid squares is small, or if the outline is very irregular. The simpler differencing schemes lose less information in this way, and for the type of highly variable data often collected in aerial surveys smooth, large-scale trends are uncommon.

When the value of n^* is small, say less than 20, satisfactory error estimates are impossible to obtain. This, however, is seldom the case. Error estimates are required for small sample areas when estimates of local density are to be mapped, but the variance estimate can then be based on the whole survey, or at least on the whole stratum of which the small area forms a part.

When an estimate of s^2 has been obtained, standard errors and confidence limits are calculated exactly as for random samples. Thus the S.E. of the mean of n observations is $s/n^{1/2}$ and a confidence interval can be calculated in the usual way. The only difference in the interpretation of random and systematic samples lies in the estimate of the error variance, the random variability not associated with large-scale spatial variation.

-0,25	0,5	-0,5	0,25
0,5	-1	1	-0,5
-0,5	1	-1	0,5
0,25	-0,5	0,5	-0,25

Fig. 3.3. Yates' "balanced difference" scheme for a square of 16 points. The appropriate divisor - the sum of squares of coefficients - is 6.25.

e) Transformations.

In the analysis of experimental data, transformations are widely used; in sampling theory scarcely ever. The reason is clear. A main objective of sampling is to obtain an estimate that is a simple linear function of the observations, and transformations introduce bias. In work on experimental analysis, something has been done on obtaining nearly unbiased estimates from means of transformed data (Quenouille, 1955), but in sampling theory this is rather pointless. The main effect of transformation is to reduce the influence of extreme observations, such as a count of a large herd of animals, and in estimating animal populations this is certainly not desirable.

There are two aspects of the sampling problem in which transformations may be useful, especially in the context of estimating animal populations from systematic aerial samples. In the first place, the assumption of local linearity, used to eliminate (or reduce) the systematic element in the variation of y, may be more nearly valid for suitably transformed counts. Secondly, the assumption of normality used in obtaining confidence intervals for the final estimate may be unrealistic; usually y and the estimate Y will both be skewed to the right, and unless samples are very large a symmetrical confidence interval may be misleading.

The transformation generally used for counts of this type is

$$z = \log(a + y)$$

Here, log may represent either the common or natural logarithm, and a is a small constant, necessary for the transformation of zero counts. Usually, a is taken as 1, but any positive value may be chosen, and it may be possible to stabilize the variance more effectively by a different choice.

Sometimes a square root transformation is used for counts, but it is appropriate only where the error is of the Poisson form, with the variance proportional to the mean. This is certainly not true for counts of gregarious animals, and the transformation does not provide such simple confidence intervals as the log form.

For observations in the form of proportions, the arc sine transformation, $z = \arcsin(y^{\frac{1}{2}})$ is sometimes used. This is theoretically justified when the observations follow a binomial distribution; that is, when the proportions are counts of independent events out of a fixed number of trials. Under these circumstances, the transformation has a variance-stabilizing effect. There is no real justification for using this transformation merely because the observations are in the form of proportions, though in the analysis of experimental results it is quite often used for such data on a purely ad hoc basis. In sampling theory, it has the disadvantages of all non-linear transformations, and there seems to be no case for ever using it.

In systematic sampling, the variance is estimated from the deviations of interior points from the average of neighbouring points. In the previous section, it was suggested that this average should be a simple arithmetic mean of the four nearest neighbours, or a weighted

arithmetic mean of the nearest neighbours and diagonal neighbours. There is no reason, however, why some non-linear average should not be used. Consider, for example, the following scheme:-

(i) For each point (i,j) with four nearest neighbours, take the mean value of $z = \log(a + y)$ for the four nearest neighbours.

(ii) Transform this average back by calculating $Y = \exp(\bar{z})+k$, where k is chosen so that the mean values of y and Y, taken over the n* interior points, are the same.

(iii) Estimate the variance by

$$s^2 = \Sigma(y-Y)^2/(n^*-1)$$

This gives an estimate of variance that differs from the one suggested in the last section only in the way in which the "predicted value" Y is calculated. (The value of k makes suitable allowance for the bias in the geometric mean of y+a compared with the arithmetic mean, and the divisor n*-1 is used because the means of y and Y have been made exactly equal). The estimate will be better if it is more effective in eliminating the systematic component of the variability. Roughly speaking, this will be the case if the transformed variable $z = \log(y+a)$ varies more smoothly over the sample area than does the original variable y.

The log transformation is only one of many that could be used in this way. Provided there are no periodic effects, all the estimates of variance are biased downwards, because they do not remove the systematic component of variance completely. In large samples, it is reasonable to conclude that the smallest variance estimate is the best, since it removes the systematic component most effectively. In smaller samples, however, there is a danger of introducing a downward bias by selecting a method that gives a low estimate merely by chance.

A second application of transformations in sampling theory is in the calculation of confidence intervals. If the variable being studied has a very skew distribution - as have counts of animals or dwellings, as a rule - the usual confidence interval for the mean may be misleading. It is true, of course, that in sufficiently large samples the distribution of the mean approaches the normal distribution, but for data of this sort it may do so very slowly. The 95% confidence interval may then have a true probability different from 0.95 of including the true mean, and in particular the probabilities of lying entirely above and entirely below the true value, nominally both $2\frac{1}{2}\%$, may be quite different from each other. The calculated confidence interval is then no longer a "central" confidence interval.

The situation may be improved by a transformation. If x is an estimate of a parameter, a confidence interval for any function of x, say f(x), has the appropriate probability of including the same function of the parameter. Thus we can calculate a confidence interval for f(x), and assert that the parameter lies between the inversetransformatio of the ends of the interval. Further, if f(x) is more nearly normally distributed than x, the confidence interval will be more reliable.

Consider, for example, the function $f(x) = \log_e(x)$, where x is a sample mean with S.E. $s/n^{\frac{1}{2}}$. Then it is easy to show that the standard error of f(x) is $s/xn^{\frac{1}{2}}$. Confidence limits for f(x) are obtained as f(x) 1.96 times this S.E. (note that there is no justification for using t in this case), and this may be transformed back to a confidence interval for x by taking exponentials.

Alternatively, the power transformation $f(x) = x^w$ has S.E. $wx^{w-1}s/n^{\frac{1}{2}}$, and a confidence interval may be constructed for f(x) and transformed back by raising to the power -w. This, again, gives an approximate confidence interval for x, and if the distribution of f(x) is nearer to a normal distribution than that of x, the approximation is likely to be better, giving a size closer to the nominal 95%, or whatever other level is chosen, and a more nearly equal probability of lying completely above or below the true value.

Experience suggests that when the data are counts, confidence intervals are much better approximations when transformations of this type are used. For the typical highly skewed distributions, where a log transformation is used in experimental work for the data, a log transformation for the mean is useful, but a small negative power transformation, say $w=-\frac{1}{2}$, is probably better.

The proposals in this section are tentative. More research is needed, on theoretical aspects and on the practical uses of these transformations. It is clear, however, that it is possible to improve upon standard estimation techniques when data are grossly non-normal.

f) Stratification.

The subdivision of a population into groups, which are then sampled, is known as stratification. Each group, or stratum, is chosen to be as homogeneous as possible, and to differ as much as possible from the others. In the present context, this means that the area of the survey is subdivided, and each subdivision is then sampled, using random or systematic sampling. The estimates from each of the sections are then combined to give an estimate for the whole area. This has two main advantages:-

(i) If the total areas of the strata are known, the variability of estimates depends on the variability within the strata, and if these have been sucessfully chosen to be homogeneous, this may be much smaller than the variability of the whole population. Under these circumstances, a stratified random sample gives much more accurate results than an unstratified random sample.

(ii) The sampling fraction can be chosen to be different in different strata, and this may give more accurate estimates for the same cost of sampling. The optimum sampling fraction is proportional to variance within the stratum, and inversely proportional to cost.

In aerial survey work, stratification is only worth while when a very large area is sampled, and it is easily divisible into regions that are very different one from another. The first advantage refers specifically to the comparison of random and stratified random samples; it does not apply to systematic samples. If random sampling is chosen, however, the advantages can be very great.

One possibility that is often mentioned here is that of post hoc stratification. Sometimes it becomes obvious only after the sample has been taken that some areas have high density and some low density of the character being studied. It may be easy to break the area up into high and low density strata, basing the division on the observations, not on previous knowledge of the terrain. Analysis of the results as if they constituted a genuine stratified random sample may then give much lower error estimates.

Is this legitimate? It should be obvious that it can lead to an underestimate of error; simply dividing observations into different ranges and estimating error as if these were strata is clearly wrong. Further, the exact divisions between these "strata" are arbitrary, and so their areas are not known exactly.

Nevertheless, the technique should not be rejected out of hand. When it is clear that there is large-scale variation of this sort, post hoc stratification may give better estimates, and more reasonable variances, than the usual treatment of a random sample. It is important, though, that the strata are large and compact; subdivision of a large-scale survey into blocks of a hundred or so observations is unlikely to be misleading, but finer subdivision could well be.

The second advantage applies to systematic as well as to random samples. In counting objects, the main difference in variability is associated with differences in abundance. The same is true of estimating proportions of area, at least when the proportions are small. The point is therefore not to waste effort in intensive surveys of deserted areas. Further, if some areas are difficult (more costly) to sample efficiently, they should be sampled less intensively.

It must be said, however, that the advantage is very slight unless differences among strata are extreme. A constant sampling fraction is appropriate unless heterogeneity is extreme.

g) Fitting distributions.

A problem of some theoretical interest is that of fitting statistical distributions to the observations. This cannot lead to better estimates of totals or mean values, nor can it improve the quality of maps; it may, however, give alternative error estimates based on the distributions fitted, and it may throw some light on the factors affecting the distributions. Assefa and Bille (SRDU unpublished report) have studied histograms of counts of animals and dwellings, and attempted to fit smooth curves to the histograms.

The fitting of distributions usually assumes that the observations in the sample can be regarded as independent. This assumption is obviously unlikely to be valid for data of this sort; there is environmental variation over the area studied, and adjacent observations are likely to be positively correlated. If a distribution is found to fit the observations, there remains some doubt about how it should be interpreted.

If the objects counted behave independently, and if the area sampled can be regarded as uniform, a Poisson distribution of counts would be expected. Neither of these conditions are likely to apply, and the resulting distribution is likely to show much greater variability than the Poisson distribution (which has the variance equal to the mean).

Discrete distributions for counts of this sort have been suggested; they fall in the general category of "contagious" distributions, of which the most familiar are the compound Poisson distributions. These distributions are derived from Poisson distributions in which the mean value itself varies according to some known distribution. Thus, if objects occur randomly but the average density varies over the area sampled, one might expect a distribution of this type.

The best known of these compound distributions is the negative binomial distribution. The mean of the Poisson distribution is supposed to follow a gamma distribution - the type to which the chi-squared distribution belongs - and this gives a distribution with two parameters in which the variance is greater than the mean. This distribution has been fitted to a great variety of data of different types, such as accident data, disease incidence and plant counts.

It is tempting to fit such a distribution to counts, and then to try to interpret the underlying gamma distribution in terms of environmental heterogeneity, but this ignores the possible correlations between observations and the geographical features of the variation. There is a further difficulty; the distance between sample units and their size will affect the distribution in ways that are not easy to predict. Typically, combining observations into groups of, say, four will give different parameters, and interpretation will be complicated.

Assefa and Bille have realised these difficulties and tried to circumvent them, but we feel they have not altogether succeeded. In fact, we are not convinced that there is any useful purpose in fitting distributions to data of this sort. The spatial variation, the random variation and the tendency to group are all combining to produce the final set of data, and fitting a distribution does not really help to sort them out.

h) Precision of systematic samples.

In general, systematic samples may be expected to give higher precision than random samples of the same area. The improvement will be greatest for large, highly heterogeneous, areas; a systematic sample ensures that all parts of the area are represented in the sample, in nearly the right proportions, while this balance may well not hold in a random sample. It is difficult to judge, however, how much is gained by taking systematic samples. If a systematic sample is available, an estimate of variance may be based on an appropriate differencing method, as described. This estimate is in general biased upwards, so that calculations based on it are conservative; with a suitable choice of estimate, the bias should be fairly small. A systematic sample gives no direct information about the precision that would be given by random sampling. Precisely because the systematic sample ensures that all parts are fairly represented, the variance among the observations of a systematic sample - treated as if they had been taken at random - tends to overestimate the variance associated with random sampling.

Of course, if a number of systematic samples and a number of random samples were taken of the same area at the same time, a direct comparison of variability would be possible. This might be an interesting exercise in simulation, but is certainly not a practical proposition. Any attempt to estimate the practical gain in precision must be based on single systematic samples. For small samples, with data as variable as counts of animals or dwellings typically are, no such comparison is of much value. For sufficiently large samples, however, a comparison of the variance estimates for a systematic sample calculated by the differencing method and calculated as if the sample were random gives at least some estimate of the relative precision of random and systematic sampling. It must be realised that both estimates are biased, but the biases are in the same direction, and for very large samples are relatively small.

The following table gives, for ten surveys taken over four large areas, the standard error as a percentage of the mean, based on Jolly's method 2 - which is appropriate for a sample based on random transects - and differences from a four-point moving average, as described above. In all cases, the survey was based on a systematic sample of parallel and equally spaced flight lines, and the individual observations corresponded to equal section of those lines, centred on a square grid.

For each survey, the data were counts of cattle, and the estimated mean density , also shown in the table, varies by a factor of ten. Gongola has an area of 43,875 square kilometres, Gourma 83,100 square kilometres, and Niger 81,155 square kilometres.

In every case, with one exception, the differencing method gives a lower estimate of the standard error, usually by a factor of about 1.4. This means that the sample size to attain the same precision would be twice as large for a random sample as for a systematic sample. In Mali, where the average counts are higher, the advantage is even greater.

If we can accept that these figures are a fair comparison of

precision, and that these eight large surveys are reasonably typical of the results to be expected from large-scale counts of cattle, the conclusion is obvious. We may expect the same precision from a systematic sample and from one based on random flight lines if the latter is twice as large. In practice the advantage will sometimes greater and sometimes less, but the figures leave no doubt that systematic sampling has very large advantages for data of this sort.

Location/season		%S.E.(Jolly)	%S.E.(Diff.)	Density/km^2.
Gongola	dry	9.6	5.4	13.6
	wet	6.9	4.28	15.7
Gowma	dry	10.0	6.75	4.13
	wet	6.9	6.92	5.40
Niger	May	7.5	5.62	3.42
	Oct	6.0	5.63	4.49
	Sept	9.7	7.50	4.00
Mali	Oct	15.7	8.05	22.81
	Mar	14.2	7.65	36.42
	Jun	11.0	7.71	22.77

Table 3.1. A comparison of the precision of random and systematic sampling. See text.

4.CONTOURING AND SMOOTHING.

a) Contouring and smoothing.

One of the aims of an aerial survey may be to produce a map showing the relative abundance of particular animals, crops or types of building over an area. Such a map may be presented in various ways, but all depend essentially on some process equivalent to drawing contours of equal density.

It is convenient to divide the process into two stages, which are conceptually different, though they may be performed in a single step. Interpolation consists in estimating values at points where actual observations are not available - usually some sort of weighted mean of the nearest observations. Smoothing involves adjusting the actual observations to reflect, as far as possible, the underlying trend, while reducing the random variation in the observation.

In what follows, we shall assume that observations are made at a point (whereas they really correspond to an area centred on that point), and that the points lie on a regular square or rectangular grid. The first assumption greatly simplifies the discussion without introducing any error as large as that inherent in the sampling. The second restricts us to systematic samples; in fact, all the methods discussed can be applied to random observations, but the arithmetic is more difficult and the accuracy of the map varies according to the position of the samples.

Interpolation methods were developed for use with mathematical tables (see, for example, Pearson, 1920). The method is essentially one of fitting a polynomial to a grid of points surrounding the point to be estimated. When we are concerned with observations subject to random error, elaborate techniques are inappropriate; usually the best method is to fit a plane (by least squares) to the four points at the corners of the rectangle in which the point is situated. Thus, given values over a grid, it is easy to estimate the value at any point within the grid.

Contouring uses the interpolation procedure inversely, to find all points for which the estimates have a particular value. These form smooth curves within any grid square, and computer programs are available, with varying degrees of sophistication, to produce smooth contour lines over the whole area.

Obviously, interpolation by itself is unsuitable for data from a spatial sample. The fact that a herd of animals is seen at one point on a particular day does not mean that the prevalence of animals is very high in that immediate neighbourhood, and falls to zero in a neighbouring square in which no animals were seen. There is random variation in the sampling, and, in the case of animals, random day-to-day variation. Some smoothing is necessary, and it involves a compromise depending on the random variation present. Undersmoothing will produce a highly irregular map, with spurious fluctuations that are purely random; undersmoothing will produce an over-simplified map in which real features are obscured.

b) Moving averages.

The simplest method of smoothing, developed in connection with time series analysis, is the use of moving averages. The observed value at each point is replaced by a weighted average of the value itself and the observations at nearby points. As an example, one might take half the actual value, plus one-eighth the sum of its four nearest neighbours.

The main problem is to decide how many points to include in the moving average, and what weights to give them. These decisions determine the degree of smoothing, and must be made subjectively. An experienced analyst can probably make a good guess, based on his knowledge of the type of random variation present - but it remains a guess.

A second point is that moving averages necessarily tend to flatten out real maximum and minimum values. The process reduces the random errors, but also tends to flatten out the underlying trend.

Nevertheless, moving averages provide a simple and easily understood method of smoothing. Used with discretion, they may produce a reasonable set of modified values at the points of the grid, and these can then be used as the basis for contouring to give a satisfactory map.

c) Trend surfaces.

Another straightforward way of smoothing data is to fit a trend surface. Each point of the grid is represented by Cartesian coordinates x,y , and the observed values are fitted to some function of x and y using regression methods. The resulting surface fits as closely as possible to the observations in the least squares sense. Further, no special contouring programme is needed - the fitted function has a value at every point of the area. Smoothing and interpolation are combined in a single process.

Unfortunately, the types of function that can be easily fitted by these methods are very limited. Polynomials are the obvious choice, and as far as I know the only functions to have been used in practice. They have grave drawbacks. In the first place, the number of parameters rises rapidly with the order of the polynomial. The general polynomial of order k has $(k+1)(k+2)/2$ coefficients, and in practice that restricts the value of k to a maximum of three or four, even for very large data sets. Even polynomials of this order have a restricted number of maxima and minima, and cannot adequately describe "patchy" distributions, in which abundance depends on local soil-type or water-supply.

Further, polynomials can produce anomalous values. In particular, if there are several adjacent zero observations, it is very probable that the fitted polynomial will become negative. Small negative values can simply be ignored, but are disconcerting. They can be avoided by a suitable transformation - for example, by fitting a polynomial to $z = \log(1 + x)$, and then contouring $\exp(z) - 1$.

In conclusion, polynomial trend surfaces work well only when the underlying trend is simple, having, say, a linear or parabolic form in all directions. The functions cannot give good fits to irregular, patchy, data. This problem is well known in the one-dimensional case, and various attempts have been made to solve it. The most promising of these is the use of splines (Silverman, 1984). This is a technique for fitting polynomials, usually cubics, to the data, but the fit is <u>local</u> - the fitted cubics are different in different parts of the range, but are constrained to fit smoothly together. The approach is much more flexible, and avoids most of the drawbacks of polynomial regression. Unfortunately, the calculations are much less easy, and the technique has not been extended to the two-dimensional case.

d) Kriging

The most elaborate method of constructing contour maps is known as "kriging". It was devised in the context of mining, and has been used particularly in the search for oil. Many of the terms used reflect this origin. It is basically a method of estimation using weighted least squares, based on the observed similarities among points separated by small distances. In fact, estimated values are weighted means of nearby points, but the method differs from the simpler moving average and contouring approaches in that the weights are based on observed relationships, rather than subjective judgment. The subjective element, however, is not removed altogether; decisions must be made about the degree of complexity of the model to be fitted, and these can have a considerable effect on the eventual map.

The fundamental tool of kriging is the semivariogram. This is a plot representing the extent to which points close together in space tend to have similar values. The abscissa is the distance between pairs of points; the ordinate is half the mean squared difference between observations a distance x apart. Typically, one might expect a semivariogram estimated from points on a square grid to look like Fig. 1. The x values for which there are observations correspond to the distances between pairs of points on a square grid. When $x = 0$, the semivariance is zero, by definition. As x increases, the semivariance increases, eventually tending to a limit. This limit corresponds to the variance of an observation at a randomly chosen point in the area studied.

The semivariogram, in fact is very closely related to the autocorrelogram widely used in time series analysis. As the autocorrelogram falls from 1 at zero lag to zero for long lags, so the semivariogram rises from zero to its limiting value.

Several points must be noted here. In the first place, we are assuming that the observations constitute a "stationary" series. If there is an obvious trend, the semiovariogram, defined as we have done disregarding the trend, has little meaning. Secondly, the observations are supposed to be taken at points; if, in fact, they are values for areas, the distances between them are not simply defined, and the variances are, at best, an average over a range of distances. Thirdly, it is assumed that the semiovariogram depends on distance only, and that this dependence is the same in all directions. In some circumstances, the semiovariograms for N-S and E-W distances might be quite different. Finally, if the sample points are random, each x value will occur only once, and average values at fixed distances cannot be calculated. All these complications are discussed in books and papers about the method, and suitable modifications to the calculations have been worked out, but for the present we shall ignore them.

The semivariogram is then used to find appropriate weightings to estimate the value of the observed variable at any point in the area. Since the semivariogram drops to zero at $x = 0$, the values at the grid points themselves are unchanged. In fact, kriging using a semivariogram like that in Fig. 1 is a method of <u>interpolation</u>, without any smoothing. This is clearly unsuitable for the present purpose.

This problem has long been recognized in geostatistics. Semivariograms in practice often look more like Fig. 2 than Fig. 1. The value at x = 0 is necessarily zero, but the values for higher values of x seem to lie on a smooth curve that is dropping towards a value greater than zero. There are presumably two sources of variation, one very short range in comparison with the distances between observations, and the other a smooth variation over longer ranges. In geostatics, this is known as the "nugget effect", representing very small areas of high concentration, such as gold nuggets. In the context of spatial sampling, the nugget effect is related to the random errors of sampling, while the rest of the semivariogram represents, mainly, the systematic variation over the area.

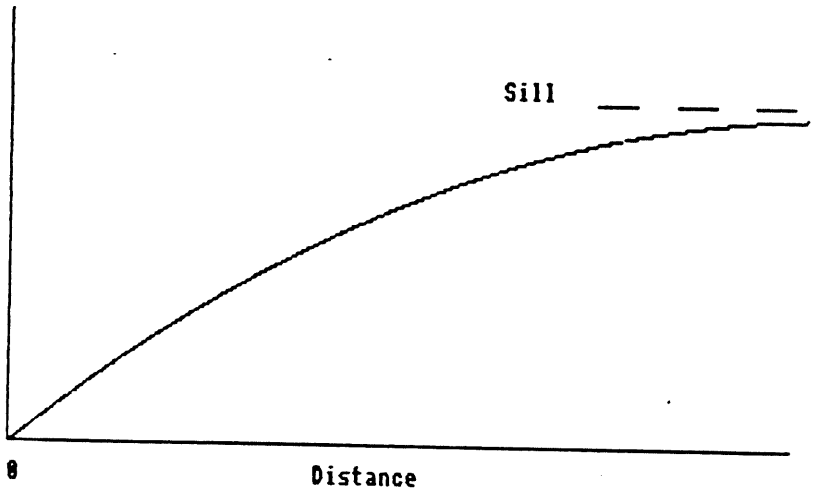

Fig. 4.1. A semivariogram; the half-variance of differences rises from zero to a limiting value or "sill".

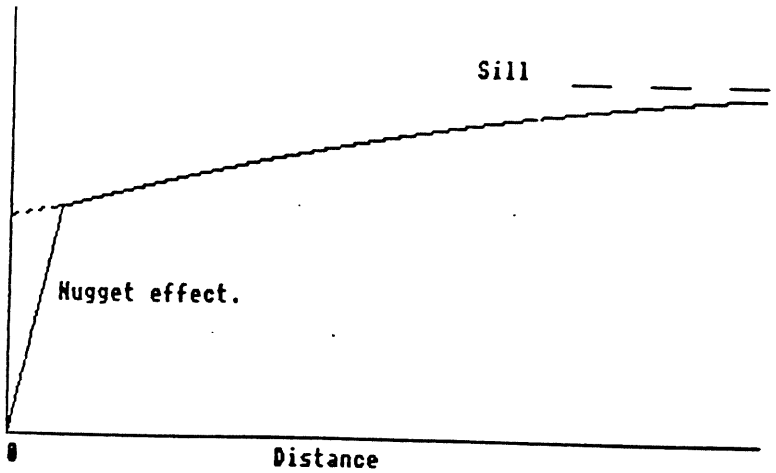

Fig. 4.2. A semivariogram showing the "nugget effect". The curve seems to drop steadily from the sill towards a non-zero value at zero distance, though there must be a final sharp drop.

The next step in the estimation process is to fit a suitable curve to the semivariogram. When there is a clear limiting value, a "sill", as in Fig. 1, one of two models is usually chosen:-

The exponential model

$$y = a(1 - \exp(-kx))$$

or the spherical model

$$y = 3x/2a - x^3/2a^3, \quad x<a$$

$$y = a, \quad x>a.$$

Both models have a sill at $y = a$; in the exponential model the sill is approached as x tends to infinity, in the spherical model, it is reached at $x = a$. There is little practical difference between the two models, and they usually produce nearly identical estimates.

When a nugget effect is present, as in Fig. 2, the same models can be used, modified so that $y = b$ when $x = 0$. Thus they become:-

$$y = b + a(1 - \exp(-kx))(1 - b/a)$$

and $$y = b + (3x/2a - x^3/2a^3)(1 - b/a), \quad x<a$$

$$y = a, \quad x>a$$

respectively.

These models can be fitted to the observed values of the semivariance. When there is a nugget effect, the three-parameter models are fitted to the semivariances for $x \div 0$; the nugget effect represented by b is thus inferred from the trend of the semivariances ignoring the zero value. Estimation using this type of semivariogram involves smoothing as well as interpolation; the fitted surface no longer passes through the observations.

Thus kriging provides a method of contouring in which the degree of smoothing is objectively determined from the semivariogram. It must be realised, however, that the critical value b, which determines the degree of smoothing, may not be very accurately estimated; the observed semivariogram may be reasonably consistent with quite a large range of values of b. The degree of smoothing is no longer a matter of guesswork, but there is no guarantee that the fitted surface will accurately reflect underlying trends with the random variation smoothed out.

The procedure of producing a fitted surface from the observations and the semivariogram is arithmetically complex, and requires a computer program. In principle, values are fitted by least squares, the weights being determined from the semivariogram. The fitting process is iterative, and involves quite heavy computation. Strictly, each fitted value is a weighted mean of all the observations; in practice, however, the value depends only on the nearby points. The weights attached to points further off are so small as to be negligible. This "screening effect" means that the values produced are usually similar to those given by moving averages and contouring.

e) Problems in kriging.

The previous section describes the simplest form of kriging. This involves, as has been mentioned. strong assumptions that are often

unrealistic in practice. Many modifications have been worked out to allow for these complications. They are available in the sophisticated computer programs available for kriging, and are fully described in textbooks (see, for example, Journel and Huijbrechts, 1978). In this section, they will be mentioned very briefly.

(i) Anisotropy.

The form of the semiovariogram may depend not only on distance, but on the direction in which the distance is measured. In geostatistics, this is typically related to the geological formation. The correlations between neighbouring points may not be the same in the N - S and E - W directions. In aerial surveys, the same thing may happen as a result of features of the terrain. More importantly, however, there is no reason to suppose that the correlation between neighbouring grid squares on the same transect will be the same as that between squares on adjacent transects, even though their mid-points are the same distance apart.

It may therefore be necessary to estimate separate semivariograms for the direction of transects and the direction at right-angles to it. This is easily accomodated in kriging programs, but of course the semivariograms will be based on fewer points and accuracy will be lost if in fact there is homogeneity.

(ii) Random samples.

If sampling is random, rather than systematic, the estimation of the semivariogram is more difficult, to achieve. Again, however, the programs can handle the problem, and produce appropriate maps. The main disadvantage is the variable accuracy over the area. A map of variances of estimates for a systematic sample will give relatively constant values over the area, with some loss of accuracy near the edges. If sampling is random there may be parts of the area that are relatively undersampled, and estimates for these parts will be correspondingly inaccurate. When the main purpose of the survey is to produce a map, rather than a single estimate of a total population, the advantages of systematic sampling are even greater.

(iii) Fitting trends.

Often the assumption of stationarity is unacceptable. There may be an obvious trend present, of a fairly simple form, as opposed to unstructured patchiness. In this case, it is possible to fit simultaneously a trend surface and a superimposed stationary model. The process is known as "universal kriging", as opposed to "simple kriging.

This widens the choice of models available. There is an analogy with the problems encountered in time series; a model may be fitted more "parsimoniously" with a simple trend and a simple correlation structure than with either separately. Again, universal kriging is an option available in kriging programs.

(iv) Transformations.

The values used in a kriging program need not, of course, be the raw observations. They may be transformed, and the map may thus be based on transformed values, with the only difference that the contours must be appropriately labelled. Indeed, as the method is based on least squares, a transformation will usually be appropriate when the random error is expected to be related to the mean. Thus, for counts of clustered objects, the usual logarithmic transformation will often be appropriate.

f) Conclusions.

Kriging was developed in the context of mining engineering. Here, the cost of sampling is extremely high, observations are usually of high accuracy, and the rewards from a reliable map are very great. Consequently, no effort is spared in carrying out the best possible statistical analysis of the data. This involves sophisticated computer programs, and considerable skill is required in using them. The resulting map is obtained objectively, in the sense that once decisions have been made about the type of semiovariogram, the type of trend, anisotropy, and so on, the computer does the rest. The outcome is a map of estimated values over the area, with a map of estimated accuracy. There is no doubt that it is the best method available for handling this type of data.

In aerial survey work, observations are as a rule subject to much greater random error. Sometimes the main purpose of the survey is to produce overall estimates, and a map is a by-product, intended merely to give a general impression of the location of areas of high density. The nugget effect, generally a minor irritation in the original context, here becomes a major concern - the degree of smoothing is here paramount.

In these circumstances, the the additional complication and expense of kriging may not be judged worth while. Simpler subjective methods may be quite adequate for the purpose, and the analyst may prefer his subjective judgment to computer output based on possibly dubious assumptions.

5. OTHER STATISTICAL TECHNIQUES.

Survey work is mainly involved with overall estimates of numbers or areas, and with local estimates used for producing maps. The corresponding statistical techniques form the main subject of this report. Survey data, however, is sometimes analysed by other methods, and this section discusses briefly some possible applications of such methods to clarify the interpretation of survey data.

In fact, statistical analysis often stops short at giving estimates and standard errors. More detailed analysis is perhaps felt to be unnecessary, or unduly complicated. Nevertheless, an aerial survey generates a considerable mass of multivariate data, at considerable expense; if more elaborate statistical analysis can extract further information from it, the extra effort is certainly worth while.

The main fields of application are regression analysis and multivariate analysis. In both, the grid squares are regarded as independent observations - that is, their spatial coordinates and any effects of spatial correlation are ignored in the analysis - on a number of possibly related variables. This assumption of independence is not strictly justified, and significance tests and confidence intervals are invalidated; these calculations should be regarded as exploratory data analysis, designed to generate, not test, hypotheses.

(a). Multiple regression.

The aim of multiple regression is to predict a random variable y as a linear function of a set of "regressor variables" $x_1 \ldots x_p$. The "dependent variable" y might be, for example, the number of cattle counted on a grid square, and the x-variables might represent vegetation conditions, distance to water, number of dwellings, and so on counted on the same square. The reference to a linear function may seem rather restrictive, but is not so really; the x-variables need not be the original observations, but variables constructed from them. Thus, they might include log counts rather than raw counts, vegetation indices rather than actual percentages of different vegetation types, and dummy variables indicating presence or absence of a character rather than a quantitative variable.

There is no assumption involved about the distribution of the x's. Multiple regression is not a multivariate technique; y is the only variable whose distribution is of interest, and the underlying assumption is that y is distributed with constant variance about the value predicted from the x's. A transformation is often necessary to satisfy, at least approximately, the constant variance condition - a logarithmic transformation is usually appropriate for counts.

In the present context, the aim of multiple regression is not so much to obtain an accurate prediction of y, as to investigate which of the x variables are most closely associated with it. This presents an intractable problem. Often the x variables are themselves closely correlated, and quite different subsets of them may work equally well in predicting y, yet not give any improvement in prediction when they are combined. This is, in the jargon of the economists, the problem of

"multicollinearity".

The most usual way of dealing with multicollinearity is to select a subset of the x variables that gives an adequate prediction of y - that is, one effectively as good as could be obtained with all the x's - and ignore the remaining x's. There may be several such subsets; if two of the x's are highly correlated, there is no way of telling, on statistical grounds, which one is the better predictor. That could only be determined from samples in in other areas in which the regressor variables were less highly correlated.

The usual method of selecting regressor variables is "stepwise". First, the x that correlates most strongly with y is selected, and y is adjusted by regression on that variable. The next x variable selected is the one that correlates best with the adjusted y, the part of y uncorrelated with the first x. This process continues until none of the remaining x's is strongly correlated with the adjusted y, and the x's then in the equation are accepted as a suitable subset for the prediction of y.

This procedure is described in detail in many computer package manuals - see, for example, the BMDP manual (Dixon, 1983). The theory is given in Draper and Smith, 1981. A good example of the application of stepwise multiple regression to aerial survey data is given in Milligan,1983, which investigates the environmental conditions that most affect density of livestock of different types.

In conclusion, multiple regression is a useful technique for investigating relationships among the variables observed in aerial survey work. It is important to realise, however, that relationships obtained, whether by stepwise methods or otherwise, are not, in general, unique. Multicollinearity is almost always a problem, and when it is present more than one interpretation is possible. This is particularly important in comparing relationships found in different surveys; very similar data sets may give quite different sets of regressor variables, without any indication of real differences in the factors that are important. Considerable skill and experience are needed for the interpretation of the results.

(b) Principal components.

When data are collected on two variables, they can be usefully represented in a scatter diagram. With more than two variables, this sort of representation becomes increasingly difficult. Three variables may be represented in a three-dimensional model, and further variables can be included by using different symbols, or different colours, for different parts of their ranges, but very soon diagrams become confusing and the point is lost.

Multivariate data are often highly correlated, and it is reasonable to hope that the information in a data set may reasonably be represented in a smaller number of dimensions. A projection of the high-dimensional plot onto a space of lower dimensions may lose very little of real value, while reducing the random variability; or "noise".

The most important technique for representing multivariate data in a space of lower dimensionality is principal component analysis. Suppose there are p variables observed. These are replaced by p new variables, linear combinations of the original set. The first of these variables is chosen as the linear combination that has maximum variance, subject to a constraint on the coefficients, usually that their sum of squares is unity. The second is uncorrelated with the first, and has maximum variance subject to that condition, and so on. There are altogether p of these new variables (unless the original observations are strictly linearly related), they are all uncorrelated, and their variances decrease. These are the principal components.

The idea of the analysis is that the first few principal components may account for most of the variation in the data, and the rest may be discarded without much loss of information. In fact, the sum of the variances of the principal components is the same as that of the original variables, and so it is possible to calculate the percentage of the total variance contained in the first two or three components; if the variances fall rapidly, the analysis is probably successful in reducing dimensionality without much sacrifice.

It is important to realise that the procedure is not scale-independent. In fact, it is quite pointless to carry out this procedure on unstandardized variables unless they are all of the same type - for example, all proportions with similar variability, or all logarithmically transformed counts. Before calculating principal components, it is nearly always necessary to standardize by dividing the original variables by their standard deviations; this is equivalent to carrying out the analysis on the correlation matrix, instead of the matrix of variances and covariances. Further, variables should be transformed before standardization if the distribution is very skew; again, the log transformation is indicated for counts.

There is no simple rule about the number of components to retain and the number to discard. Usually, those with variance less than 1 (working with the correlation matrix) can safely be ignored, and usually it is unnecessary to consider more than three or four, but the technique is purely empirical and there is no underlying mathematical theory.

Each observation on the original set of variables is now replaced by a set of values of the first, second ... principal component. Plotting principal component values in pairs may cast some light on the structure of the observations. For example, they may fall into fairly obvious groups, and, particularly if the groups fall into compact clusters on the ground, they may suggest a clustering into different types of terrain. Further, the coefficients of the variables in the principal components may suggest an interpretation. Thus the first principal component may have positive coefficients for numbers of animals, number of dwellings, nearness to water, and vegetation cover, and may be interpreted as a contrast between rich, populous, land and poor land unable to support much life. The other principal components must then be uncorrelated with the first, and may contrast different types of land management.

Mapping the values of principal components may be interesting. Because they tend to consist of sums of correlated variables, they may have rather smoother contours than similar maps of individual variables, and if they have a natural interpretation the maps may be very informative.

The computational details of principal component analysis are given in all textbooks of multivariate analysis, and the method is implemented in all the standard computer packages. Webster (1977) has an excellent account of their use in soil science, including their use in preparing soil maps - a section that is closely applicable to aerial survey data. Principal components may also be used to alleviate the multicollinearity problem in multiple regression; see the paper by Newsome, Dudzinski and Low in ILCA, 1981.

Again, it must be emphasized that this type of analysis is exploratory, and little is known of its statistical properties. In particular, the stability of principal components in samples from the same population is a difficult area that has not been satisfactorily resolved. There is a particular problem when two components have similar variances; in another sample, they may be interchanged, or replaced by two quite different linear combinations, that are nevertheless equivalent to them as a pair. This is the problem of rotation, which plays an even larger part in factor analysis, but will not be discussed further here.

(c) Canonical variables.

The multiple regression model may be extended to the relationship between two sets of variables. This is a truly multivariate problem, and one that arises very commonly; it may be required, for instance, to relate numbers of animals of different species to various environmental factors.

The appropriate statistical technique is known as canonical correlation analysis. Given two sets of variables, say

$x_1 \ldots x_p$ and $y_1 \ldots y_q$

the first step is to find linear functions of the x's and y's that have maximum correlation. There is a unique pair of such functions, subject to arbitrary scaling factors, and they are the first canonical variables. The second pair of canonical variables can then be defined as the pair of linear functions, uncorrelated with the first pair, that have maximum correlation, and so on.

The distributional theory of canonical variables and canonical correlations is well known, but involves assumptions that are unlikely to be satisfied in applications in connection with aerial surveys. Like the other techniques discussed in this section, canonical analysis is best regarded as an exploratory, empirical, method, that may throw some light on the relationships between the two sets.

In general, there are p or q, whichever is less, pairs of canonical variables defined by the process described. The correlations between these pairs decreases from the first onwards, and only those pairs with high correlations need be considered. Again, the main purpose of the analysis is to interpret these pairs of variables and so describe the types of association between the two original sets.

In this sort of analysis, the x and y variables are treated in exactly the same way. It is concerned with correlations, not with prediction of one set from the other, nor with the effect of one set on the other. Often, the natural interpretation of the results will be in terms of cause and effect; environmental factors do determine the distribution of wild and domestic animals to a great extent. This, however, is not a mathematical property of the analysis, which is merely concerned with associations.

Canonical correlation analysis is a powerful technique, which has been surprisingly little used in investigations of this sort. Webster (1977) gives an interesting discussion of applications in soil science. It would seem ideally suited to the analysis of satellite data, and particularly to investigations of the relationships between satellite signals and ground conditions. In fact, most of the Landsat methodology is based on pattern-recognition techniques - presumably because of constraints on computer time. See, however, Honey et al, (1981), relating Landsat signals to observations on the vegetation in Western Australia.

(d) Cluster Analysis

Cluster analysis is the term used for a number of different techniques for dividing observations into groups. The general requirement is that group members should be as much alike as possible, on the basis of the observations made upon them. There may also be an attempt to find "natural" groups, on the assumption that there is some underlying structure that suggests a particular way of subdividing the data and a particular number of groups. In the case of aerial survey data, however there seems no reason for such an expectation. Cluster analysis is best regarded as a convenient way of classifying observations into groups, with membership of a particular group conveying as much information as possible about the item.

The very large literature on cluster analysis cannot disguise the fact that the methods used in practice are extremely crude and simplistic. The first step is to define a measure of dissimilarity (or, equivalently, of similarity). This can be thought of as a distance between points representing the observations, and the idea is to join up points that are close together to form compact groups. The definition of dissimilarity is obviously crucial, and is completely subjective. Usually the dissimilarity between two observations is based on the differences between them on each variable; the sum of squares of these differences is the squared Euclidean distance between the points, and the sum of the absolute differences is known as the City-block distance, or the Manhattan distance. The differences may be weighted in various ways. Usually, it is meaningless to use the crude differences between the original variables; some transformation may be needed, and standardization to the same standard deviation is essential in most cases. The weights attached to different variables are then chosen to reflect the experimenter's views about the relative importance of the different variables. It is sometimes suggested that the choice of equal weights is "objective" - it is not, it is merely one of a set of possible subjective choices, and is often adopted from inertia rather than conviction. Once weights have been chosen, the appropriate distance is calculated.

The next step is to join the two nearest observations, then the next two, and so on. Here a further important decision is needed; the distance between two individual observations has been defined, but we also need the distance from an individual to an already formed group, and the distance between two such groups. These distances may be defined in various ways, and the final clustering depends critically on the definition adopted. The three commonest definitions of the distance between two groups are the distance between the nearest points in each group, between the furthest points in each group, and the distance between the centroids. Adopting the first option gives "single-linkage" clustering, while the second is known as "complete-linkage". Once a definition has been chosen, the joinig-up procedure continues. If there are n individual observations, after n-r steps they fall in r groups.

The difference between single-linkage and complete-linkage clustering shows up most obviously in the type of groups that they

generate. When there are in fact clearly distinguished natural groups, both methods will find them. Otherwise, complete-linkage tends to give compact groups of similar size, while single-linkage tends to give one or two large groups with isolated points unattached, forming single point groups. Where there are natural groups, but they are not clearly separated without overlap, the two types of clustering both tend to distort the grouping, in the direction of compact, equal groups for complete linkage, or of straggling, unequal groups for single linkage. The centroid method is intermediate between these two extremes, and is a compromise intended to avoid both forms of distortion. There are a number of other intermediate methods, claimed by their inventors to have advantages in at least some situations.

There is one technical point of some importance. If there are ties among the distances, so that there may have to be arbitrary decisions about which points to join first, the structure of the resulting clusters may be quite different according to the decisions made. This is true of all clustering methods except single linkage. When the variables concerned are continuous, or counts of large numbers, so that ties are unlikely, the point may seem irrelevant; but it has a further implication that single-linkage clustering is likely to be more robust to small changes in the distances, in the sense that such changes are likely to have only minor effects on the final clusters.

The final decision to be made is the choice of the number of clusters. In the present context, it has already been suggested that grouping is largely arbitrary, and the choice is primarily one of convenience. For mapping purposes, probably 4-8 is a suitable range of numbers; the final choice may be guided by the dissimilarities between the objects joined at each stage of the clustering. If there is a sharp rise in dissimilarity, this suggests that very dissimilar objects are being grouped, and clustering should stop before this step. Another guide is the degree of fragmentation of the map. If the next step in combining groups produces a much more regular map, with fewer isolated patches, it should probably be made.

There is a large literature on cluster analysis, and a large number of different techniques. Many of them are implemented in the computer package CLUSTAN. The success of cluster analysis can be judged from the results; a method that gives a grouping consonant with common sense, with groups having obvious features in common, forming reasonably compact areas on the ground, may be judged to have worked. Discussions of the advantages and disadvantages of different definitions of distance and different clustering techniques are largely irrelevant; usefulness of the result is the real criterion. There is a good account of the most important techniques in Webster (1984). This is presented in the context of soil mapping, and much of the discussion is immediately relevant to aerial survey methods.

REFERENCES

Dixon,W.J.(ed.)(1983). BMDP Statistical Software. University of California Press, Berkely and London.

Draper,N. and Smith,H. (1981). Applied Regression Analysis. (2nd. ed.) John Wiley and Sons, New York.

Honey,F.R., Campbell,N.A., Tapley,I.J., Burnside,D.G. and Holman,W.F. (1981). Range inventory and monitoring using Landsat data in the semi-arid perennial graslands of north-western Australia. Proc. Landsat 81 Conf.

ILCA.(1981). Low-level Aerial Survey Techniques. International Livestock Centre for Africa, Addis Ababa.

Journel,A.G., and Huijbregts,Ch.J.(1978). Mining Geostatistics. Academic Press, London.

Milligan,K. (1983). An Aerial Reconnaissance of Livestock and Human Population in Relation to Land Use and Ecological Conditions in the SORDU Project Area of Southern Ethiopia. JEPSS Research Report No. 5, Addis Ababa.

Norton-Griffiths,M.(1978). Counting Animals. (Second edition). African Wildlife Leadership Foundation, Nairobi.

Pearson,K. (1920). On the Construction of Tables and on Interpolation. Tracts for Computers, Nos. 2 and 3, Cambridge University Press, London.

Quenouille,M.H.(1953). The Design and Analysis of Experiment. Griffin, London.

RIM.(1984). Farming and Settlement Patterns in Bauchi State, Nigeria. Resource Inventory and Management, St. Helier.

RIM(1984). Livestock and Land Use in Southern Gongola State, Nigeria. Resource inventory and Management, St. Helier.

Silverman,B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. JRSS B, 47 1-52.

Webster,R. (1977). Canonical correlation in pedology: how useful? Journal of Soil Science, 8 196-221.

Webster,R. (1984). Quantitative and Numerical Methods in Soil Classification and Survey. (2nd. ed.) Clarendon, Oxford.

Yates,F.(1981). Sampling Methods for Censuses and Surveys. (Fourth Edition). Griffin, London.