SUBSAMPLING OF LARGE LIGHT TRAP CATCHES OF *CULICOIDES* (DIPTERA: CERATOPOGONIDAE)

H. VAN ARK and R. MEISWINKEL, Onderstepoort Veterinary Institute, Onderstepoort, 0110

ABSTRACT

VAN ARK, H. & MEISWINKEL, R., 1992. Subsampling of large light trap catches of *Culicoides* (Diptera: Ceratopogonidae). *Onderstepoort Journal of Veterinary Research*, 59., 183–189 (1992)

Analyses of 2 light trap catches comprising 6 041 and 1 598 *Culicoides* showed that the reliability of subsampling such catches increased with subsample size, while the subsampling error decreased with an increase in the number of individuals per species present in a subsample. Subsamples comprising approximately 500 *Culicoides* are deemed sufficient for comparing population densities between sites and sampling occasions and also give an acceptable indication of relative species abundance at a site.

It is recommended that species for which the mean number of specimens in subsamples originating from 3 catches at a site is less than 7, should not be included in biometric analyses aimed at comparisons of population densities. For all other species a test level of 1 % rather than 5 % should be used for such comparisons. When species ratios obtained from subsampled catches are employed as indicators of abundance, the chi-squared test should be utilized at a 1 % level if the ratios originate from 3 catches at a site and at a 0,1 % level if only one catch per site is made. Due to poor representativeness of small catches, it is suggested that species for which fewer than 7 individuals are present in a single subsample, be excluded from chi-squared tests.

A 5-point procedure for subsampling a large light trap collection of *Culicoides* is given.

INTRODUCTION

Worldwide light traps of various models are routinely used as the quickest and simplest way to sample adult flying *Culicoides* to obtain ecological and epidemiological information on the presence, relative abundance and age structure of species that may transmit diseases of veterinary importance. In South Africa catches made during the warm summer months are usually very large (1 500– 500 000 *Culicoides*) and even though only a single light trap is usually operated per site the labour involved to exhaustively sort such catches and identify them to species is prohibitive. In a research programme recently initiated in the Kruger National Park, South Africa, to investigate the possibility that the zebra (*Equus burchelli*) is a reservoir of African horsesickness virus (AHSV), it is planned to survey different sites for a period of at least 1 year to obtain data on the prevalence and abundance of *Culicoides* found in association with zebras. For comparisons of *Culicoides* populations to be reliable 3 light-traps will be operated at each site once a month. As subsampling catches would greatly lighten the task of sorting and identifying, it was decided to investigate the variance increasing effect that subsampling has and to establish at which confidence limits the results remain valid and representative.

The relative abundance of various *Culicoides* species at a site is usually expressed as percentages of the total number caught (Nevill & Anderson, 1972; Jupp, McIntosh & Nevill, 1980; Phelps, Blackburn & Searle, 1982; Nevill, Venter, Edwardes, Pajor, Meiswinkel & Van Gas, 1988; Venter & Sweatman, 1989; Kitaoka & Zulu, 1990). It is reasonable to suggest that the ratios for the more abundant species in light trap catches reflect both the characteristics of the environment at a site and the abundance of specific host animals. Species ratios may therefore contribute to identifying areas where the transmission of livestock diseases are most likely to occur. The effect of subsampling on the reliability of species ratios was therefore also investigated.

MATERIALS AND METHODS

Two routine light trap catches were used to study the error introduced by subsampling large catches. The catches were made by means of a light trap consisting of a 220 V, 8 W blacklight tube and a down-draught suction motor. Entry of large insects was restricted by means of a 2 mm mosquito gauze fitted to the trap. The insects were collected in a phosphate buffered saline solution to which "Savlon"[1] was added to break the surface tension of the liquid and to preserve the specimens.

A large light trap catch of 6 041 *Culicoides* made during April 1989 on the Kawalazi estate nr. Mzuzu in north-eastern Malawi was used to initiate the investigation into the reliability of subsampling. This mixed catch (all insects passing through the fitted mosquito gauze being included) was suspended in alcohol by gently shaking the container. Twenty subsamples of equal volume (comprising the total catch) were then drawn from the suspension by means of a pipette. A second smaller catch of 1 598 *Culicoides*, made during October 1991 at the research camp in Skukuza, Kruger National Park was subsequently selected to verify the subsampling results obtained for the Malawi catch. All non-*Culicoides* were removed from this catch before subsamples were drawn. The catch was also suspended in alcohol and gently shaken before the insects were allowed to settle to the bottom of the

---

[1] ICI Chemicals SA Pty. Ltd.

TABLE 1 Composition of the 2 light trap catches used for the evaluation of subsampling

Malawi, April 1989			KNP, October 1991		
Species	Number	Ratio (%)	Species	Number	Ratio (%)
C. fulvithorax	3 560	58,93	*C. exspectator*	311	19,46
C. imicola	866	14,34	*C. leucostictus*	290	18,15
C. glabripennis	730	12,08	*C. tropicalis*	249	15,58
C. albovenosus	420	6,95	*C. imicola*	220	13,77
C. sp. # 3	113	1,87	*C. ravus*	123	7,70
C. leucostictus	87	1,44	*C. similis*	65	4,07
C. accraensis group	62	1,03	*C. bedfordi*	62	3,88
C. miombo	54	0,89	*C. tororoensis*	55	3,44
C. pereittii	43	0,71	*C.* sp. # 54	46	2,88
C. sp. # 9	38	0,63	*C. nivosus*	36	2,25
C. dubitatus	21	0,35	*C. bolitinos*	26	1,63
C. trifasciellus	18	0,30	*C.* sp. # 50	21	1,31
C. tropicalis	15	0,25	*C.* sp. # 3	18	1,13
C. citroneus	4	0,07	*C.* sp. # 35	14	0,88
C. dutoiti	3	0,05	*C. pycnostictus*	13	0,81
C. milnei group	2	0,03	*C. coarctatus*	13	0,81
C. hortensis	2	0,03	*C. kobae*	11	0,69
C. similis	2	0,03	*C. micheli*	9	0,56
C. sp. # 50	1	0,02	*C. nigeriae*	5	0,31
			C. brucei	3	0,19
			C. schultzei	2	0,13
			C. sp. # 107	2	0,13
			C. pseudopallidipennis	2	0,13
			C. quinquelineatus	1	0,06
			C. zuluensis	1	0,06
Total	6 041			1 598	

container. Eleven subsamples of 1 mℓ (comprising the total catch) were then drawn by means of a pipette inserted in the centre of the settled insects.

The *Culicoides* present in each subsample were identified to species and counted. The species composition and ratios of both total catches are given in Table 1.

In order to determine the effect of different subsample sizes on the subsampling error, individual subsamples were randomly pooled to obtain subsamples of increasing size. This random selection was repeated 3 times in order to obtain more representative results. Three replicates for each subsample size were therefore available for analyses. The total for all species in each subsample was regarded as repesentative of a very abundant species and also included in the analyses. For each subsample size the number of individuals of each species was multiplied by the reciprocal of the subsample fraction, rounded to the nearest whole number and used as the estimated total catch for that species.

Percentage error of subsampling

The actual and estimated total catch for each species in each subsample fraction were transformed to $\log_{10}(X+1)$ before the percentage error was calculated (see Discussion).

The reliability of subsampling for each species in each subsample was expressed as the percentage error of the estimated total catch in relation to the corresponding actual total catch (Van Ark, 1975), viz:

$$Y_{(i)} = \frac{\left| X_{(i)} - Y'_{(i)} \right|}{X_{(i)}} \times 100$$

where $X = \log_{10}$ (actual total catch + 1)
and $Y' = \log_{10}$ (estimated total catch + 1)

Negative exponential regressions were fitted for the percentage error on the number of individuals of each species in a subsample, utilizing all 3 replicates. The PC-program LINREG.EXE, written by H.v.A., was used for fitting the regressions. The linearised form of the equation is:

$\log_{10}(Y) = a + bZ$
where Y = the percentage error
and $Z = \log_{10}$ (number of individuals + 1 of a species in a subsample)

Species for which fewer than 10 individuals were present in the total catch (<0,2 % and <0,6 % for the Malawi and KNP catches respectively) were not considered and replicates having 100 or 0 % error were also ommitted. These omissions were necessary to improve representativeness and goodness of fit of the regressions (see Results).

Species ratios

For species ratios the actual subsample catches were utilized. The species ratios expressed as percentages for each subsample were calculated as follows:

$$Y_{(i)} = \frac{X_{(i)}}{XT} \times 100$$

Where X = the number of individuals of a species;
and XT = the total number of *Culicoides* in a subsample.

To determine if species ratios obtained from subsamples corresponded closely with those of the total catches, the ratios of the 7 most common species in each subsample were compared with the corre-

sponding ratios in the total catch by means of chi-squared tests (see Discussion). The PC-program FREQ.EXE written by H.v.A., was used for these comparisons.

RESULTS

Percentage error of subsampling
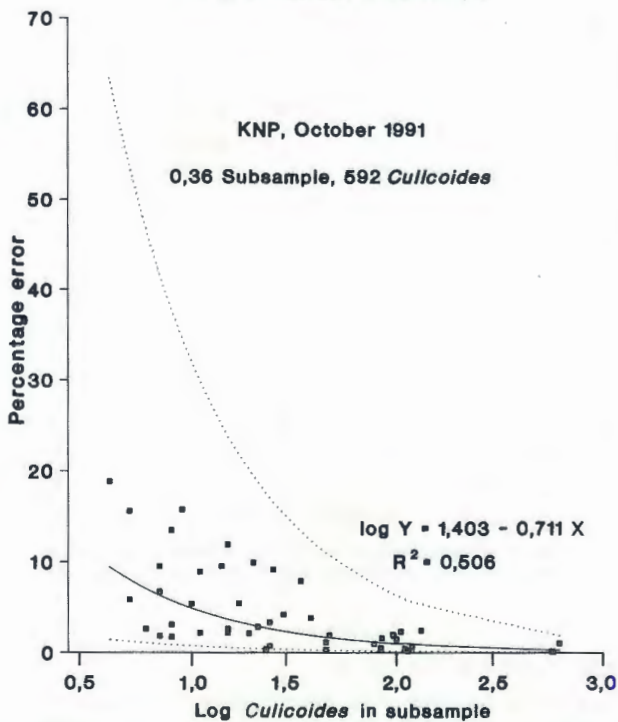
The percentages error for species caught in very

small numbers varied considerably, ranging from zero to 100 %. Errors of 100 % occurred most often for the smaller subsamples, due to the absence of individuals of such species from the subsamples. On the other hand, for those species caught in small numbers, zero errors occurred with increasing frequency as the size of the subsample increased. This is hardly suprising considering the increased probability of estimating the total catch exactly for those species represented by small numbers.

All negative exponential regressions fitted very highly significantly (P <0,001), with the coefficients of determination ranging from 0,31 to 0,86 for increasing subsample sizes. This indicates that the percentage error increased negative exponentially with a decrease in the number of individuals of a species present in a subsample. Two examples of these curves are shown in Fig. 1.

According to Southwood (1966), a sampling error of approximately 10 % is tolerable for ecological studies and 25 % for purposes where large differences must be distinguished. The expected number of individuals of a species in each subsample corresponding to an error of 10 % and 25 % and the associated 95 % confidence limits, were calculated from the exponential regressions. The expected values for a 10 % error are plotted in Fig. 2. Although these values tended to be somewhat larger for the KNP catch than the Malawi catch, overlapping occurred and no real differences for either the 10 % or 25 % error values were evident between the 2 catches. The number of individuals of a species in a subsample at which an error of 10 % was obtained did not decrease with increasing subsample size and varied from 1 to 6 individuals over all subsample

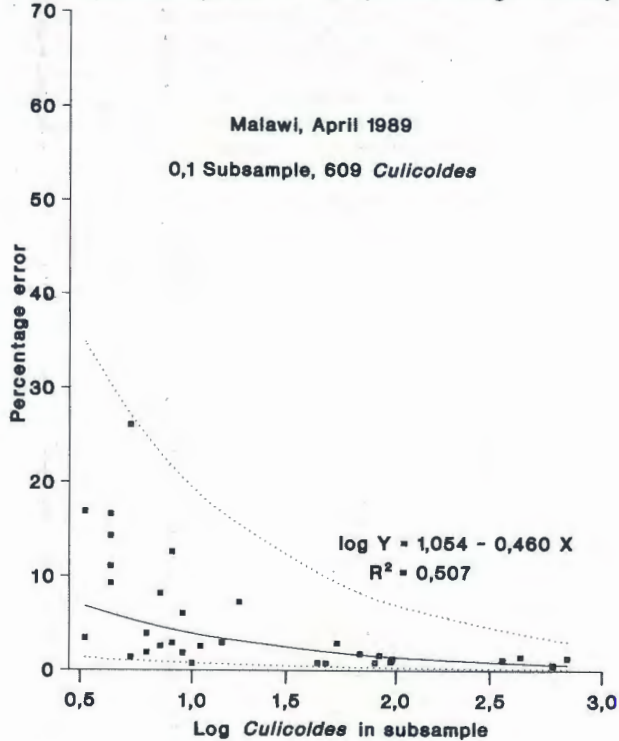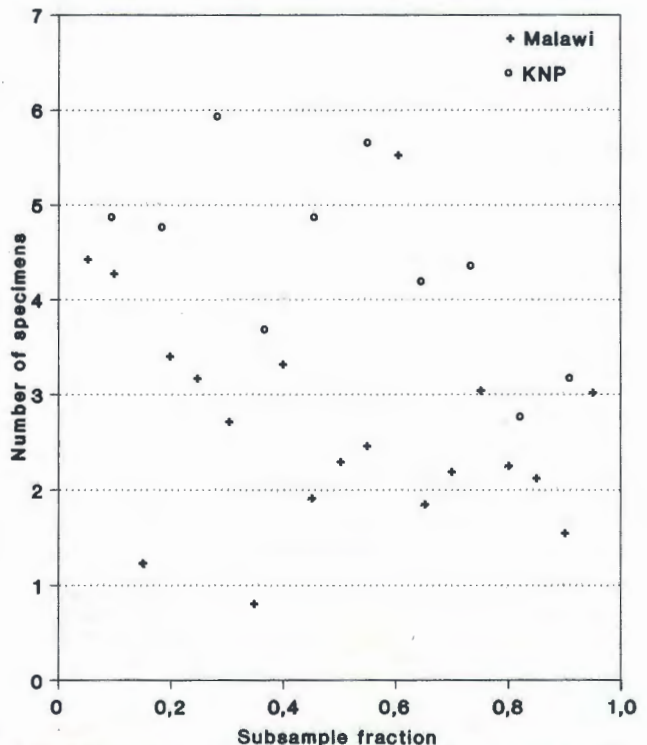FIG. 2 Expected number of specimens in a subsample at a subsampling error of 10 %

FIG. 1 Examples of the nagative exponential relationship between the percentage error and the logarithm of the number of specimens per species in a subsample

185

sizes. Similar results were obtained for the 25 % error, with the number of individuals varying from 1 to 2. However, the associated 95 % confidence limits for both error levels decreased linearly with increasing subsample size. The expected number of individuals at both error levels as well as the fitted linear regression for the upper limits for the 2 catches combined (Malawi and KNP) are shown in Fig. 3. The regression for the 95 % limits associated with the 25 % subsampling error fitted rather poorly (P = 0,012 and R^2 = 0,22). The upper 95 % confidence limits for the 10 % error calculated from the linear regression for selected subsample fractions are given in Table 2.

values were significant at P = 0,05, indicating that the ratios of species for all replicates of all sub-sample sizes were similar to those of the corresponding total catch. The chi-squared values (each with 6 degrees of freedom) comparing species ratios in the subsamples with those in the total catches of both sites combined, are plotted against the sub-sample fractions in Fig. 4. The linearised form $[\log_{10}(Y) = a + b(SF)]$ of the negative exponential regression fitted for the chi-squared values on the subsample fractions for both catches pooled is also given in Fig. 4. The expected chi-squared values calculated from this regression for selected sub-sample fractions are given in Table 3.

TABLE 2 Expected number of individuals of a species at the 95 % confidence limits for a subsampling error of 10 %

Subsample fraction	Upper 95 % confidence limit
0,05	72
0,1	69
0,2	63
0,3	57
0,4	51
0,5	45
0,6	39
0,7	33
0,8	27
0,9	21

TABLE 3 Expected chi-squared values for species ratios

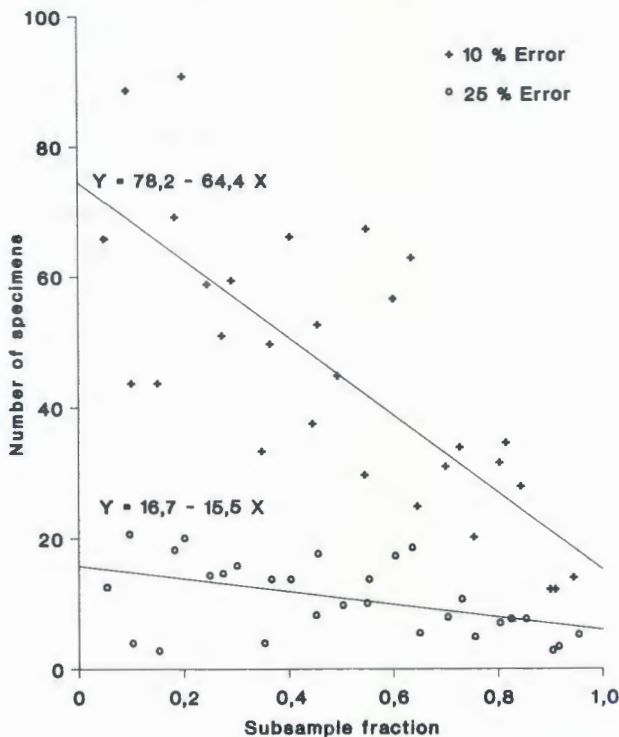| Subsample fraction	Chi-squared value
0,05	4,84
0,1	4,21
0,2	3,18
0,3	2,40
0,4	1,82
0,5	1,37
0,6	1,04
0,7	0,78
0,8	0,59
0,9	0,45

FIG. 3 Linear relationship between the upper 95 % confidence limits for the expected number of specimens in a sub-sample and the subsampling fraction, calculated for sub-sampling errors of 10 % and 25 %
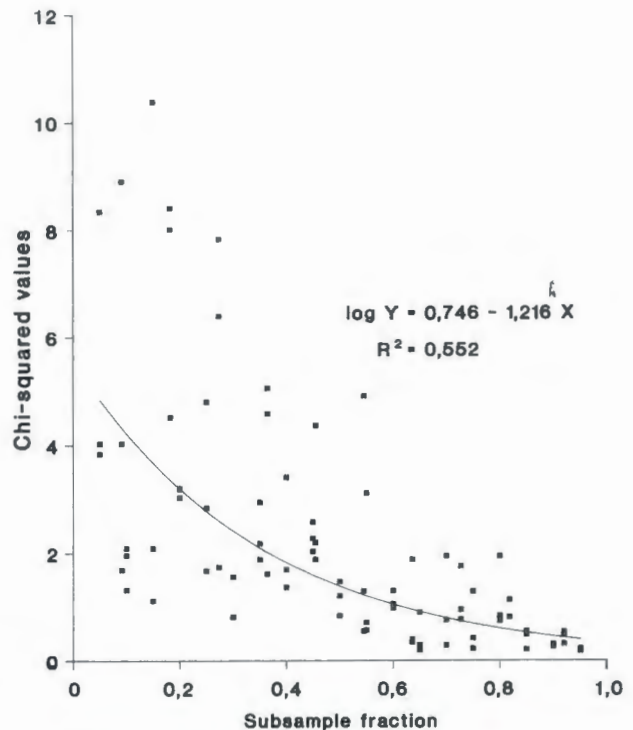

FIG. 4 Negative exponential relationship between chi-squared values for species ratios and the subsample fraction

Species ratios

The general decrease in magnitude of the chi-squared values for increasing subsample sizes were similar for both catches. None of these chi-squared

DISCUSSION

Logarithmic transformation

Light trap catches may vary largely in size and in species composition, tending to simulate a conta-gious distribution. It is common practise to use an appropriate transformation (usually logarithms) to

normalize such data before biometric tests are carried out (*inter alia* Williams, 1937; Taylor, 1961; Van Ark, 1975, 1976; Petney, Van Ark & Spickett, 1990). Future ecological and epidemiological studies will include comparison of catches between sites and trapping periods. Such comparisons require simultaneous use of at least 3 light traps at any site chosen, with the obtained catches transformed to logarithms. The inflationary effect of subsampling on the experimental error was therefore studied on this basis.

Percentage error of subsampling

In general the data conformed to the already established trend of exponentially decreasing subsampling errors with increasing numbers of specimens of a species present in a subsample (Fig. 1) (*inter alia* Clark, Tucker & Turton, 1971; Venrick, 1971; Van Ark, 1975). By the same law, this introduced error decreased with an increase in the subsample size.

Subsampling *Culicoides* catches made by means of the specific screened light traps described here was considerably more effective than subsampling catches of all available night flying insects made by means of unscreened light traps. Van Ark (1975) found, for example, that a 10 % subsampling error for a subsample fraction of 0,2 could be expected for a mean estimated total catch of 81 insects at one site and 97 insects at another site. For the uncleaned Malawi catch the corresponding mean number of *Culicoides* was 16, and for the cleaned KNP catch 20. This is entirely reasonable considering the extreme heterogeneity of insect sizes in unscreened light trap catches compared with relatively homogeneous insect sizes in the presently screened light traps used for catching *Culicoides*. The distribution of small equal sized insects suspended in a liquid would tend to be more random than for dry insects varying greatly in size (from minute to very large) for which the catch is mixed by hand.

Cleaning catches

After analysing the uncleaned Malawi catch (containing a total of 15 544 non-*Culicoides*, mainly Chironomidae), it was thought that cleaning catches would reduce the subsampling error. The cleaned KNP catch, containing approximately 1 700 non-*Culicoides*, was used for this comparison. Comparisons of the negative exponential regressions indicated no differences in the magnitudes of the subsampling errors between the 2 catches. The percentages error for comparable subsample fractions of the 2 catches overlapped nearly completely. This similarity is probably due to the fact that the non-*Culicoides* present in the catches differed little in size from the *Culicoides,* thus having little or no effect on the distribution of *Culicoides* in the suspensions. Cleaning catches before subsampling seems to be unnecessary provided that the catches are not completely swamped by non-*Culicoides*. Manual cleaning of catches is time consuming and increases the chance of damage to *Culicoides* specimens because the catch must be handled twice (see recommended subsampling procedure).

Limitations due to subsampling

A light trap catch is merely a sample of the *Culicoides* available for capture at a specific site. The representativeness of this sample in relation to the population densities of different species present depends largely on a number of factors such as activity patterns, the stage of gonotrophic development, physical environmental factors, the attractant properties of the light trap and the number and placement of the light traps used. The exact sampling error relating to the population variance is unknown unless catch sizes are substantiated by other, perhaps more reliable, sampling methods. Therefore the contribution of the subsampling error to the total sampling error remains largely unknown and will probably vary for the different species present at each site depending on the prevailing circumstances. The subsampling errors indicated by this study should consequently be used only as guidelines to determine catch sizes for which the inference based on biometric analyses may lead to doubtful conclusions.

Culicoides surveys are mainly made for ecological or epidemiological purposes, thus necessitating the use of a 10 % error limit of subsampling for exclusion of small catches of a species. The subsampling process is accurate enough not to have to resort to using the 25 % error level as the exclusion limit (Fig. 2 and 3).

Using the upper 95 % confidence limits of the 10 % error for excluding catches (Table 2) from biometric analyses would probably be much too severe a limitation in relation to the total sampling error involved. Using these limits would also eliminate a relatively large number of species from analyses and useful ecological information could be lost for such species. It is therefore recommended that the expected subsampling error, calculated from the negative exponential regressions, be used for this purpose, i.e. species which have a mean of less than 7 individuals in subsamples from 3 catches, regardless of the subsample size, should be excluded from biometric analyses aimed at establishing differences in catch sizes between comparable sites or sampling occasions. Catches may vary greatly between sites and over time and the number of light trap catches that can be sorted is limited to the labour available. Small subsamples would therefore be favoured, and, in addition to the exclusion of species represented by small numbers, differences should be tested at a more conservative test level of 1 % instead of at the customary 5 % level, thus increasing confidence that obtained differences really do exist. This recommended procedure may better suit the researcher who is not mathematically inclined rather than utilizing the formulae given by Venrick (1971) for estimating the total error variance when subsampling is undertaken. Moreover, computer programs generally available for routine analyses make no provision for such revised estimates of variance.

Catches of particular *Culicoides* species are sometimes subsorted to sex, while the females may be further subsorted into nulliparous, parous or gravid individuals to obtain other information on their

187

potential role in disease transmission. The recommended exclusion limit of less than 7 individuals and a test level of 1 % should also be observed for each of these categories.

Species ratios

Environments which are partly controlled by man (stables, kraals, etc.) seem to limit *Culicoides* representation to a relatively small number of species, but with some present in very large numbers and thus considered to be of potential vectorial importance (Nevill *et al.*, 1988; Venter & Sweatman, 1989). More natural environments like the Kruger National Park, on the other hand, harbour more species at more equal abundance rates (Table 1). Utilizing only the 7 most abundant species for determining the effect of subsampling on species ratios should be sufficient to cover most of the types of environments with which specific *Culicoides* species are associated.

Although variation between replicates for a specific subsample size existed, the chi-squared values for testing discrepancies in species ratios between the total catch and subsamples generally decreased negative exponentially with an increase in the subsample size (Fig. 4). As for the percentage subsampling error, it is considered reasonable to regard the expected chi-squared value and not the 95 % confidence limits as the criterion for the limitation subsampling introduces. Considering the fact that subsample fractions of 0,1 and somewhat larger will probably be routinely obtained from 3 catches at a site, comparing species ratios between sites or sampling occasions at a 1 % level would make provision for most of the introduced variation. If ratios originating from single subsampled catches are to be compared, a test level of 0,1 % is recommended to compensate for the relatively poorer representativeness of a single catch than 3 catches in relation to the existing population density sampled.

The chi-spuared test has a weak discriminatory power when small expected frequencies are used and frequencies smaller than 5 should normally be excluded from analyses (Siegel, 1956). Small numbers of a species present in a catch in relation to the total numbers caught (approximately less than 2 %, see Table 1) are generally considered to be unimportant and are either non-representative of the population density of that species or are of a species which never occurs in high numbers such as tree-hole breeders. An exclusion limit of 5 specimens for a species can safely be set for each of 3 subsampled catches made at a site and to 7 specimens for a single subsampled catch. These limits should not cause serious loss of important ecological information.

Occurrence of rare species

Specimens of new or little known species are of great importance to the taxonomist. Population densities of different species differ greatly and species remaining undetected usually occur in small numbers and are therefore poorly represented in light trap catches. The probability of inclusion of specimens of a particular species in a subsample diminishes sharply with a decrease in abundance of

TABLE 4 Mean number of specimens of a species in the total catch corresponding with recovery of at least one specimen in all 3 replicate subsamples

Subsample fraction	Mean number
0,05	22
0,1	18
0,2	12
0,3	8
0,4	6
0,5	4
0,6	3
0,7	2
0,8	1
0,9	1

the species in the total catch (Table 4). If loss of specimens of rare or new species cannot be tolerated, the complete catch should be sorted.

Recommended subsampling procedure

No differences in effect of subsampling were observed between the 2 methods employed. However, a standardized method for all future collections made would be preferable if data accumulated over many years is to be comparable. No subsampling advantage was gained from the laborious cleaning of catches, but some method of removing most non-*Culicoides* would reduce labour especially if the latter occur in very large numbers. Increasing the subsample size reduced variation but did not result in more favourable exclusion limits for species at the 10 % error limit. Subsamples consisting of approximately 500 specimens should therefore be practical and suitable. The following subsampling procedure is recommended, viz:

1. Separate *Culicoides* from non-*Culicoides* by means of a 1,65 mm Endicott sieve.

2. Suspend the *Culicoides* and other insects which passed through the sieve, in a known volume of alcohol by means of lightly shaking the container; the volume of alcohol should be adapted for enormous catches.

3. Take a subsample of known volume from the middle of the suspension by means of a pipette with an aperture of at least 2 mm and count and identify all specimens.

4. Continue subsampling until approximately 500 specimens are obtained.

5. Calculate the volume of the subsample in relation to the volume of the total catch and determine the subsample fraction for calculation of the estimated total catch.

REFERENCES

CLARK, C. J., TUCKER, A. M. & TURTON, J. A., 1971. Sampling technique for estimating roundworm burdens of sheep and cattle. *Experimental Parasitology*, 30, 181–186.

JUPP, P. G. McINTOSH, B. M. & NEVILL, E. M., 1980. A survey of the mosquito and *Culicoides* fauna at two localities in the Karoo region of South Africa with some observations on bionomics. *Onderstepoort Journal of Veterinary Research*, 47, 1–6.

KITAOKA, S. & ZULU, F. P., 1990. Species composition of *Culicoides* (Diptera: Ceratopogonidae) found at Chilanga near Lusaka, Zambia. *Bulletin of the Niigala Sangyo University*, 4, 197–206.

NEVILL, E. M. & ANDERSON, D., 1972. Host preferences of *Culicoides* midges (Diptera: Ceratopogonidae) in South Africa as

determined by precipitin tests and light trap catches. *Onderstepoort Journal of Veterinary Research*, 39, 147–152.

NEVILL, E. M., VENTER, G. J., EDWARDES, M., PAJOR, I. T. P., MEISWINKEL, R. & VAN GAS, J. H., 1988. *Culicoides* species associated with livestock in the Stellenbosch area of the western Cape Province, Republic of South Africa (Diptera: Ceratopogonidae). *Onderstepoort Journal of Veterinary Research*, 55, 101–106.

PETNEY, T. N., VAN ARK, H. & SPICKETT, A. M., 1990. On sampling tick populations: The problem of overdispersion. *Onderstepoort Journal of Veterinary Research*, 57, 123–127.

PHELPS, R. J., BLACKBURN, N. K. & SEARLE, L., 1982. *Culicoides* (Diptera: Ceratopogonidae) catches and virus isolations from them in the Mukwadzi Valley, Zimbabwe. *Journal of the entomological Society of southern Africa*, 45, 195–200.

SIEGEL, S., 1956. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Co. Inc.

SOUTHWOOD, T. R. E., 1966. Ecological methods, with particular references to the study of insect populations. London: Methuen.

TAYLOR, L. R., 1961. Aggregation, variance and the mean. *Nature*, 189, 732–735.

VAN ARK, H., 1975. The error involved in subsampling of large light trap catches of insects. *Phytophylactica*, 7, 1–8.

VAN ARK, H., 1976. The effect of dieldrin coverspraying on populations of night-flying insects. *Phytophylactica*, 8, 31–36.

VENRICK, E. L., 1971. The statistics of subsampling. *Limnology and Oceanography*, 16, 811–818.

VENTER, G. J. & SWEATMAN, G. K., 1989. Seasonal abundance and parity of *Culicoides* biting midges associated with livestock at Roma, Lesotho (Diptera: Ceratopogonidae). *Onderstepoort Journal of Veterinary Research*, 56, 173–177.

WILLIAMS, C. B., 1937. The use of logarithms in the interpretation of certain entomological problems. *Annals of Applied Biology*, 24, 404–414.