Sampling Statistics and Detection or Estimation of Diverse Populations of Stored-Product Insects[1]

DAVID W. HAGSTRUM, ROBERT L. MEAGHER,[2]
AND LAWRIE B. SMITH[3]

USDA–ARS, U.S. Grain Marketing Research Laboratory,
Manhattan, Kansas 66502

Environ. Entomol. 17(2): 377–380 (1988)

ABSTRACT Sampling data for a variety of stages and species of insects, sampling devices, locations in the marketing systems, and types of grain fit a single model with narrow confidence limits, indicating that the same sampling statistics might be used with diverse populations of stored-product insects. The progressively faster increase in variance and the progressively slower increase in the fraction of samples with insects as insect population density increased can both be explained by a purely probabilistic increase in the chance of a sample having more than one insect.

KEY WORDS Insecta, stored grain insects, modeling

SAMPLING POPULATIONS to estimate density is an integral part of any study of population dynamics. Ruesink & Kogan (1982) have reviewed the most pertinent literature related to sampling insect populations. The relationship between the mean number of insects per sample and the sample-to-sample variance or the percentage of samples with insects provide two measures of the uniformity of insect population distribution over an area. The more evenly a population is distributed, the fewer samples are needed for detection or estimation. These measures can thus be used to calculate the number of samples needed for detection or estimation of the insect population in that area. Hagstrum et al. (1985) show how they can be used to calculate the number of grain trier samples required for detection or estimation of insect populations in a bin of farm-stored wheat. Hagstrum (1987) showed that equations used for these calculations applied over much of the storage season. Lippert & Hagstrum (1987) showed that these equations also worked for probe traps.

This paper presents sampling data for all developmental stages of *Tribolium castaneum* (Herbst), and for eggs of *Cadra cautella* (Walker), re-analyzes data from five published studies on diverse populations of stored-product insects, and compares sampling statistics from these seven studies. The mechanism underlying such sampling statistics is discussed.

Materials and Methods

The distribution of *T. castaneum* among samples was studied by taking 9 0.12-kg samples of wheat with a grain trier from each of two 0.7-bushel (24.7 liters) lots of wheat at 27°C and 70% RH, 30, 50, 70, 90, 110, and 130 d after initial infestation with five pairs of adult beetles. The insects were sieved from the wheat with an oblong-holed grain sieve (0.18 by 1.27 cm, Seedburo, Chicago, Ill.) and then the wheat samples were washed with 70% ethanol to remove eggs adhering to the wheat. The insects were sorted and counted by stage and larval instar. The pupae and adults were further sorted and counted by sex.

The distribution of *C. cautella* eggs between samples was studied by allowing individual females to oviposit during a 48-h period at 25°C and 50% RH on 7.5 kg of inshell peanuts spread evenly over an area of 1.2 by 1.2 m. Eggs were counted by dividing the area into 1,600 3-by-3-cm subdivisions, and the number of eggs in each sampling unit were entered into the computer as a 40 × 40 matrix using Cartesian coordinates of sample locations. Increases in egg density were simulated by adding like elements of the matrices together for 2, 4, 8, or 16 individual females. The similarity of the distributions created by overlaying to the actual distributions for groups of 2, 4, 8, or 16 females was confirmed by running four replicates at each of these female densities.

The Statistical Analysis System (SAS Institute 1982) was used to calculate means, confidence limits, and frequencies, fit regression equations, and plot data. The distributions of *C. cautella* eggs created by overlaying were compared with those for groups of females using model comparison procedures of Draper & Smith (1981). For Meagher

---

[1] The article reports the results of research only. Mention of a proprietary product does not constitute an endorsement or a recommendation for its use by USDA.

[2] Department of Entomology, P.O. Box 7626, North Carolina State University, Raleigh, N.C. 27695.

[3] Agriculture Canada, Research Station, Winnipeg, Manitoba R3T 2M9 Canada.

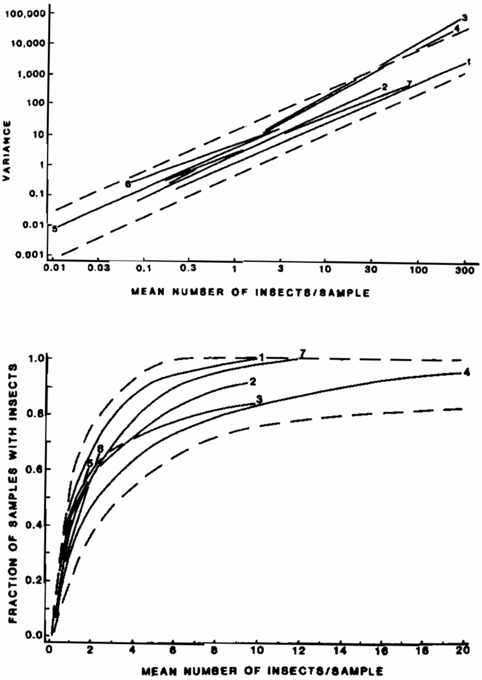

Fig. 1. Plot of the seven regression models in Table 1 (top) and Table 2 (bottom) with the prediction limits for the combined-data model in dashed lines.

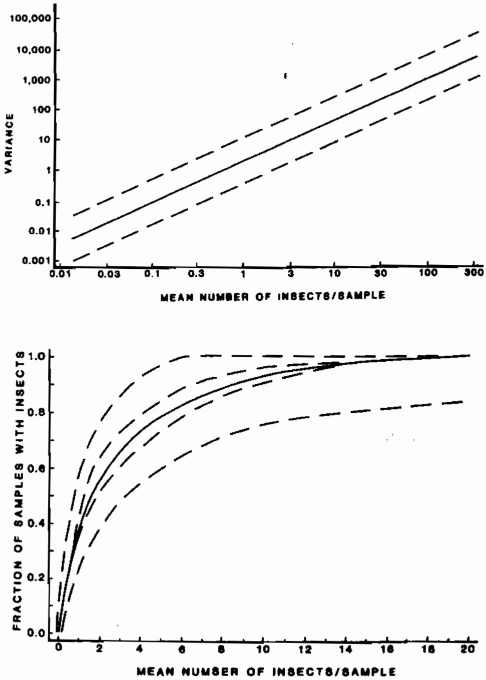

Fig. 2. Plot of the combined-data models in Table 1 (top) and Table 2 (bottom) with the 95% confidence and the wider 95% prediction intervals. The confidence limits for variance-mean model were too narrow to be seen on a logarithmic scale ranging from ±4% of mean at one insect per sample to ±10% of mean at both extremes.

et al. (1986), the means, variances, and percentages of samples with insects were calculated for each of three species at each of 5–7 depths in each of six bins using the nine samples taken at each depth, and for Smith (1985) they were calculated for the numbers of *Cryptolestes ferrugineus* (Stephens) in the 4–86 samples taken during the loading of a railroad car.

Results

The data from two new data sets and five published studies all indicated that the variance increased progressively faster than the mean (Fig. 1; Table 1) and that the fraction of samples with insects increased progressively slower as population density increased (Fig. 1; Table 2). In fact, the narrow 95% confidence limits about the models fitted to combined data from seven studies indicated that both of these measures of the uniformity of the population distribution over the area sampled increased in a similar manner for diverse populations of stored-product insects (Fig. 2). Further, the models for each study were generally within the 95% prediction interval for the combined-data model (Fig. 1).

The progressively faster increase in variance, and the progressively slower increase in the fraction of samples with insects that occurred as insect population density increased, can both be explained by a purely probabilistic increase in the chance of a sample having more than one insect. If the *C. cautella* eggs in excess of one in each sample were ignored, the slope of the logarithm of variance against logarithm of mean regression was 0.75 instead of 1.13. This indicates that additional eggs laid in sampling units that already contained eggs had resulted in the variance increasing faster than

Table 1. Regression models for logarithm of variance against logarithm of mean number of insects per sample

Data set	n	Slope	Intercept	r^2
1. Hagstrum et al. (1985)	539	1.32	0.061	0.8489
2. Hagstrum (1986)	357	1.43	0.34	0.8826
3. Lippert & Hagstrum (1987)	190	1.69	0.53	0.9424
4. Meagher et al. (1986)	73	1.61	0.51	0.9272
5. Smith (1985)	165	1.33	0.48	0.9070
6. *Cadra cautella*[a]	136	1.13	0.68	0.9377
7. *Tribolium castaneum*	204	1.17	0.39	0.9145
Combined data	1,718	1.32	0.32	0.8587

[a] The regression model for groups of females was not significantly different from model for densities simulated by overlaying data for individual females ($F_{133,135} = 0.0073$; $P < 0.005$).

Table 2. Regression models for percentage of samples with insects against mean number of insects per sample

Data set	n	A[a]	B[a]	C[a]	r^2
Hagstrum et al. (1985)	96	0.29	2.77	0.41	0.9347
Hagstrum (1986)	176	0.32	2.60	0.21	0.8353
Lippert & Hagstrum (1987)	368	0.52	1.41	0.11	0.8645
Meagher et al. (1986)	100	0.55	0.12	0.70	0.8932
Smith (1985)	240	0.30	0.70	0.90	0.9270
Cadra cautella	137	0.087	1.74	0.36	0.9762
Tribolium castaneum	282	0.19	1.18	0.32	0.9591
Combined data	1,406	0.31	1.86	0.23	0.8953

[a] These are coefficients in the double-logarithmic model $q = 1 - [A \exp^{-Bx} + (1 - A)\exp^{-Cx}]$.

the mean, i.e., slope >1. The progressively slower increase with density in the fraction of samples with eggs was also a result of an increase in the percentage of eggs being laid in sampling units that already contain eggs. The percentages of eggs laid in sampling units that already contained eggs were 53, 55, 59, 65, and 72 with 1, 2, 4, 8, and 16 females laying 221, 435, 870, 1,742, and 3,485 eggs, respectively. Similar calculations could be made for data from other studies. However, we used the data for *C. cautella* because with these data the possibility of density-dependent changes in moth behavior explaining the results was eliminated by overlaying egg distributions of single females to increase density. These data clearly demonstrate that the changes in variance and the percentage of samples with insects are a result of purely probabilistic changes in the chance of a sample having more than one insect.

Discussion

The similarity of sampling statistics for a variety of stages and species of insects, sampling devices, locations in the marketing system, and types of grain indicates that a single model might be used with confidence in many situations. The diversity among the seven studies considered in this paper can be characterized as follows. New data for *T. castaneum* represented all developmental stages. Data from Smith (1985) included larval and adult stages of *C. ferrugineus* and data from Hagstrum et al. (1985) and Hagstrum (1986) include adult stages of nine species and larval stages of two species. Overall, 14 species of insects were represented in the combined data set. Grain sampling techniques included a scoop inserted into grain stream leaving the elevator (Smith 1985) and a grain trier (Hagstrum et al. 1985, Hagstrum 1986), deep bin cups or probe traps (Lippert & Hagstrum 1987), and commercial pneumatic grain sampler (Meagher et al. 1986) in storage bins at the farms. Types of grain included wheat, grain sorghum, oats, barley, and flax.

Alternatively, some of the differences among the seven studies favor the use of different models in

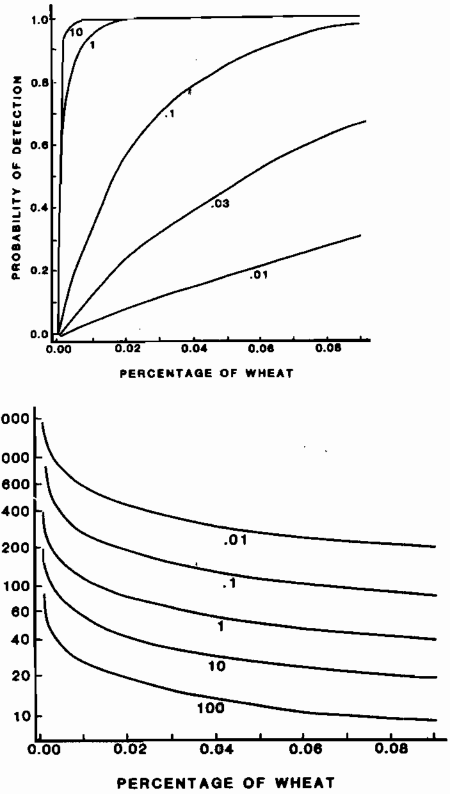

Fig. 3. The theoretical probabilities of detection (top) and accuracies of estimates expressed as percentage of mean (bottom) predicted by combined-data models for sampling rates ranging from an examination of 0.0009–0.09% of wheat for insects. Predictions are given for infestation levels (insects per 0.5-kg sample) indicated by numbers below each curve.

different situations. Variance in insect numbers increased faster relative to the mean when sampling devices such as the deep bin cup, probe trap, and pneumatic sampler were used to sample grain at a single point in the bin than for the trier, which sampled grain from a 1.27-m transect. Also, in the homogeneous laboratory environment, the increase in variance relative to the mean was lower for *T. castaneum* and *C. cautella* than in other studies.

In an earlier paper, Hagstrum et al. (1985) used the relationship between the mean number of stored-grain insects per sample and variance or percentage of samples with insects to calculate the number of samples needed to estimate infestation level or probability of detection. The equations used to calculate the number of samples required for detection or estimation can also be used to calculate changes in the probability of detection or the accuracy of estimates as the percentage of wheat that is examined for insects is increased (Fig. 3). Hagstrum (1987) used one of these equations to calculate the changes in the accuracy of estimates when sampling a growing insect population with a fixed number of samples. The expected accuracy of the estimate (c) expressed as a percentage of the mean was calculated by solving an equation from Ruesink & Kogan (1982) for c

$$c = 100t_{\alpha}\sqrt{\frac{a\bar{x}^{b-2}}{n}} \quad (1)$$

where n is the number of 0.5 kg samples or the percentage of wheat examined for insects/f, f is the percentage of 1,000 bushels (35,240 liters) represented by a 0.5-kg sample or 0.0018, t is the value of t distribution for confidence limit α, a is the antilogarithm of intercept, and b is the slope of the regression of the logarithm of the variance against the logarithm of the mean ($\bar{x}$). The probability of detection (P) can be calculated from the binomial distribution

$$P = (1 - q)^{n} \quad (2)$$

where n is the number of 0.5 kg samples or the percentage of wheat examined for insects/f, f is the percentage of 1,000 bushels (35,240 liter) represented by a 0.5-kg sample or 0.0018, and q is the probability that a sample will have at least one insect. The value of q can be calculated from the double-logarithmic model, $q = 1 - (A \exp^{-Bx} + (1 - A) \exp^{-Cx})$. The double-logarithmic model mechanistically explains the distribution of insects over an area as a two-stage process. The first stage is a logarithmic increase in the number of sampling units occupied as insect density increases. The second stage is a logarithmic increase with increasing density in the number of sampling units with more than one insect, which also reduces the rate at which additional sampling units become occupied.

As the percentage of wheat examined for insects increased, the probability of detection and accuracy of estimate improved (Fig. 3). Both depended upon the density of insect infestation. The improvements in detection resulting from increased sampling intensity decreased as the percentage of wheat examined for insects increased. At densities of 10 insects/0.5 kg., there was little improvement in detection when more than 0.00009% (0.025 kg/1,000 bushels) was examined. At all densities, the accuracy of estimates was improved less by each additional increase in the amount of wheat examined for insects. Each 10-fold increase in sampling intensity increased the accuracy of estimates 3-fold.

In this paper, sampling statistics are shown to be similar for a variety of stages and species of insects, sampling devices, locations in the marketing system, and types of grain. The similarity is perhaps a result of the changes with increasing density in both variance and percentage of samples with insects being explained mechanistically by the same simple two stage process in all seven studies. Simulation studies showed that improvements in detection with increased sampling intensity decreased as the percentage of wheat examined for insects increased, and that each 10-fold increase in sampling intensity increased the accuracy of estimates only 3-fold. These insights will allow us to plan more reasonably sampling strategies for ecological studies and pest management programs.

References Cited

Draper, N. R. & H. Smith. 1981. Applied regression analysis, 2nd ed. Wiley, New York.

Hagstrum, D. W. 1987. Seasonal variation of stored wheat environment and insect populations. Environ. Entomol. 16: 77–83.

Hagstrum, D. W., G. A. Milliken & M. S. Waddell. 1985. Insect distribution in bulk-stored wheat in relation to detection or estimation of abundance. Environ. Entomol. 14: 655–661.

Lippert, G. E. & D. W. Hagstrum. 1987. Detection or estimation of insect population in bulk-stored wheat with probe traps. J. Econ. Entomol. 80: 601–604.

Meagher, R. L., R. B. Mills & R. M. Rubison. 1986. Comparison of pneumatic and manual probe sampling of Kansas farm-stored grain sorghum. J. Econ. Entomol. 79: 284–288.

Ruesink, W. G. & M. Kogan. 1982. The quantitative basis of pest management: sampling and measuring, pp. 315–352. *In* R. L. Metcalf & W. H. Luckmann [eds.], Introduction to insect pest management. Wiley, New York.

SAS Institute. 1982. SAS user's guide: statistics. SAS Institute, Cary, N.C.

Smith, L. B. 1985. Insect infestation in grain loaded in railroad cars at primary elevators in southern Manitoba, Canada. J. Econ. Entomol. 78: 531–534.

Received for publication 24 March 1987; accepted 13 November 1987.