

CITRUS RUST MITE (ACARI: ERIOPHYIDAE) COUNTS ON
FRUIT AND THE NEGATIVE BINOMIAL DISTRIBUTION

D. G. HALL¹, C. C. CHILDERS², J. E. EGER³ AND J. C. ALLEN⁴

¹Research Department, United States Sugar Corporation
P.O. Drawer 1207, Clewiston, FL 33440

²University of Florida, Citrus Research & Education Center
Lake Alfred, FL 33850

³DowElanco, Tampa, FL 33607

⁴University of Florida, Department of Entomology and Nematology
Gainesville, FL 32611-0620

ABSTRACT

Count data for the number of citrus rust mites per cm² on fruit across a 4-ha (10-acre) area of orange trees followed the negative binomial probability distribution 79% of the time based on chi-square tests. A correlation of $r = 0.993$ was found between observed counts and counts projected based on the distribution. A common k of 0.149 was computed but generally appeared more suitable for mean densities of 3.0 to 55.0 than 0.5 to 3.0 citrus rust mites per cm². For mean densities of from 0.5 to 55 citrus rust mites per cm², the parameter k of the negative binomial was related to the mean density (\bar{x}): $k = 0.081 + 0.1139 * (\log_{10} \bar{x})$. Estimated k -values were used to draw expected count data profiles for several mean densities ranging from 1 to 40 citrus rust mites per cm². Due to the skewness of the count data, the number of mites per cm² expected in most individual samples was always considerably smaller than the average density. Based on the negative binomial, mean rust mite densities could be estimated from the percentage of samples with at least one mite. Results of the study provide a means to predict the relative frequency histogram of densities associated with a mean density of citrus rust mites per cm² across an area of trees.

Key Words: citrus rust mite, *Phyllocoptruta oleivora*, sampling, negative binomial distribution

RESUMEN

Los datos del número de *Phyllocoptruta oleivora* por cm² en frutos de naranja en un área de 4 ha (10 acres) siguieron una distribución de probabilidad binomial negativa en el 79% de los casos, basada en pruebas de chi-cuadrada. Fue encontrada una correlación de $r = 0.993$ entre los conteos observados y los proyectados sobre la base de la distribución. Una k común de 0.149 fue computada, aunque en general pareció ser más adecuada para densidades medias de 3.0 a 55.0 que para densidades de 0.5 a 3.0 ácaros por cm². Para densidades promedio de 0.5 a 55 por cm², el parámetro k de la binomial negativa estuvo relacionado con la densidad promedio (\bar{x}): $k = 0.081 + 0.1139 * (\log_{10} \bar{x})$. Los valores estimados de k fueron usados para calcular los perfiles de los datos de conteo esperados para varias densidades medias en el rango de 1 a 40 ácaros por cm². Debido a la desviación de los datos de los conteos, el número de ácaros por cm² esperado en la mayoría de las muestras individuales fue siempre considerablemente menor que la densidad promedio. Tomando como base la binomial negativa, las densidades medias de ácaros podrían ser estimadas a partir del porcentaje de muestras con al menos un ácaro. Los resultados del estudio proveen medios para predecir

el histograma de frecuencia relativa de densidades asociadas con una densidad promedio de *P. oleivora* por cm² en un área con árboles.

Average densities of the citrus rust mite (CRM) [*Phyllocoptruta oleivora* (Ashmead)] per cm² on fruit across an area of orange trees can be estimated from counts of the number of mites present within a one-cm² surface area per fruit (Hall *et al.* 1994). The number of fruit and trees that must be sampled depends upon both the desired precision of estimates and the density of mites at which this precision is required.

If the probability distribution (e.g., see Gomez & Gomez 1984) associated with CRM count data is known, the frequency histogram of individual counts associated with a particular mean density can be projected. This would be useful for projecting damage by a CRM population and for establishing control levels. Histograms of CRM counts taken within individual trees usually followed the negative binomial probability distribution (Hall *et al.* 1991). No information was available on probability distributions describing CRM count data from fruit over an area of trees.

We had a considerable amount of CRM count data from fruit samples taken across 4-ha (10-acre) areas of 'Hamlin' and 'Valencia' orange trees in Florida (Hall *et al.* 1994). Previous analyses of the data indicated that the counts usually followed an aggregated distribution (Hall *et al.* 1994). Because aggregated dispersions often follow the negative binomial probability distribution (Southwood 1978), and because CRM count data from individual trees usually followed the negative binomial, we evaluated this distribution for projecting the frequency histograms of our count data.

MATERIALS AND METHODS

Count data were obtained on the number of CRM per cm² on fruit across 32 4-ha blocks of 'Hamlin' and 'Valencia' orange trees using a transect sampling plan (Hall *et al.* 1994). This plan consisted of 192 1-cm² samples per block - two samples per fruit, four fruit per tree (1 from each compass quadrant), 12 trees along one transect between the northeast and southwest corners of the block, and 12 trees along a second transect between the northwest and southeast corners of the block. All CRM except eggs within a 1-cm² sample were counted using a 10X magnifier fitted with a 1-cm² grid of 25 equal-sized subdivisions. In cases where >35 CRM per cm² were present, the number of mites was sometimes estimated by counting the number of mites in a diagonal row of five grid subdivisions and multiplying by 5. The block samples were taken during May through December within several different citrus growing areas in Florida. The only treatment applied to the blocks during the study was a summer spray of copper and oil. No samples were taken until at least 6 wk after this treatment.

The negative binomial probability distribution is characterized by two parameters, the mean (\bar{x}) and a coefficient k (Johnson & Kotz 1969). The value of the k parameter defines the shape of the negative binomial distribution and serves as a general indicator of aggregation, with smaller values of k indicating increased aggregation (Southwood 1978). An iterative solution was used to manually estimate k for each block averaging at least 0.5 CRM per cm² (25 blocks):

$$N \cdot \log_e \left[1 + \frac{\bar{x}}{k} \right] = \sum \left[\frac{Ax}{k+x} \right] \quad (1)$$

with N = total number of samples, \log_e = natural logs, and Ax = the sum of all frequencies of sampling units containing more than x individuals (Bliss & Fisher 1953, Southwood 1978).

The 25 k estimates were then evaluated using regression procedures presented by Bliss & Owen (1958) and Bliss (1958) to determine if a single, *common* k (k_c) existed. This involved regressing two statistics for each block, $y' [= s^2 - \bar{x}]$ on $x' [= \bar{x}^2 - (s^2/N)]$, where \bar{x} was the mean, s^2 the variance and N the number of individual counts per block. The regression was forced through the origin, and k_c was estimated from the inverse of the slope of the regression. The adequacy of this k_c estimate was evaluated using a regression analysis of $1/k$ on $\log_{10}(\bar{x})$: a trend between these variables discredits the suitability of a single k_c (Bliss & Owen 1958, Southwood 1978).

One way to write negative binomial probabilities is:

$$p_x = \frac{(x+k-1)!}{x!(k-1)!} \cdot \left[\frac{k}{k+\bar{x}} \right]^k \cdot \left[\frac{\bar{x}}{k+\bar{x}} \right]^x \quad x = 0, 1, 2 \dots \quad (2)$$

where p_x is the probability of a sample having x mites (Williamson & Bretherton 1963). To determine the histogram of CRM counts expected in each block according to the negative binomial, we used observed mean densities and estimates of k in the following iterative probability formula:

$$p_{x+1} = \frac{k+x}{x+1} \cdot \left[\frac{\bar{x}}{k+\bar{x}} \right] \cdot p_x \quad x = 0, 1, 2 \dots \quad (3)$$

with the probability of no mites ($x=0$) being

$$p_{x=0} = \left[\frac{k}{k+\bar{x}} \right]^k$$

where p_x = the probability of a sample containing x mites. Equation (3) was obtained by writing successive terms for p_0, p_1, p_2, \dots from equation (2) and noting the common multiplier. Using this iterative method avoids brute force calculation of the combinatorials which often cause computer overflow for large values of x . We programmed SAS (SAS Institute Inc. 1990) software to compute the successive probabilities. Chi-square tests ($\alpha = 0.05$) according to guidelines presented by Gomez & Gomez (1984) and correlation analyses were used to test the fit of the observed CRM counts to those expected under the negative binomial based on estimated k -values.

RESULTS AND DISCUSSION

The mean density of CRM observed in the 25 blocks ranged from 0.5 to 112.5 per cm^2 . A regression analysis indicated that the maximum density of CRM (\bar{y}) observed in each block could be estimated from the mean density (\bar{x}): $y = 32.5 + 17.3^*x$; $r^2 = 0.85$, $F = 123.1$, $PR > F = 0.0001$, $d.f. = 23$.

Individual estimates of the negative binomial k ranged from 0.0199 to 1.58 ($\bar{k} = 0.2147$, $s = 0.3025$) (Fig. 1). With respect to investigations into k_c , an initial plot of y' on x' indicated that one data point clearly deviated from the main trend of the regression (Fig. 2). This data point, which was associated with a mean density of 112.5 CRM per cm^2 and a k value of 1.58, was excluded from further investigations into k_c but indicated that CRM aggregation may substantially decrease as population densities increase to as high as 100 or more CRM per cm^2 .

Among the 24 sets of count data retained for k_c determination, mean densities ranged from 0.5 to 54.9 CRM per cm^2 ($\bar{x} = 12.96$ per cm^2 , $s = 16.9$). The individual k estimates for the blocks varied from 0.0199 to 0.3580 ($\bar{k} = 0.158$, $s = 0.1052$). A k_c of 0.149 was calculated ($F = 141.7$, $\text{Pr} > F = 0.0001$, $r^2 = 0.86$, d.f. = 24) (r^2 corrected for the mean = 0.83, d.f. = 23) (Fig. 2). A statistically insignificant relationship ($\alpha = 0.05$) was found between $1/k$ and \bar{x} number CRM per cm^2 , but a weak relationship ($r^2 = 0.395$) was found between $1/k$ and $\log_{10}(\bar{x})$ number CRM per cm^2 (Fig. 3), which indicated the k_c of 0.149 may not have been a suitable substitute for all of the individual k -values. A similar problem was reported with respect to determining a k_c associated with a set of wireworm counts (Bliss & Owen 1958). Variability in individual k -values associated with small mean CRM densities (e.g., 0.5 to 3.0 CRM per cm^2) was responsible for this trend; no significant trend was found between $1/k$ and $\log_{10}(\bar{x})$ number CRM per cm^2 among mean densities of from 3 to 55 CRM per cm^2 ($N = 15$), and the same k_c (0.149) was calculated across these densities.

Because k_c tended to be a poor substitute for individual k -values at mean densities below around 3.0, as an alternative to k_c we conducted a regression analysis and determined an equation for estimating k across different mean densities (\bar{x}): $k = 0.081 +$

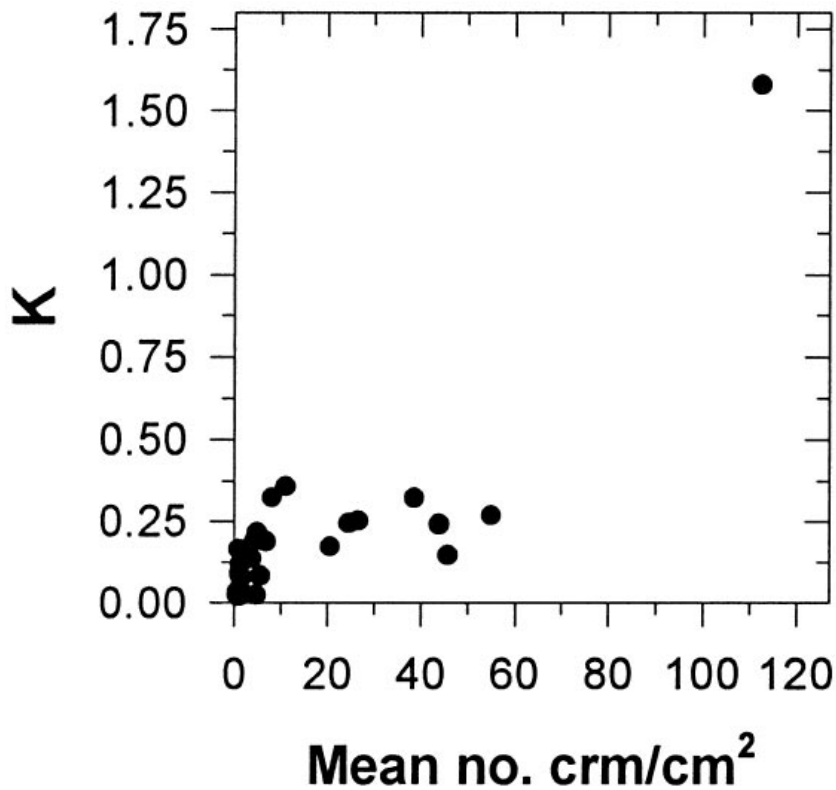


Fig. 1. Negative binomial k -values associated with observed mean densities of citrus rust mites (crm) per cm^2 on fruit across 4-ha areas of orange trees.

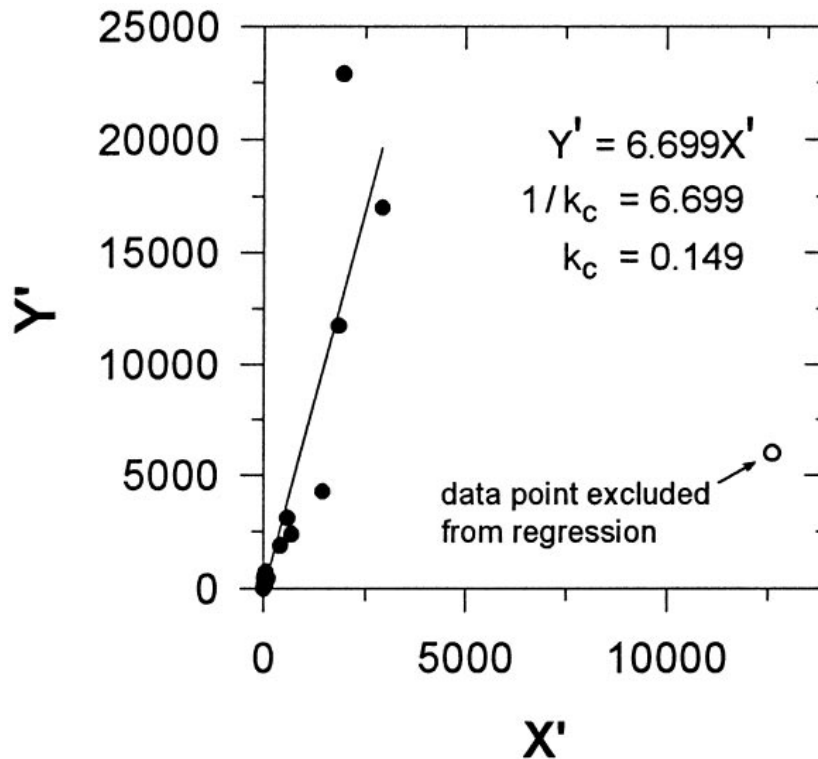


Fig. 2. Regression analysis used to calculate a single, common k (k_c) for a negative binomial distribution describing the number of citrus rust mites per cm^2 on fruit across 4-ha areas of orange trees. The open circle represents a data point excluded from the regression; this data point was associated with a mean of 112.5 mites/ cm^2 and a k of 1.58.

0.1139*($\log_{10} \bar{x}$); $F = 27.11$; $\text{Pr} > F = 0.0001$; $r^2 = .55$; d.f. = 23. A comparison of some histograms generated from individual, common and regressed k -values is presented in Fig. 4.

Chi-square tests indicated that CRM counts across a 4-ha area of trees followed the negative binomial distribution in 19 of 24 (79%) areas based on individual k -values and in 16 of 24 (67%) areas based on either k_c or regressed k -values. Among the observed count histograms that did not follow the negative binomial based on chi-square tests, these histograms visually resembled the distribution (e.g., Fig. 4). Over all 24 sets of CRM count data, the correlation between observed counts and counts projected using the negative binomial was 0.993 based on individual k -values, 0.976 based on regressed k -values, and 0.965 based on the k_c estimate. Among the 24 count sets, the lowest correlation between observed counts and counts expected under the negative binomial was 0.929, 0.921 and 0.872 based on individual k -values, regressed k -values and k_c , respectively.

Overall, counts of the number of CRM per cm^2 on fruit across a 4-ha area of trees appeared to be at least reasonably described by the negative binomial distribution when mean densities were in the range of 0.5 to 55 per cm^2 . The distribution in con-

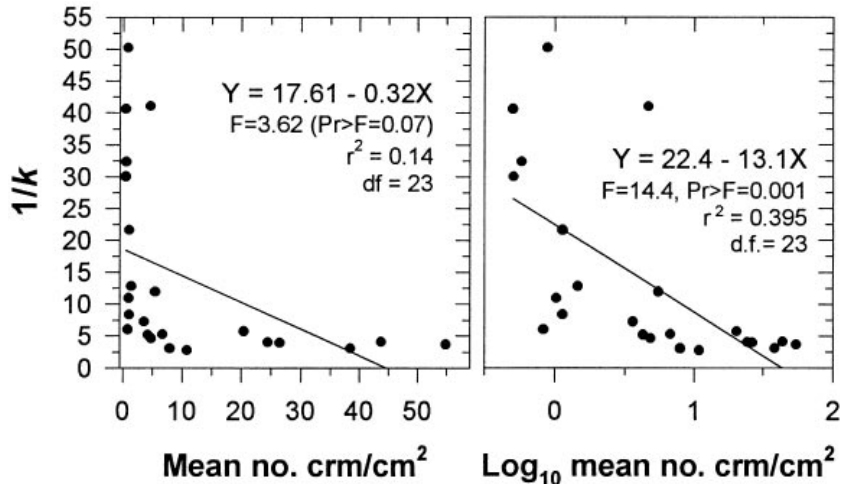


Fig. 3. Relationship between $1/k$ and \bar{x} compared to $1/k$ and $\log_{10}\bar{x}$ number of citrus rust mites (crm) per cm^2 on fruit across a 4-ha area of orange trees.

junction with the k parameter could therefore be used to project the frequency histogram of CRM densities at any mean density within this range, which in turn could be used in combination with models projecting how much surface damage to fruit a given density of CRM will cause (e.g., see Allen 1976, Yang *et al.* 1995) for an overall estimate of damage a CRM population will cause. The distribution could also be used to develop a sequential sampling plan (Southwood 1978), which might reduce the cost of sampling CRM. While histograms projected based on k_c were similar to those based on regressed k values, overall our analyses favored histograms based on the regressed estimates. As a word of caution, k values and the goodness-of-fit of the negative binomial distribution could be negatively influenced by extraneous factors that affect mite dispersion in a grove (e.g., chemical applications).

Expected profiles of CRM counts for a number of mean densities ranging from 1 to 40 mites per cm^2 were projected based on the negative binomial using regression estimates of k in formula #3 (Fig. 5). Differences were relatively small between means of 1 to 40 mites per cm^2 with respect to the projected probability of any individual count in the range of 5 to 15 mites per cm^2 . Due to the skewness of CRM count data, the number of mites per cm^2 expected in most individual samples was always considerably smaller than the average density. For example, at an average density of 5 CRM per cm^2 , fewer than 5 mites per cm^2 were expected to be present in around 80% of the individual samples across a 4-ha area. This information would be important to a citrus grower who might mistakenly assume that an average density based on scouting data reflects the midpoint of densities present. As the mean density increased, the probability increased that any particular large count would be observed. For example, the expected percentage of counts above 30 mites per cm^2 increased from about 4% at a mean of 5 mites per cm^2 up to about 18% at a mean of 20 mites per cm^2 . The skewness of CRM count data supported contentions made by McCoy *et al.* (1976), namely that a control threshold should take into consideration the frequency histogram of mite counts.

Given that count data follow the negative binomial and k is known, expected mean densities can be estimated from the percentage of samples containing at least one an-

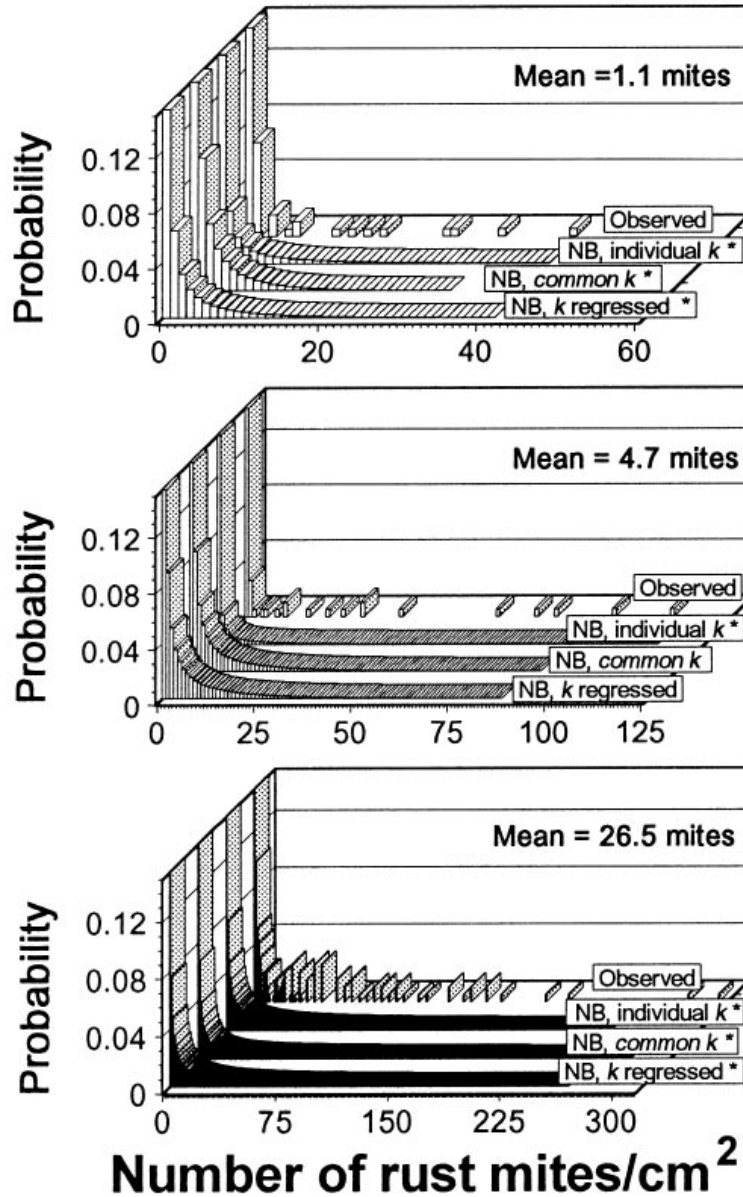


Fig. 4. Probability (proportion) of observed counts of citrus rust mites per cm² compared to counts projected from negative binomial (NB) distributions derived from individual, common and regressed k -values (largest observed counts shown, probabilities below 0.0001 or above 0.15 not shown). An asterisk (*) indicates the observed histogram followed the projected histogram based on a chi-square test ($\alpha = 0.05$).

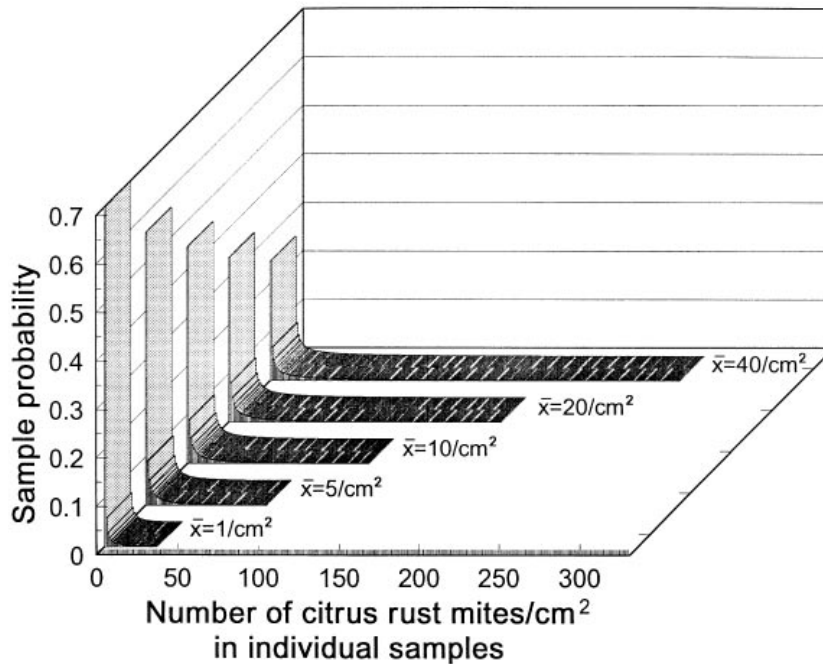


Fig. 5. Projected frequency distributions for citrus rust mite counts per cm^2 on fruit across a 4-ha area of orange trees based on the negative binomial and regressed k -values, means of 1, 5, 10, 20 and 40 mites per cm^2 (class probabilities below 0.00025 and above 0.7 not shown). The zero-class probability extended up to 0.81 at a mean of 1 mite per cm^2 .

imal (see Southwood 1978). The relationship between mean CRM density and the percentage of infested samples based on the negative binomial is presented in Fig. 6. Similar relationships between percentages of infested samples and mean CRM densities have been observed without the use of a probability distribution (Knapp & Fasulo 1983, McCoy *et al.* 1976). The percentage of samples infested became increasingly poorer as an indicator of mean density as CRM densities increased. Benefits and precautions associated with using the percentage of infested samples as an indicator of mean CRM densities have been discussed (McCoy *et al.* 1976).

ACKNOWLEDGMENTS

The authors acknowledge and thank Clay McCoy, Jorge Peña and Phil Stansly for their constructive reviews during the preparation of this manuscript. This article is Florida Agricultural Experiment Stations Journal Series No. R-05232.

REFERENCES CITED

- ALLEN, J. C. 1976. A model for predicting citrus rust mite damage on Valencia orange fruit. *Environ. Entomol.* 5: 1083-1088.

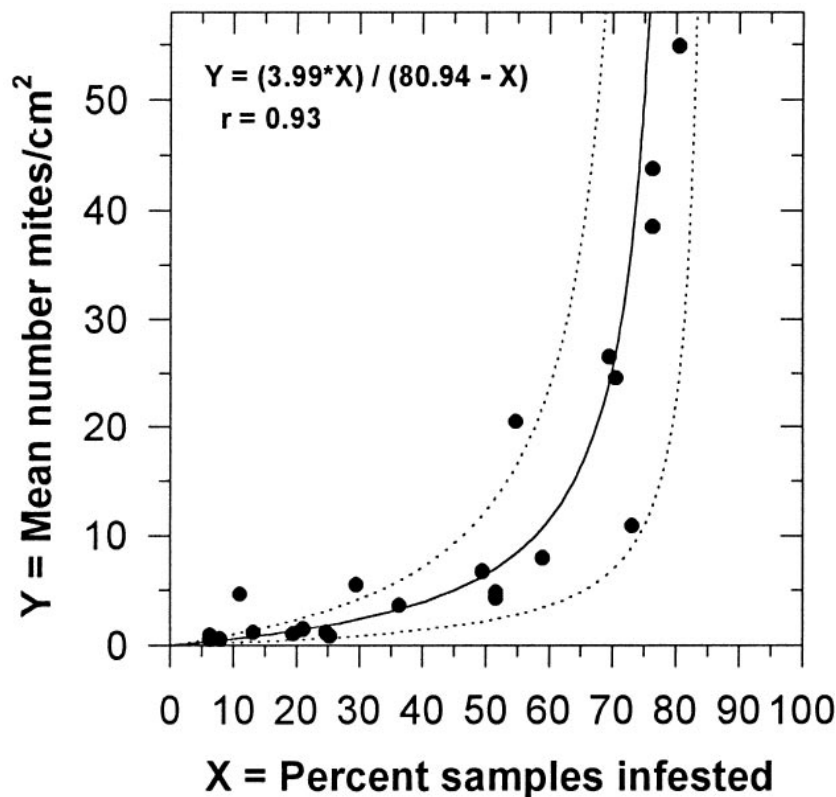


Fig. 6. Relationship between the mean density of citrus rust mites per cm^2 on fruit across a 4-ha area of trees and the percentage of 1- cm^2 samples with at least one rust mite. Parameter estimates (standard error) associated with the hyperbola were 80.94 (1.166) and 3.994 (0.1905). A 95% confidence interval is given by the dotted lines. Percent samples infested (x) can be estimated from the mean number of mites/ cm^2 (y) by: $x = (80.94 * y) / (3.99 + y)$.

- BLISS, C. I. 1958. The analysis of insect counts as negative binomial distributions. *Proc. X Int. Congr. Entomol.* 2: 1015-1032.
- BLISS, C. I., AND R. A. FISHER. 1953. Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. *Biometrics*. 9: 176-200.
- BLISS, C. I., AND A. R. G. OWEN. 1958. Negative binomial distributions with a common k . *Biometrika*. 45: 37-58.
- GOMEZ, K. A., AND A. A. GOMEZ. 1984. *Statistical Procedures for Agricultural Research*, 2nd ed. Wiley & Sons, New York. 680 pp.
- HALL, D. G., C. C. CHILDERS, AND J. E. EGER. 1991. Estimating citrus rust mite (Acari: Eriophyidae) levels on fruit in individual citrus trees. *Environ. Entomol.* 20: 382-390.
- HALL, D. G., C. C. CHILDERS, AND J. E. EGER. 1994. Spatial dispersion and sampling of citrus rust mite (Acari: Eriophyidae) on fruit in 'Hamlin' and 'Valencia' orange groves in Florida. *J. Econ. Entomol.* 87: 687-698.

- JOHNSON, N. L., AND S. KOTZ. 1969. Discrete Distributions. Houghton Mifflin Co., Boston. 328 pp.
- KNAPP, J. L., AND T. R. FASULO. 1983. Citrus rust mite. *In* Florida citrus integrated pest and crop management handbook. SP-14, Florida Coop. Ext. Serv., IFAS, University of Florida, Gainesville.
- MCCOY, C. W., R. F. BROOKS, J. C. ALLEN, AND A. G. SELHIME. 1976. Management of arthropod pests and plant diseases in citrus agroecosystems. Proc. Tall Timbers Conf. Ecol. Animal Control Habitat Management. 6: 1-17.
- SAS INSTITUTE INC. 1990. SAS Language: Reference, Version 6, First Edition. Cary, NC. 1042 pp.
- SOUTHWOOD, T. R. E. 1978. Ecological Methods, 2nd ed. Wiley/Halsted, New York. 524pp.
- WILLIAMSON, E., AND M. H. BRETHERTON. 1963. Tables of the Negative Binomial Distribution. John Wiley & Sons, N.Y. 275 pp.
- YANG, Y., J. C. ALLEN, J. L. KNAPP, AND P. A. STANSLY. 1995. Relationship between population density of citrus rust mite (Acari: Eriophyidae) and damage to 'Hamlin' orange fruit. Environ. Entomol. 24: 1024-1031.

