Acta entomologica serbica, Supplement, 2006, 89-96 UDC 595.771 : 591.34] : 574

SAMPLING AND STATISTICS IN UNDERSTANDING DISTRIBUTIONS OF BLACK FLY LARVAE (DIPTERA: SIMULIIDAE)

JOHN W. MCCREADIE[1], PETER H. ADLER[2], MARIA EUGENIA GRILLET[3] AND NEUSA HAMADA[4]

[1]Department of Biological Sciences, Life Sciences Building, University of South Alabama, 307 University Blvd., Mobile, AL 36688-0002, U.S.A.
[2] Department of Entomology, Soils, and Plant Sciences, 114 Long Hall, Clemson University, Clemson, SC 29634-0315, U.S.A.
[3]Laboratorio de Biologia de Vectores, Instituto de Zoologia Tropical, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1041-A, Venezuela
[4]Instituto Nacional de Pesquisas da Amazonia-Entomologia, 69011-970 Manaus, AM, Brazil.

ABSTRACT – We demonstrate by example a standardized set of protocols for collecting both larval black flies (Diptera: Simuliidae) and the accompanying stream data. By following a few simple procedures, powerful data sets can be readily collected. The protocols have been developed as a result of sampling over 500 streams in both North and South America. In addition, simple statistical procedures for producing species distribution models also are presented.

KEY WORDS: Black flies larvae, sampling, statistics, distribution, North America, South America

INTRODUCTION

The underpinning of any study in community ecology is the appreciation that patterns of species distribution vary with scale (SCHNEIDER, 1994). By extension, the patterns of species distributions appear to offer an effective means to determine the underlying mechanisms that structure species assemblages (or communities), especially in circumstances where experimentation is impractical or impossible (MCARDLE, 1996; GASTON AND BLACKBURN, 2000). However, in order to identify the mechanisms responsible for either the distribution of a single species or the structuring of an entire species assemblage, one must first clearly delineate the scale of study.

Larval blackfly distribution has been studied from the scale of the substrate (Colbo and Wotton, 1981; Craig and Galloway,1987; Mccreadie and Colbo, 1993) tothat of zoogeographic regions (Adler et al., 2004; Mccreadie et al., 2005). At the level of the substrate, velocity and water depth have been considered important factors influencing larval distribution (Craig and Galloway, 1987). In streams punctuated by outlets or other types of impoundments, larval abundance can vary significantly over a few meters downstream from the outlet (Adler and McCreadie, 1997). By combining velocity, depth, and distance from an outlet, McCreadie and Colbo (1993) were able to demonstrate that *Simulium truncatum* Lundström and *S. rostratum* Lundström, over a 10 - 20 m section of stream, showed species specific distribution patterns. Although not rigorously defined, the distribution of larvae at this scale is often referred to as the distribution with a microhabitat (Adler and McCreadie, 1997). Not only does abundance change over the first few meters from a lake outlet (Wotton 1979; Richardson and Mackay, 1991), but species composition can vary dramatically over a scale of tens of meters (McCreadie and Colbo, 1991).

The next scales of study, and the primary focus of this paper, include the distribution of species between sections of a stream and between streams. Traditionally, such scales of study for blackflies have been on the scale of kilometers to tens of kilometers (McCreadie and Colbo, 1991; 1992; Adler and McCreadie, 1997; McCreadie and Adler, 1998). At these scales, a variety of factors are correlated with both species richness and the distribution of individual species. Such factors include stream size, the presence/absence of lake outlets, stream bed, water chemistry, substrate availability, elevation, and landscape characteristics (McCreadie and Alder, 1998; Adler et al., 2004; McCreadie et al., 2004, 2005).

The intent of this paper is to provide a standardized larval sampling protocol using a small data set from Venezuela. Then goal is to demonstrate by example our standardized method of collecting both larval and stream data so that these data can be compared among various scales and across different laboratories. By following a few simple procedures, powerful data sets can be readily collected. The protocols discussed below have been successfully used on over 500 streams sampled in both North and South America. Equally important, by following a ‘standarized’ method of stream characterization and larval collection, data from various researchers in different locations can be pooled, thereby permitting the analysis of larval distribution over large scales.

Although the protocols were largely developed for streams less than 100 m in width, with proper modification such protocols also work for larger rivers. In addition, simple statistical procedures useful for producing species distribution models also are presented.

MATERIALS AND METHODS

The study area included streams (05.06° - 06.53° N, 66.98° - 67.78° W) in the vicinity of Puerto Ayacucho, Amazonas, Venezuela. Streams in this area flow into the Orinoco River.

Sampling Protocols

The variables selected in our ‘stream characterization’ have been shown to be useful predictors of blackfly larvae along stream reaches (McCreadie and Colbo, 1992; McCreadie and Adler, 1998; McCreadie et al., 2005). Additional variables can also be added for the specific needs of different studies. We strongly suggest that among other variables to be measured, GPS reading should be included.

Stream width (w) can be measured either with a meter tape or by simply 'pacing' the stream; this measurement is taken at a single point and will also serve as a plumb line in which to measure water depth and velocity. Stream depth, in meters, is measured at equal distance points along the plumb line with a stout steel ruler. Our rule of thumb is to obtain two (streams < 1 m in width) to five (streams up to 100 m in width) measures of depth. Mean depth is then calculated and designated as $\mathbf{D_1}$. At the same location where water depth is measured, water velocity can be estimated by moving the steel rule at right angles to the stream bank and noting the distance water moved up the wide face of a steel ruler. The mean is calculated and designated as $\mathbf{D_2}$, and water velocity is then estimated using the formula $U = \sqrt{2g(\mathbf{D_2} - \mathbf{D_1})}$, where U = water velocity ($U\ ms^{-1}$) g = the force due to gravity ($9.8\ ms^{-2}$). In some streams, it is not practical to use this method, for example in large fast rivers or small trickles. In such cases, the time a cork takes to move a prescribed distance can be employed. HAMADA AND MCCREADIE (1999) showed that both methods produced similar estimates of velocity ® = 0.931, $P < 0.001$). Depth ($\mathbf{D_1}$), width (w) and velocity (U) measurements are then used to estimate discharge (i.e., $Q\ m^3s^{-1} = \mathbf{D_1} \times w \times U$).

Conductivity (µS/cm @ 25°C, a measure of the total amount of ions in the water), pH, and water temperature in our studies are measured using a combined digital measure. We generally use the Cole-Parmer pH/CON 10 models. Based on measurements from 297 streams throughout North America, the mean percent saturation of oxygen was 97.5%, i.e., in most cases streams were near or often exceeded oxygen saturation. This variable therefore is not critical to measurements and accordingly is not included here.

Stream bed particle size, riparian vegetation, and canopy cover can be visually estimated following the protocols of MCCREADIE AND COLBO (1991). Streambed particle size can be classed as mud/silt, sand, small stones, rubble, boulders, and bedrock (Table 1). Although a variety of particle sizes can occur, one of these types usually predominates; thus, one measures the dominant stream bed particle size. Ranked measurements of the dominant substrate (Table 1) can then be used in statistical analyses. Table 2 is used to determine the dominant riparian vegetation. Here the emphasis is on 'form', not species identification. To estimate canopy cover over the stream, we used the following guidelines. If stream-side vegetation from opposite sides of the bank touch over the stream or if vegetation is found only on one side of the bank but it extends to the other side, the canopy is complete. If stream side vegetation extends over less than 10% of the stream, it is open.

Table 1. Classification of streambed particle size used by MCCREADIE AND COLBO (1991) and the ranking system for statistical analysis.

Category	Particle diameter (mm)	Ranking
Mud/silt	< 1	1
Sand	1-2	2
Small stone	2-32	3
Rubble	32-256	4
Boulder	> 256	5
Bedrock	-	6

Table 2. Classification of riparian vegetation used by McCreadie and Colbo (1991) and the ranking system for statistical analysis.

Category	Vegetation form	Ranking
Open	pasture, grassland, bogs, meadows	1
Brush	extensive herbaceous growth, scattered trees, saplings etc.,	2
Forest	continuous border of trees along stream banks	3

Otherwise, the canopy is partial. Ranked measurements of riparian vegetation type (open, brush, forest; ranked 1-3), and canopy cover (open, partial, complete; ranked 1-3) can also be used in statistical analyses.

Each site was sampled by hand-collecting larvae from all available natural substrates. The rationale for this sampling protocol is given in McCreadie and Colbo (1991). The intent is to collect a minimum of 30 specimens at a site. Also, time spent collecting at a site should be noted (see below). As in similar studies (McCreadie, 1991; McCreadie et al., 1995), it was assumed that species found in the hand-collected sample from each site were representative of local occurrences.

Either implicit or explicit in studies examining the distribution of blackfly larvae among stream types is the idea of equal sampling effort in terms of area or time among the local collections to be compared. This ideal situation is unfortunately not possible under most if not all sampling protocols. For example, equal-timed collections among stream sites, which are supposed to represent equal sampling effort, in fact do not. An investigator sampling a shallow, slow-moving 1-m wide stream over a standard period of time would sample more habitat (and thus more specimens) than in a fast-flowing, silt-bottomed, 100-m wide stream over the same period of time. Furthermore, there is no artificial substrate that will work equally well among all stream types for blackflies (Colbo, 1987; Adler et al., 2004). Hence, to reduce bias in handing collection methodologies, the following approach is useful. First, correlate collection times at each sites with the number of species collected (i.e., species richness). If a significant correlation is found, start removing samples one at a time at the extremes, i.e., the samples with longest and shortest sampling times, while recalculating the correlation at each removal. Continue until the correlation between species richness and sampling time is nonsignificant and use these remaining samples for any further analysis. Another 'rule of thumb' is to use only those sites from which a minimum of 30 larvae were found (McCreadie and Adler, 1998).

Statistical Analysis

Regression analysis is a useful way to link the distribution of individual species to stream site conditions. Because larval abundance is at best measured only semi-quantitatively in any blackfly study, the response variable should be species presence/absence. It has been shown (Corkum, 1989; McCreadie and Adler, 1998; Feminella, 2000) that binary data from single-point collec-

tions are robust enough to detect faunal differences among streams. Binary data can also avoid certain problems associated with abundance data. Abundance depends not only on the abiotic conditions of the habitat, but also on species interactions such as competition, disease, predation, and symbiosis. Therefore, changes in abundance reflect not only abiotic characteristics of the habitat (accounted for in this study), but also constantly fluctuating biotic factors (not taken into account), which represent a source of unaccounted error.

Accordingly, each species at each site was recorded as a binary variable (0 = species absence, 1 = species presence in a sample). Forward logistic multiple regression was used to estimate the probability of a species being present at a site, given measured site conditions. Under the logistic function, $p_i = e^L / (1 + e^L)$, and $L = B_0 + BX_{1i} + ...B_j X_{ji,}$ where p_i is the probability that a species is present at the i-th site, $X_{1i} .....X_{ji}$ are the predictor (independent) variables, and $B_1....B_j$ are the regression coefficients for the linear combinations of predictors. Entrance of each variable into the model was arbitrarily set at $p = 0.15$, with the significance of each predictor ($p < 0.05$) assessed using maximum likelihood estimation (HOSMER AND LEMESHOW, 1989). Because of the low number of sites examined (n = 15), only the four most commonly encountered species were subject to regression analysis. In studies with a larger number of sites, it was noted that using species occurring at a frequency of less than 20% resulted in regressions with a lack of power (type II error) due to the large number of zero values (MCCREADIE ET AL., 2005). Thus, uncommon species were not considered in this type of analysis. Concordance was used to assess the fit of regression models to the observed data. Concordance pairs all values of the response variable that are different (i.e., 0,1), and then counts the number of times that the member of a pair with the higher predicted probability of a species being present was correct (SAS, 1987). Results are expressed as a proportion of the total number of pairs compared, with the higher the percent, the stronger the model. Loosely speaking, it can be viewed in the same manner as is the R^2 value is when using ordinary least squares regression.

Stream variables are often highly intercorrelated and such multicolinearity affects confidence intervals and significance tests of regression coefficients; hence, the resulting models are unreliable (NETER ET AL. 1990). To avoid problems associated with multicolinearity, principal component analysis (PCA) was used to transform stream variables into a set of statistically independent principal components or PCs (MCCREADIE AND ADLER, 1998; QUINN AND KEOUGH, 2002). The use of PCA can also allow broader ecological interpretations of habitat variables (MCCREADIE AND ADLER, 1998; HAMADA ET AL., 2002). Thus, PCs with eigenvalues greater than 1.0 (NORUSIS, 1985), replaced the original stream variables as predictors in the regression analyses above. Variables not normally distributed are subjected to appropriate transformations ($\log_{10}$, square root, etc.) before entering a PCA. It has been our experience that the easiest way to interpretate PCs is to correlate each PC with the original stream variables (MCCREADIE AND ADLER, 1998). To produce a rigorous interpretation, we interpret the meaning of any particular PC in relation to only those stream variables in which the correlation is significant at $p < 0.01$.

RESULTS

No significant correlation between collection times (15 - 75 min) and species richness (1 - 5 species per site) was found ($r = -.098$, $p = 0.729$). We therefore concluded that there was no gross bias in collecting methodology. The most frequently collected species were *Simulium quadrifidum* Lutz (11 sites), *Simulium subpallidum* Lutz (eight sites), *Simulium maroniense* Floch & Abonnenc

(eight sites), and *Simulium incrustatum* Lutz (six sites). The mean (± 99% CI) number of species per site was 3.3 ± 0.7.

The range of all stream variables measured is given in Table 3, as is the PCA. Three principal components had eigenvalues > 1.0 and together accounted for 73.1 % of the variability among sampling stations. In the first place, PC-1 (largely a measure of channel conditions and vegetation cover) explained 34% of the intersite variability in stream conditions. Higher PC-1 values (or scores) identified faster, smaller streams with less riparian vegetation than along streams with lower PC-1 values. On the other hand, PC-2 (accounting for an additional 23% of site variability) was largely a measure of stream size. Sites with higher PC-2 values were smaller (as indicated by both

Table 3. Results of PCA and Spearman's rank correlation analysis of stream variables and derived principal components (PCs) for all collections (n = 15) taken during October, 1996.4.1

	Stream Sites			Principal components		
Variables	Range		Mean (± SE)	PC-1	PC-2	PC-3
Velocity (m s^{-1})	0.62 -	1.77	0.99 ± 0.10	0.688*	-0.328	0.076
Discharge (m3 s^{-1})	0.15 -	112.05	9.42 ± 7.34	0.175	-0.870*	-0.032
Depth (m)	0.05 -	1.75	0.50 ± 1.40	-0.414	-0.849*	0.079
Width (m)	1.7 -	21.0	6.8 ± 0.19	-0.657*	-0.011	0.265
Temperature °C	25.1 -	28.3	26.1 ± 0.23	-0.451	0.502	0.371
pH	3.9 -	5.7	4.6 ± 0.12	-0.436	-0.228	0.798*
Conductivity (µS cm-1, 25°C)	4.1 -	17.2	8.5 ± 1.01	-0.540	0.059	0.662*
Stream bed [1]	sand -	bedrock		-0.632	-0.461	0.343
Riparian vegetation [1]	open -	forest		-0.806 *	-0.373	-0.271
Canopy [1]	none -	complete		-0.752*	-0.243	-0.332
% variance explained						
Proportion				34.0	23.0	16.0
Cumulative				30.7	57.1	73.1

[1] Ranked variables: stream bed 1 - 6, riparian vegetation 1 - 3; canopy 1- 3. Rankings followed McCreadie & Colbo 1991.
* $p < 0.01$;

lower discharge and shallower depth) than streams with lower PC-2 values (Table 3). Finally, PC-3 accounted for another 16% of site variability. This principal component was a measure of water chemistry, sites with higher PC-3 values identifying less acidic waters with higher conductivity than streams with lower PC-3 values. It is noted that neither temperature nor streambed was strongly associated with any of the three PCs considered in this analysis.

Two species, *S. subpallidum* and *S. incrustatum* showed a significant association with PC-3. Regression models were as follows:

S. subpallidum

$p_i = e^{L} / (1 + e^{L})$,

where L = 67 + 172PC-3, G = 20.73, df = 1, $p < 0.001$, concordance = 100%

S. incrustatum

$p_i = e^{L} / (1 + e^{L})$,

where L = -0.7 + 1.5PC-3, G = 7.06, df = 1, p = 0.008, concordance = 83%

In the case of *S. subpallidum*, concordance indicates that we were able to predict those streams with and without this species with 100% accuracy. The positive coefficient in front of the PC-3 term ($B_1 = 172$) indicates that this species was found in streams with higher PC-3 values, i.e., at sites that were less acidic and had higher conductivity. As for *S. incrustatum*, it also had a positive coefficent in front of the PC-3 term ($B_1 = 1.5$), indicating that it too was most often found in less acidic streams with higher conductivity. In this case, we were able to predict the presence or absence of *S. incrustatum* with 83% accuracy.

DISCUSSION

The results of this study are clear, the larval distribution of two of the four simuliid species examined in the Amazonas region of Venezuela were strongly associated with stream conditions. Given these results and those of previous works (McCreadie et al., 1995; McCreadie and Adler, 1998; Hamada and McCreadie, 1999; Hamada et al., 2002; McCreadie et al., 2004, 2005; McCreadie and Adler, 2006) in both North and South America, we have consistently shown that the distributions of most blackfly larvae so far investigated are strongly associated with stream conditions. In addition, the distribution of each of these species can be modelled with a fair degree of accuracy using relatively simple techniques.

The primary purpose of this paper was to demonstrate a standard protocol for measuring stream variables and collection of larvae using a small data set and simple yet powerful statistical techniques to examine the distribution of simuliid larvae across sampling locations. By following this protocol, distribution patterns of species, structures of species assemblages, and changes in diversity can examined and modelled on scales ranging from the watershed to the continental biosphere. (Adler and McCreadie, 1997; McCreadie and Adler, 1998; Adler et al., 2004; McCreadie et al., 2004, 2005). Futhermore, data collected from different laboratories can be pooled and used in analysis on scales ranging from the stream reach to global distribution patterns.

In any study of larval distribution, there are implicit and/or explicit caveats or assumptions that must be made in order to set the boundaries of intrepretation of the results. It is instructive to discuss several of these assumptions and caveats in relation to the protocols considered in the current paper. First, any particular species of larvae occurs in any particular stream because the female chooses to oviposit in that particular location (McCreadie, 1991; McCreadie et al., 2005). Investigators rarely considered this implicit fact when researching the distribution of blackflies, and this most likely represents a significant source of error in larval distribution models. One could argue that the most evolutionarily stable strategy (ESS) for an adult female would be to oviposit in streams where the likelihood of larval survival is maximized. However, as our knowledge of simuliid ovipositional cues is rudimentary (Crosskey, 1990), investigators have used easily measured stream variables as predictors of larval occurrence. Given our poor understanding of ovipositional cues, there is no compelling evidience that cues used by females are strongly correlated with the conditions the larvae would experience. Again, the only aurgement for such a correlation would be ESS. Given our current state of knowledge of oviposition behavior, we are left with treating the larval distribtution in 'isolation' from the adult. However, it should be kept in mind that studies on larval distribtuion are an indirect study of female habitat selection.

In our studies of local larval species assemblages, we have assumed that all species used in the analysis were capable of reaching any stream site under consideration; that is, species were not

precluded from stream sites as a result of dispersal abilities. This assumption is appears to be reasonable given the widespread distribution of the species used in our analysis, the limited distances between sites, and the known dispersal abilities of many blackflies (Adler et al., 2004).

However, at higher levels of study (such as study of the differences of assemblages between floristic provinces or between mainland territory and oceanic islands), such a blanket assumption is not possible.

REFERENCES

Adler, P. H., D. C. Currie & Wood, D. M. 2004. The Black Flies (Simuliidae) of North America. Cornell University Press, Ithaca, NY. xv + 941 pp. + 24 color plates.

Adler, P. H. & McCreadie, J. W. 1997. The hidden ecology of black flies: sibling species and ecological scale. American Entomologist 43: 153-161.

Colbo, M. H. & Wotton, R.S. 1981. Preimaginal blackfly bionomics. *In*: Blackflies: The Future For Biological Methods In Integrated Control. Edited by M. Laird. Academic Press, London, pp. 209-226.

Corkum, L.D., 1989. Patterns of benthic invertebrate assemblages in rivers of northwestern North America. – Freshwat. Biol. 21: 191-205.

Craig, D. A. &, Galloway, M.M. 1987. Hydrodynamics of larval black flies. *In*: Blackflies: Ecology, Population Management, and Annotated World List. Edited by K. C. Kim and R. W. Merritt. Pennsylvania State University, University Park, pp. 171-185.

Feminella, J.W., 2000. Correspondence between macroinvertebrate assemblages and four ecoregions of the southeastern USA. – J. N. Am. Benthol. Soc. 19: 442-461.

Gaston K.J. & Blackburn, T.M. 2000. Pattern and Process in Macroecology.Blackwell Science, Oxford.

Hamada, N. & McCreadie, J.W. 1999. Environmental factors associated with the distribution of *Simulium perflavum* (Diptera: Simuliidae) among streams in Brazilian Amazonia. Hydrobiologia 397: 71-78.

Hamada, N., McCreadie, J.W. & Adler, P.H. 2002. Species richness and spatial distributions of black flies (Diptera: Simuliidae) among streams of Central Amazonia, Brazil. Freshwater Biology 47: 31-40.

Mcardle, B.H., 1996. Levels of evidence in studies of competition, predation and disease. N.Z. J. Ecol. 20: 7-15.

McCreadie, J. W. & Adler, P. H. 1998. Scale, time, space, and predictability: species distributions of preimaginal black flies (Diptera: Simuliidae). Oecologia 114: 79-92.

McCreadie J.W. & Adler, P.H. 2006**.** Ecoregions as predictors of lotic assemblages of blackflies (Diptera: Simuliidae). Ecography (in press).

McCreadie, J.W., Hamada, N. & Eugenia-Grillet, M. 2004. Spatial-temporal distribution of preimaginal blackflies in Neotropical streams. Hydrobiologia 513: 183-196.

McCreadie, J.W., Adler, P.H., & Hamada, N. 2005. Patterns of species richness for blackflies (Diptera: Simuliidae) in the Nearctic and Neotropical regions. Ecological Entomology, 30: 201-209.

Richardson, J.S. & Mackay, R.J. 1991**.** Lake outlets and the distribution of filter feeders: an assessment of hypotheses. Oikos, 62: 370-380.

Schneider, D.C., 1994. Quantitative ecology: spatial and temporal scaling. Academic Press, San Diego.

Wotton, R.S., 1979. The influence of a lake on the distribution of blackfly species (Diptera: Simuliidae) along a river. Oikos, 32: 368-372.