Apr 19th, 4:15 PM

Strip Adaptive Cluster Sampling with Application to Cave Crickets

Kurt Lewis Helf
I&M Division/NPS, kurt_helf@nps.gov

Tom Philippi
I&M Division/NPS

Bill Moore
I&M Division/NPS, bill_moore@nps.gov

Lillian Scoggins
I&M Division/NPS

Strip Adaptive Cluster Sampling with Application to Cave Crickets

Kurt Lewis Helf[1], Tom Philippi[1], Bill Moore[1], and Lillian Scoggins[1]

[1] I&M Division/NPS

Introduction

Most cave ecosystems depend on the transport of organic matter from the surface by both passive (e.g., water) and active (e.g., cave crickets) agents (Culver and Pipan 2009, Schneider 2009). Subsidized ecosystems are vulnerable to perturbations that affect the production, transfer, and use of those subsidies (Riley and Jefferies 2004). Perturbations on a regional and local scale can affect productivity on the surface and the ability of surface-feeding cave organisms to access it and thus alter the amount of the subsidy being transferred to the subsurface. Important insights into the individual and collective effects of local changes on actively subsidized cave terrestrial ecosystems in the southeast can be gained through assessing the modulation of cave cricket entrance populations.

Cave crickets (*Euhadenoecus* and *Hadenoecus sp.*) are commonly found roosting in high densities just inside cave entrances throughout the southeastern United States. They are omnivores that feed on the surface and transfer nutrients-in the form of guano, eggs, and bodies-into the subsurface habitat.

In the Mammoth Cave region, cave crickets (*Hadenoecus subterraneus*) are a keystone species in that their entrance populations subsidize up to three separate cave invertebrate communities through the active, regular transfer of organic matter from the surface to the subsurface (Poulson and Lavoie 2000, Lavoie et al. 2007). The communities they subsidize can include rare, sometimes endemic, obligate cave-dwelling invertebrates (Culver et al. 2000).

Perturbations affecting the availability of surface resources to cave crickets, such as contingent climatic conditions, can alter the amount of nutrient subsidies they transfer to their dependent subsurface communities. Poulson et al. (1995) showed conditions favorable to cave crickets foraging on the surface (i.e., warm winters, cool summers, and above average precipitation) were correlated with the highest abundance and diversity of the cave invertebrate community dependent on cave cricket guano and declined in years with cold winters, hot summers, and below average precipitation.

Helf's (2003) study provided rigorous support for Poulson et al.'s (1995) data in that it showed a significant inverse relationship between cave crickets' use of artificial bait patches and precipitation among growing seasons, strongly suggesting cave crickets fed more on artificial bait patches due to decreased primary productivity on the surface.

Extremes in maximum temperature and precipitation events across the Southeast, predicted by mid-century (Fisichelli 2013, Kunkel et al. 2013), could lead to reduced primary productivity on the surface. While precipitation and primary productivity are often positively correlated the predicted concomitant temperature increases may increase evaporation and so lead to a net loss in moisture available to surface communities

(Young et al. 2011). Drier surface conditions may directly reduce the amount of organic material available to cave crickets or indirectly reduce its availability by creating suboptimal foraging conditions that preclude cave cricket foraging bouts (Studier et al. 1987, Poulson et al. 1995, Helf 2003, Lavoie et al. 2007).

On the other hand, minimum temperatures below freezing are also predicted to decrease by 20-25 days/year (Fisichelli 2013) which suggests increased foraging opportunities for cave crickets during winter months. Increases in winter foraging opportunities may compensate for decreased foraging opportunities in summer.

Management actions, such as altered cave entrance configuration, can also affect the flow of allochthonous organic matter into caves due to their effects on cave cricket foraging behavior and population structure (Fry 1996, Poulson et al. 2000, Helf 2003). Indeed, from 1993-1996, Mammoth Cave National Park (MACA) facilities and resources management personnel retrofitted cave entrance doors with airlocks to mitigate the negative effects of cold, dry winter air on the growth and formation of speleothems and biological communities (Fry 1996).

To assess the potential effects of this program MACA funded visual censuses of cave cricket populations at nine cave entrances, six with varying degrees of anthropogenic modification and three without, from 1994-1998. Among all cave entrances overall cave cricket abundance declined significantly from 1994-1997 (Poulson et al. 2000).

Monitoring Objectives

The Cumberland Piedmont Network's primary monitoring goal is to assess status and trends of MACA's cave cricket entrance populations and their habitat use; we have three monitoring objectives:

- **Monitoring Objective 1:** To determine the status and trend of cave cricket entrance population size, life stage, and sex ratio among 15 developed and undeveloped cave entrances at Mammoth Cave National Park during biannual visits.
- **Monitoring Objective 2:** To determine effects of management decisions (e.g., alteration of cave entrances) at Mammoth Cave National Park on cave cricket populations within selected developed caves. Specific monitoring foci will include assessment of the impact of cave-entrance modification on cave cricket population size and structure and localized impacts of infrastructure installation/ improvement on cave cricket habitat use.
- **Monitoring Objective 3:** To determine if a correlation exists between cave temperature, relative humidity and air flow trends, surface temperature, relative humidity and precipitation trends and: 1) trends in cave cricket entrance population size, life stage, and sex ratio, and 2) trends in spatial distribution within 15 developed and undeveloped cave entrances in Mammoth Cave National Park using biannual and continuous automated sampling.

Field Methods

For this protocol the overall statistical population of interest is the set of cave crickets using a set of cave entrances in MACA. Inferences will be made comparing cave cricket entrance populations between developed (i.e., entrances with bat gate or door(s), significant modification to its entrance/passage or significant infrastructure, such as a lighting system, or regular tours) and undeveloped entrances (i.e., entrance with or without bat gate, light or no modification to its entrance/passage or no infrastructure or no tours).

Because neither a complete census of cave entrances nor a complete census of cave cricket entrance populations is possible, this monitoring protocol requires two separate sampling frames: the selection of which cave entrances to monitor and defining how to sample within cave entrances. Such multi-stage sampling designs (Thompson 2002) are common for large-scale environmental surveys. At the broad level of cave entrances our sample frame consists of 15 cave entrances within MACA's boundary stratified by whether they are developed or undeveloped.

Because neither a complete census of cave entrances nor a complete census of cave cricket entrance populations is possible, our target population requires a multistage, adaptive sampling design (Thompson 2002, Salehi and Seber 2013) for defining how to sample within cave entrances. The within-entrance component of cave cricket sampling is designed to provide estimates of the total number of crickets in that entrance, separate estimates of numbers of individuals by life stage and sex, and estimates of counts as a function of distance from the opening to the surface.

For sampling rare, clumped distributions adaptive cluster sampling and related methods have the potential to be much more efficient than simple random sampling in that their variance declines with sample size relative to simple random sampling (Thompson 2002). In addition to the estimates of total population size, adaptive cluster sampling automatically partitions the population size into components of cluster size and numbers of clusters, which can be informative for interpreting temporal changes in population size within each entrance.

Thus, this protocol uses a combination of a linear transect, (i.e., baseline) running down the length of the passageway from the entrance toward the depth of the cave, and strip adaptive cluster sampling (Thompson 2002) with strip locations defined by positions along that baseline. Generalized Random Tessellation Stratified (GRTS) sampling is used to select strip locations along the baseline to provide spatial balance to the survey.

During a sampling event one crew, comprised of two individuals, surveys a randomized selection of two cave entrances per day. A fiberglass measuring tape, placed in the same location each sampling event, serves as the baseline on which the randomized strips are positioned. The strips are defined by two red laser lines separated by 10cm, perpendicular to the baseline, and projected on the walls and ceiling of the passageway (Figure 1).

When a cricket is detected within a strip we use a plotless adaptive cluster sampling design (Mosquin and Thompson 1998). That is, for each cricket in a strip, any other crickets within 10cm are added to that cluster, and any crickets within 10cm of those crickets, recursively, until no additional crickets are within 10cm of any cricket in the cluster (Figure 2).

Digital images of each cave cricket clusters are captured. From these images counts of cave crickets, both inside and outside the strip, are obtained during subsequent image analysis. Data on cave cricket entrance populations are derived from a careful

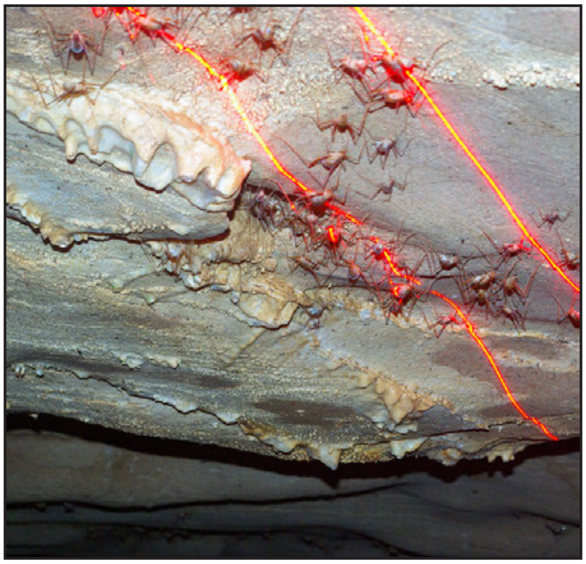

Figure 1: Cluster of cave crickets captured by 10cm wide laser strip projected on the ceiling at Frozen Niagara entrance.

analysis of the digital images as shapefiles in ArcMap. Ancillary data on clusters include mapping the location of each sampled roosting cluster, the width (i.e., extent) of sampled roosting clusters, and roost site descriptive characteristics (e.g., located on wall or ceiling).

Sampling events for cave cricket monitoring are conducted biannually. In a sampling year two sets of sampling events are conducted at all 15 sampling sites. Sampling events occur within a two-week period each "shoulder season" (i.e., May-June and October-November), at each of the 15 selected cave entrances at MACA.

Previous monitoring efforts show these months are the best times of year to maximize sample size and reduce day to day variability among entrance populations because equable weather creates optimal foraging conditions on the surface and similar proportions of cave cricket entrance populations forage on any given evening (Helf 2003, Lavoie et al. 2007). Due to drought conditions during the mid-summer through late fall months cave cricket abundance on any given day is highly variable and so the potential for substantial sampling noise is greatly increased.

Prior to each sample event or group of sampling events (a grouping of cave entrances to be visited during a sampling session), the project leader conducts a GRTS draw to randomize the order in which caves are visited and the order in which locations on the baseline are surveyed during in-cave sampling.

The R code which generates these draws harvests a list of entrances to be visited, within-cave sample sizes, and the sequential order of caves visited from the previous sampling event. This code then formats and populates field data sheets in Microsoft Word™.

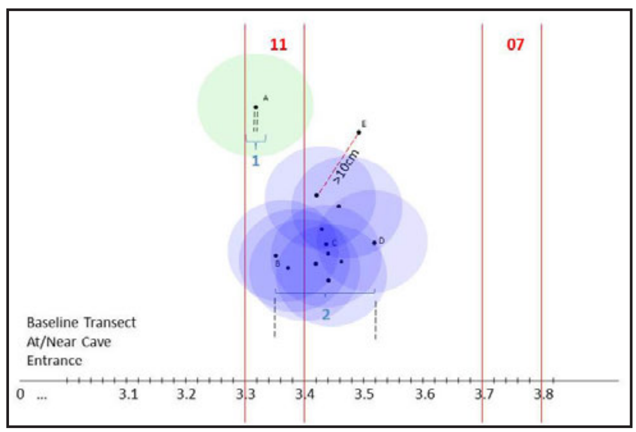

Figure 2: Schematic diagram of the plotless strip adaptive cluster sampling method used to monitor cave cricket entrance populations at Mammoth Cave National Park. Note that because the probability a cluster of crickets will be detected is dependent only on the extent of that cluster along the baseline, the grid and virtual quadrats need not exist. Any cricket intersected by a strip triggers a cluster; any crickets within 10cm radius of a cricket in a cluster are added to that cluster, and in turn have their 10cm radii searched.

In addition to generating the primary field data sheets, R code is also be used to create temporary tables in the protocol database. The values in these tables can then be utilized during the data entry process reducing manual data entry. R code is also used to pull and summarize the counts from the various shapefiles generated and append values in the temp_* tables in Access.

Data Management and Analysis

In short, the majority of data entry is not accomplished via the traditional method whereby an individual sits down at their computer with a completed field data sheet and enters each value into a similarly designed form on the computer. Instead much of the data are populated into temporary tables in the database via R code.

Thus the data entry process includes: ensuring data are accurately parsed to the correct location/event combinations in the 'permanent' tables in the database; data records are complete; and finally, entry of remaining data elements from the field data sheets (e.g., notes fields, cricket cluster locational information) is completed. A series of Quality Assurance/Quality Control checks are in place to assist in this process.

Data from the MACA cave cricket monitoring project are analyzed/summarized in multiple ways:

- Annual status summary of cave cricket monitoring highlights,
- Analysis of trends in key measures over time; typically summarized every five years,
- Evaluations of relationships between key ecosystem drivers/attributes/ stressors and key measures including cave and surface meteorology and infrastructure installation/ maintenance.

Data from the MACA cave cricket monitoring protocol support both non-adaptive estimates based on the counts inside strips and strip adaptive cluster sampling (SACS) estimates based on the counts by clusters. SACS should be substantially more efficient (i.e., lower uncertainty about estimates for a given sampling effort) than non-adaptive estimates based on just the crickets inside strips (Thompson 2002).

However, because the rules for adaptively sampling clusters are based on all crickets, strip adaptive cluster estimates of the total counts for some sub populations (e.g., juveniles) might be less efficient than non-adaptive estimates. Therefore, as is common practice in these applications, we will compute both non-adaptive estimates based on strips and SACS estimates based on clusters, for the total population of crickets, and for the subpopulations based on sex and life stage (Ver Hoef and Boveng 2007).

Given these estimates of the total numbers of crickets at each cave entrance and sampling event, temporal trends will be tested as both generalized linear mixed models (GLMM using function glmer in the lme4 R package) and generalized estimating equations (GEE using function geeglm in the geepack R package). Both of these approaches are appropriate for count data that are likely to be overdispersed relative to the Poisson error distribution expected for counts of independently occurring events. For technical reasons, the glmer approach fits overdispersed Poisson as a two-parameter negative binomial distribution. The geeglm approach adds an overdispersion parameter and treats the error distribution as quasipoisson.

These models also support tests for differences in trend among cave entrances or among groups of cave entrances (e.g., between developed and undeveloped entrances). However, because the monitored entrances are not a probability sample of any defined population of entrances, the tests support inferences about only these particular entrances, and not to unsampled developed or undeveloped entrances.

The status of cave cricket entrance populations over time is one of the objectives of this monitoring protocol and is effectively presented by a form of control chart. The estimated population size for the most recent sampling event at each entrance is plotted over a boxplot of the estimates from previous sampling events (Figure 3).

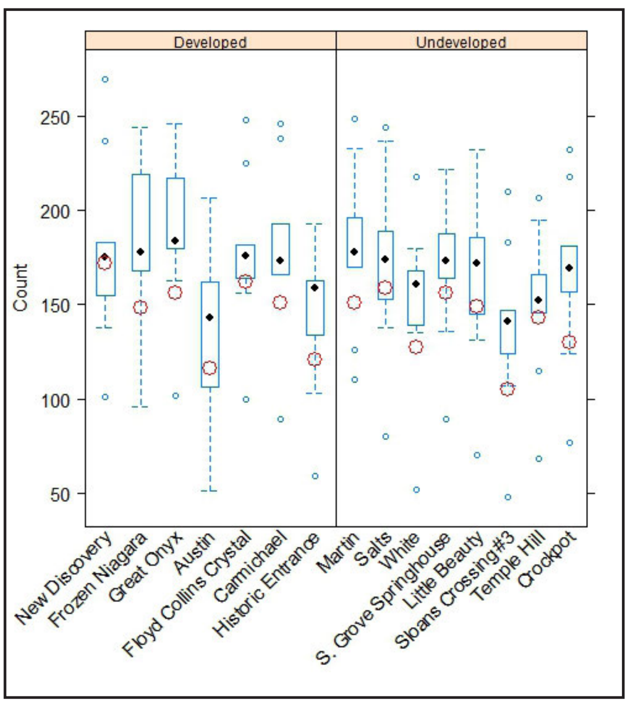

Figure 3: Mockup of control chart depicting ten years of estimates of monitored cave cricket entrance populations. The most recent sampling event (large, open circles) are plotted over a boxplot of estimates from previous sampling events. Note: dots indicate the median of the data and small, open circles are outliers.

This produces a visual representation of which, if any, of the monitored cave entrances have recent population estimates high or low relative to that cave entrance's historic range of variability. If some current values are high and some are low, there is cave entrance-specific fluctuation. If most cave entrances deviate in the same direction that suggests a region-wide driver such as surface weather or food sources.

Estimates of total cave cricket entrance populations, sex, and life stage, are only one aspect of cave cricket status in these entrances. Other aspects may also be informative of impacts of cave entrance management, climate, or other stressors. For instance, the distribution of roosting crickets as functions of distances from the cave entrance to aboveground might shift due to changes in air circulation or meteorological conditions in the first few tens of meters of the passageway.

This sampling and data collection scheme supports estimates of several such secondary aspects. Temporal changes in total cave cricket entrance population will be estimated and also partitioned into several components of numbers of clusters and the distribution of the numbers of crickets per cluster (Figure 4). The distribution of crickets as a function of distance from the surface can be characterized as cumulative distribution functions estimated for individual cave entrances and each cave entrance can support tests for shifts in those distributions over time.

To reduce the time and effort normally required to write annual status and trend analysis reports R code, used to access standard databases to produce informative tables and figures, will be added during initial report writing in MS Word™. Thus, when new data are entered into the database the R code run on those data will produce new report components.

For consistency between/among report intervals all of the formatting, boilerplate background text, and forms of tables and figures will remain the same year after year. This scripting of the workflow provides both documentation and automation, and makes the work reproducible from one year to the next.

Reports generated by this monitoring project will consist of three major types. Trip Reports will be written to briefly summarize sampling trips for park staff. Brief follow-up trip reports will be completed within two weeks after each sampling trip. Annual Status Reports and Trend Analysis Reports will provide park management and other interested parties technical and interpretive information about the status and trends being detected in the monitored resource.

The annual status report may include descriptive statistics, graphic analysis, and correlative statistics on cave cricket entrance populations and will be produced in late winter after the preceding year's monitoring events and subsequent data analyses are completed. This type of report will target MACA's superintendent and resource managers and will provide them with a view of the current status and short-term shifts in any parameter(s) of the resource. Annual status reports will be submitted to the Natural Resources Data Series for publication.

The trend analysis report will typically be generated every fifth year, beginning five years after the formal implementation of the monitoring protocol. The trend analysis report will also address patterns in cave cricket population structure and dynamics among developed and undeveloped caves, using similar components as the annual status report, but will do so with cumulative data on a scale spanning multiple years.

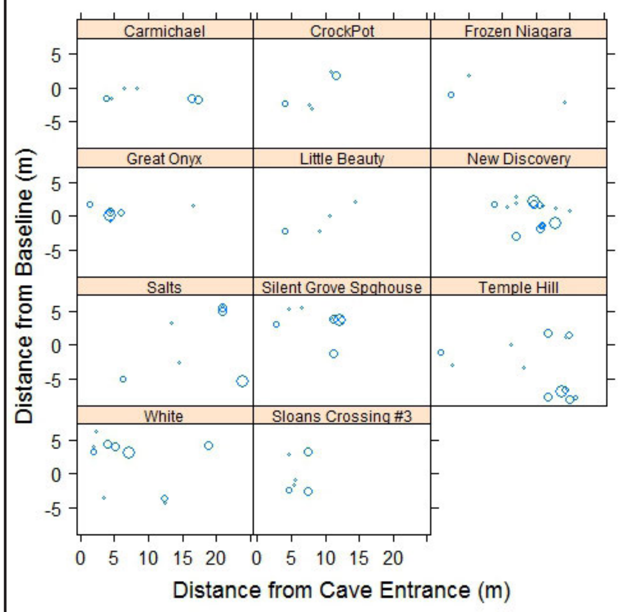

Figure 4: Proportionally sized bubble plot of cave cricket clusters from pilot data. The y-axis is the cluster's distance from the baseline and side of the passage on which they were located (positive = left and negative = right). The x-axis is the cluster's distance from the cave entrance. Proportional bubble plots are an informative way to display data on the location, number and size of cave cricket clusters.

Literature Cited

Culver, D. C., L. L. Master, M. C. Christman, and H. H. Hobbs. 2000. Obligate cave fauna of the 48 contiguous United States. Conservation Biology 14:386-401.

Culver, D. C. and T. Pipan. 2009. The Biology of Caves and Other Subterranean Habitats. Oxford University Press, New York, 254pp.

Fisichelli, N. 2013. Climate Change Trends, Mammoth Cave National Park, Kentucky. Unpublished Report, 6pp.

Fry, J. F. 1996. Eighteen cave gates and airlocks: conclusion of a three-year project to restore cave entrance dynamics at Mammoth Cave National Park in Mammoth Cave National Park's Fifth Science Conference, Mammoth Cave National Park, pp. 69-83.

Helf, K. L. 2003. Foraging Ecology of the Cave Cricket Hadenoecus subterraneus: Effects of Climate, Ontogeny, and Predation. Ph.D. University of Illinois at Chicago, Chicago, IL, 170pp.

Kunkel, K. E., L. E. Stevens, S. E. Stevens, L. Sun, E. Janssen, D. Wuebbles, C. E. Konrad, II, C. M. Fuhrman, B. D. Keim, M. C. Kruk, A. Billet, H. Needham, M. Schafer, and J. G. Dobson. 2013. 2013: Regional climatic trends and scenarios for the U.S. National Climate Assessment. Part 2. Climate of the Southeast U.S., National Oceanic and Atmospheric Administration, 94pp.

Lavoie, K. H., K. L. Helf, and T. L. Poulson. 2007. The Biology and Ecology of North American Cave Crickets. Journal of Cave and Karst Studies 69:114-134.

Mosquin, P. L. and S. K. Thompson. 1998. Frame free adaptive designs in Online proceedings of the Survey Research Methods Section. American Statistical Association, pp. 431-436.

Poulson, T. L. and K. H. Lavoie. 2000. The trophic basis of subsurface ecosystems in H. Wilkens, D. C. Culver, and W. F. Humphreys, editors. Subterranean Ecosystems. Elsevier, Amsterdam, pp. 231-249.

Poulson, T. L., K. H. Lavoie, and K. L. Helf. 1995. Long-term effects of weather on the cricket (*Hadenoecus subterraneus,* Orthoptera, Rhaphidophoridae) guano community in Mammoth Cave National Park. The American Midland Naturalist 134:226-236.

Poulson, T. L., K. H. Lavoie, and K. L. Helf. 2000. NRPP Entrance Monitoring Final Report. National Park Service, Mammoth Cave, KY.

Riley, R. H. and R. L. Jefferies. 2004. Subsidy dynamics and global change in G. A. Polis, M. E. Power, and G. R. Huxel, editors. Food webs at the landscape level. The University of Chicago Press, Chicago, IL, pp. 410-433.

Salehi, M. M. and G. A. F. Seber. 2013. Adaptive sampling designs: inference for sparse and clustering populations. Springer, Heidelberg, Germany, 70pp.

Schneider, K. 2009. How the availability of nutrients and energy influence the biodiversity of cave ecosystems. University of Maryland, College Park, 174pp.

Studier, E. H., W. D. I. Wares, K. H. Lavoie, and J. A.-M. Linn. 1987. Water budgets of cave crickets, *Hadenoecus subterraneus* and camel crickets, *Ceuthophilus stygius.* Comparative Biochemistry and Physiology 86A:295-300.

Thompson, S. K. 2002. Sampling. Second edition. John Wiley & Sons, Inc., New York, NY, 367pp.

Ver Hoef, J. M. and P. L. Boveng. 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88:2766-2772.

Young, B., E. Byers, K. Gravuer, K. Hall, G. Hammerson, and A. Redder. 2011. Guidelines for using the NatureServe climate change vulnerability index. NatureServe, Arlington, VA, 63pp.