THE RELATIONSHIP BETWEEN CUE ABUNDANCE AND CUE AVAILABILITY, AND ITS IMPACT ON DETECTABILITY DURING POINT COUNT SURVEYS:
A MONTE CARLO SIMULATION STUDY

D. Archibald McCallum

Report to
Patuxent Wildlife & Research Center
Breeding Bird Survey
12100 Beech Forest Road
Laurel, Maryland 20708-4038

Under Purchase Order No. 02ERSA0327

ABSTRACT

Monitoring territorial bird populations usually entails sampling rather than enumeration. Sample results are often negatively biased by the failure to account for birds present but not detected. One source of such bias, undefined until recently, is “availability”, the probability that a bird that is present in the count area produces a cue that is potentially detectable. Availability is a precondition to detection, which may be constrained by ambient noise, sensory acuity of the observer, etc. The probability that a cue, once given, is detected by an observer, is “detectability.” Detection probability is the product of availability and detectability.

For aural counts, which are most often used with territorial land birds, such cues are typically songs or other sounds. I explored the underlying causes of availability, and methods for estimating it, with a Monte Carlo simulation model driven by the probability of singing and an independent probability of continuing a behavior (sing/not sing), once begun. According to the results of this simulation, song production is a linear function of probability of singing, independent of the probability of continuing and largely independent of the scale at which it is measured. It is relatively simple to estimate the probability of singing from song production rates during short point counts. Availability increases, in the model, as a power function of singing rate, with high probability of continuing reducing this power. These power functions should be estimable if estimates of continuation probability can be obtained, but the most obvious short-term estimator proved uncorrelated with continuation probability. The effort to find another estimator was abandoned because availability proved to be estimable directly from brief samples with linear functions.

I investigated the feasibility of estimating availability directly from data that can be collected easily during brief point counts, such as stops on the Breeding Bird Survey (BBS). One-minute estimates of singing rate (songs per minute) were poor predictors of availability, regardless of sample size, in line with the power relationship described above. On the other hand, the probability that a bird singing in the second half of the count period was also singing in the first half, when averaged over at least 25 samples, explained 90% of the variation in the true availability of the sampled bird. This probability is used in a recently published method for estimating detection probability (Farnsworth et al. 2002) that is based on capture-recapture logic. The results of this simulation study suggest that availability can be estimated easily on the BBS by dividing the count period in half and synonymizing the singers in the second half with those in the first.

INTRODUCTION

Assessing the conservation status of wildlife populations requires monitoring. Most of the many monitoring techniques in use involve sampling rather than enumeration. Both the precision and bias of such samples must be estimated (Thompson 2002). An important potential source of bias is the proportion of animals present that are not detected during sample periods. Nondetection can result from a weak signal that may be overlooked, but also from the absence of a signal due to the inactivity of an animal. These two sources are conveniently termed "detectability" and "availability, " respectively (McCallum in press). Having only recently been identified as a parameter that should be estimated independently (Farnsworth et al. 2002), availability remains under-investigated conceptually and mathematically.

AVAILABILITY AND DETECTABILITY

McCallum (in press) defined a set of heuristic parameters designed to capture important aspects of the natural history of singing in territorial land birds, as well as the phenomenology of observation and detection of these songs by human observers. Briefly, in this approach,

$$P = p_s \, p_{d|s} \tag{1}$$

where P is the probability of detecting a bird that is present during a count period of m minutes, p_s (availability) is the probability that an average bird sings (or produces some other detectable cue) at least once during that period, and $p_{d|s}$ (detectability) is the probability that the bird is detected, given that it sings.

For detectability,

$$p_{d|s} = 1\text{-}(1\text{-}\ p_{1d})^{s} \tag{2}$$

where p_{1d} is the probability of detecting an average cue; and s is the number of cues, i.e., songs or other detectable acts, expected during a count period that contains at least one cue. Parameter p_{1d} is a measure of conspicuousness, i.e., it captures reductions in detectability due to amplitude of the cues, auditory acuity of the observer, attentiveness of the observer, and masking of the cues by other cues and extraneous stimuli (McCallum in press). Parameter s is the number of cues produced during a count period, independent of their intensity. High cue abundance mitigates all four causes of non-detection, by giving the observer multiple opportunities to make the single detection that is needed to count an individual. All these parameters are specific to the duration of the count period m (McCallum in press), so expected cue abundance can be increased by increasing m (count period duration) or s/m (singing rate, i.e., songs per minute). Detectability ($p_{d|s}$) is therefore the probability of detecting a bird at least once during a count period, given that it gives at least one cue during the count period.

Equation (2) quantifies the intuitive relationship between singing rate and the likelihood of detecting a singing individual. Even inconspicuous cues (low $p_{1d)}$ can result in detection when they are numerous (high s). For example, $p_{1d} = 0.2$, as one might find during an intense dawn chorus, translates to $p_{d|s} = 0.996$, with a realistic s of 25, or 5 songs per min in a 5-min count period. It follows that availability may be of more concern than detectability in the estimation of P.

The expected value of availability is 1 if the expected value of s is ≥ 1, so it would appear that availability is of no concern if a mere one song is expected per count period. Beyond that, further singing does not improve availability (which cannot exceed 1), although it does increase detectability through exponentiation of the probability ($1-p_{1d}$) of not detecting an average cue. But this is only the expected value of availability (p_s), under the (usually tacit) assumption of statistical independence of all songs. For example, if 60 songs are performed in an hour, one song is expected to occur each minute, and the expected value of availability for 1-min counts is 1, obviating the need to estimate it. This assumes no stochasticity, through which, say, 0, 2, or 3 songs might easily occur in some 1-min periods, despite the average rate of 1 song per minute. Stochasticity can be accounted for by an estimate of the variance in singing rate. It might prove necessary to have a singing rate of, say, three songs per minute to ensure that at least 1 song would occur every minute, with an acceptable probability of say, 0.95.

But, the natural history of avian singing indicates a more severe challenge to the equation of availability with singing rate. Birds tend to sing in bouts, during which many (e.g., several dozen) songs are produced each minute, followed by periods of silence, which may last longer than most point count periods (which seldom exceed 10 min in duration). The average value of s over a large span of count periods could be >>1, and some of them still could be songless. It is for this reason that availability must be assessed, independently of mean singing rates.

METHODS FOR ESTIMATING P

The recent flurry of publications on detection probability (Nichols et al. 2000, Buckland et al. 2001, Bart and Earnst 2002, Rosenstock et al. 2002, Farnsworth et al. 2002, and especially Thompson 2002) suggests that changes in sampling techniques may be imminent. These authors advance four largely independent methods for estimating detection probabilities. The double-observer method (Nichols et al. 2000) estimates auditory acuity and attentiveness of the observer, which are components of detectability (McCallum in press) by comparing the results from two observers. The double-sampling method (Bart and Earnst 2002) estimates availability in intensive study plots and uses it to correct rapid estimates made in these and other plots. The distance method (Buckland et al. 2001) estimates detectability by means of the fall-off in detections with distance from the observer. The removal method (Farnsworth et al. 2002) estimates availability by comparing singing activity in two subsections of a count period. This last method is of particular interest because it is the first to explicitly decouple availability and detectability, and because it claims that availability can be estimated concurrently with rapid point counts, rather than requiring intensive nest-finding and territory-mapping efforts (Bart and Earnst 2002).

In all four methods, $N = C / P$, where N is the estimated number of animals in a sample area, C is the sample count of animals in that area, and P is the correction factor or detection probability. The removal method of Farnsworth et al. (2002) uses capture-recapture logic to estimate availability. This is a promising approach. (The claim (Farnsworth et al. 2002:415-416) that the removal model estimates the product of availability and detectability is erroneous, but this does not diminish the usefulness of the model for estimating availability). The fundamental relationship used for capture-recapture estimation is $C_1/N = R/C_2$, where N is the true number of animals present and C_1 is the number captured, or counted, in the initial "capture" session. Clearly, R / C_2 is equivalent to P, where C_2 is the count of individuals in the second, "recapture" session, and R is the number of individuals in C_2 that were also in the set C_1. Estimating P as R / C_2 allows estimation of N as C_1 / P. (The *removal* model estimates N from C_1 and the number of captures in the second session that are *not* recaptures. The latter is C_2-R, so the removal and recapture approaches are mathematically interchangeable. I use the recapture approach because the focus here is on availability rather than N.)

A simple numerical example will confirm how this approach works. Suppose that 10-min point-count survey periods are divided into two consecutive 5-min blocks. Suppose that the probability of a bird singing at least once in a 5-min period is 0.5, and that singing probabilities in consecutive blocks of time are independent. For simplicity, assume further that only one bird can be present in a sample area. There are four possible outcomes for 10-min survey periods, each occurring with an equal probability of 0.25: 00, 01, 10, 11, where, for example, 01 indicates silence in the first 5 min, followed by at least one song in the second period. Under these conditions, 50% of the samples have a detection in the second session (C_2), and 50% of these are "recaptures" (R) of an individual from the first session. The method correctly recovers $P = 0.5$ from such results.

Capture-recapture methodology was developed primarily for overnight trapping sessions at least 24 hours apart. Under such circumstances, the assumption of the model that the captures on the two nights are statistically independent is at least somewhat plausible. But, given the widespread practice of singing in bouts by birds, the independence of events (e.g., songs) < 10 min apart is in question. In nature, cues are frequently clumped in time. For example, Scott et al. (unpublished) showed that once a bout of singing by a Least Bell's Vireo (*Vireo bellii pusillus*) began, it was highly likely to continue, as was a silent period (see also Hailman et al. 1985). The effect of clumping of this sort on realized detection rates and availability has not been explored.

It may be that the removal model can deliver an unbiased estimate of availability even when singing in the two sessions of a count are not independent, if the estimate is based on data pooled from a number of independent counts (e.g., stops on a BBS route). Although the mark-recapture approach is typically used to estimate N for each site, estimating an aggregate measure of abundance for a set of sites assumed to be random samples from the same population (e.g., Farnsworth et al. 2002, MacKenzie et al. 2002, Royle and Nichols 2003) is preferable for point-count data.

OBJECTIVES OF THIS STUDY

In this study I use Monte Carlo simulation to investigate the general relationship between singing rate and availability, paying particular attention to the estimation of availability using mark-recapture logic. The following questions are addressed explicitly.

1. Can availability be estimated directly from singing rates estimated during brief point counts?
   a. Can availability at a single site be estimated from the local estimate of singing rate (thereby permitting estimation of *N* for single sites)?
   b. If not, can average availability for a set of sites be estimated from an estimate of average singing rate in those sites?
2. Can availability be estimated with mark-recapture analysis of patterns of singing and silence during brief point counts?
   a. Can availability at a single site be estimated from the local estimate of R/C_2 (thereby permitting estimation of *N* for single sites)?
   b. If not, can average availability for a set of sites be estimated from pooled R/C_2 data?

Simulated data are actually more dependable than field data for preliminary explorations of this sort, because the causes of the variation in the data are explicit. In the present case, the probability of singing an individual song is completely responsible for singing rate, and the probability of continuing a behavior is, independently, responsible for bout structure. All other parameters, such as those defined above, are emergent properties of the interaction between these two.

METHODS

THE MODEL

Behavior of a single individual is entirely determined by the following simple parameters:

1. **Time increment**. The shortest period comprising a single song and one inter-song interval of silence is defined as the unit of time in the simulation. Song duration and the inter-song interval were both set at 1, resulting in a minimal time-increment of 2 sec. The simulation incremented, 2 sec at a time, through 1 hr of possible singing. An entire hour was simulated because, for many territorial land birds, several periods of extended singing and silence are accommodated in a period of this duration. Standard count periods of 3-10 min often will not accommodate several cycles of singing and silence, and hence are subject to selection bias (Thompson 2002).

2. **Probability of singing**. At each time step, the bird sang with probability PROBSING. This parameter assumed the values 0.01, 0.02, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, and 0.95 in the simulation. A state variable, SING, was set at 1 with the binomial probability PROBSING, otherwise it was set at 0. The values of PROBSING used in the simulation are equivalent to the following *s/m* values: 0.3, 0.6, 1.5, 3, 4.5, 6, 7.5, 10.5, 13.5, 16.5, 19.5, 22.5, 25.5, 28.5. That is, only the two lowest values of PROBSING yield expected values of availability that are <1.

3. **Probability of continuing**. The tendency for a bird to continue what it was doing was modeled with parameter CONTPROB. This parameter introduces different levels of serial dependence, from none to almost total (90%). The probability of continuing / discontinuing the current behavior was CONTPROB / 1 – CONTPROB. If the simulation returned a "discontinue" order, the next behavior was determined by the value of PROBSING. PROBSING was called at every time step, but the output was overridden when CONTPROB resulted in a continue order. CONTPROB was set at 0 to 0.9 in increments of 0.1. CONTPROB = 0 is the special case in which every song event is independent of every other. This condition explores stochasticity in the absence of serial dependency. Other values of CONTPROB indicate the degree of serial dependency in the SING (0/1) decision.

ASSAYS

AVAILABILITY was calculated as active minutes / 60 for each rep. An active minute is one in which at least one song occurs. One song is considered sufficient indication of the presence of an individual in most point-count protocols. Thus, AVAILABILITY captures the proportion of minutes in which a bird that is present is actually available for counting by aural means.

Songs per minute (SPM) was calculated as total songs produced in one hour divided by 60. SPM and AVAILABILITY are emergent variables that are used to characterize the simulated singing behavior. These variables could be used to estimate *P* and thereby *N*. In real point counts, however, 60 minutes are not available for estimating SPM and AVAILABILITY in this way. SONGS is defined as the number of songs simulated for each minute. SONGS is used as an estimator of SPM. R_{11} is the proportion of minutes preceded by an active minute that are also active, equivalent to R / C_2 above. It is an estimate of the probability of continuing to sing, taken during a very short segment of time. The two questions of this study address the correlation between these parameters estimated in very brief samples, and the emergent variables they are used to estimate.

STATISTICAL ANALYSES

To determine if higher values of CONTPROB biased the number of songs in a simulation run, I conducted a 2-way ANOVA, with interactions, of the effects of CONTPROB and PROBSING on total songs in a 1-hr simulated soundscape. Statistically significant results for this test could require an adjustment for overall cue abundance to make all combinations of PROBSING and CONTPROB equivalent and thereby comparable.

To address the the predictability of availability from singing rate, I averaged SONGS over 1, 2, 5, 10, 25, 50, and 60 randomly drawn 1-min samples from each simulated hour. The random draw was conducted by ordering the 60 min of each rep in random order and then retaining the first 1, 2, 5, ... 60 observations, so the average of SONGS for 60-min samples was identical to SPM. Each minute of each rep can be thought of as a random sample from a population of sites all having the same probability of singing, continuation probability, and realized availability. I regressed realized AVAILABILITY (i.e., calculated over the entire 60-min simulation) against the estimates of singing rate to build models for estimating availability from singing rate. The coefficient of determination (r^2) was used as a measure of confidence in a model.

To address the predictability of AVAILABILITY with the removal method, I randomly drew 1, 2, 5, 10, 25, 50, and 60 2-min sequences from each simulated hour, as was done with singing rate. If singing occurred in the second minute, the value of R_{11} was set to1 if singing occurred in the first minute, 0 if it did not. This probability was averaged over the number of samples. (Sample size was reduced by 1 if the first minute of the original 60-min simulation appeared in the random sample.) I regressed realized AVAILABILITY (i.e., calculated over the entire 60-min simulation) against these estimates of availability to build models for estimating availability from these estimates. The coefficient of determination (r^2) was used as a measure of confidence in a model.

RESULTS

DESCRIPTIVE STATISTICS

The entire simulation produced 2,520,000 lines of output, or, one hour in 2-sec increments for 10 replicates each of 14 values of PROBSING, and 10 values of CONTPROB.

CONTPROB had no effect on total songs, independent of PROBSING. Although PROBSING was highly significant in the full factorial 2-way ANOVA, CONTPROB was not ($F = 1.46$, $df = 9, 1260$, $p = 0.1590$), nor was the interaction term ($F = 0.85$, $df = 117,1260$, $p = 0.8633$). There was no need, therefore, to adjust for total song production in analyzing the results of the simulations. Further confidence in this statement is added by the nearly perfect correlation among PROBSING, total songs, and SPM (all > 0.998). The actual song count in each minute of the simulation was also highly correlated with PROBSING ($r = 0.914$), the 9% decrement presumably indicating the effect of stochasticity on the 1-min samples.

AVAILABILITY increased with PROBSING (Fig.1), and decreased with CONTPROB (Fig. 2), as expected. Of interest to designers of monitoring schemes is the value of PROBSING required to make AVAILABILITY = 1. Given that the time increment used throughout the simulation was 2 sec, a maximum of 30 songs could occur in any minute, and therefore a probability of singing of only 0.033 would be required to make AVAILABILITY unity, if all songs are independent events. In fact, the observed mean AVAILABILITY was approximately 0.8 for PROBSING = 0.05, instead of the expected value of 1. This 20% decrease in AVAILABILITY, when CONTPROB = 0, is the contribution of stochasticity to the problem. Non-independence of singing events exacerbates this problem (Fig. 2), leading to an observed AVAILABILITY of around 20% with low probability of singing and very high continuation probability. This result from a PROBSING with an expected AVAILABILITY of 1 crystallizes the joint problems of stochasticity and non-independence for sampling in general.

QUESTIONS

Question 1. Short-term Singing Rate Predicts Availability

Figure 1 shows that AVAILABILITY is a power function of the probability of singing, with the power determined by CONTPROB. This relationship, coupled with the high correlation between SONGS and PROBSING ($r = 0.91$) predicts poor performance of a linear prediction model of AVAILABILITY on SONGS, which was the case. The coefficient of determination (r^2) was nearly constant (0.39 to 0.44) across the range of sample sizes (Fig. 3). Needless to say, estimates of *N* for single point counts, corrected with this estimator of availability, will not be great improvements over the raw counts.

Question 2. Short-term Presence/Absence of Song Predicts Availability

Estimates of availability obtained during two consecutive minutes of standard point counts (R_{11}) were highly correlated with realized AVAILABILITY calculated over the entire 60-min period of simulated singing. When all minutes of a simulated hour were averaged, the resulting regression equation was AVAILABILITY = $1.022 * R_{11} - 0.0263$ ($F = 37019$, $df = 1, 1392$, $p < 0.0001$, $r^2 = 0.9637$). This is equivalent to the 60-sample estimate. The coefficient of determination increased with sample size (Fig. 3), reaching 0.9 at a sample size of 25 (i.e., 90 % of the variation in AVAILABILITY is explained with 25 random 2-min samples). The low predictive power of the single samples ($r^2 = 0.4271$), however, does indicate that accurate estimates of *N* for single point counts are not attainable with this estimator of availability.

DISCUSSION

SOURCES OF BIAS

Conclusions drawn from the results of a simulation model have two major sources of bias. As with all data, simulated data can be mishandled, analyzed incorrectly, and interpreted illogically. A variety of institutional structures are in place to minimize this source. Additionally, simulation models must be both logically consistent and biologically realistic.

Errors in data handling and analysis can only be caught by careful proof-reading and conducting redundant analyses with alternative data sets. This model was built in stages, and preliminary results were analyzed at each stage. After completion of this study, the code was cleaned up, modules were concatenated, and the entire simulation was run from start to finish. New code was written for all analyses, and discrepancies among results of earlier and final runs were eliminated. Nevertheless, the way one author conceptualizes certain processes is likely to contain some hidden biases. If the results of this study are important, then it should be replicated by another modeler, from scratch, to assess the robustness of the results reported here.

As for biological realism, it would have been a simple matter to vary parameters for availability and detectability, independently, and see the range of bout structures that resulted. Such results would be self-fulfilling prophesies, though, and they would be open to the criticism that availability < 1 was forced into some replicates. The route actually taken, of simply varying the tendency to sing, and the serial dependence of that tendency on the immediately-preceding behavior, is not subject to that criticism. AVAILABILITY was an emergent property of the more naturalistic song production properties, as it presumably is in nature.

It is not known if a bout of birdsong is produced by the interplay between probability of singing and probability of continuing, as modeled here. This is a plausible model, but it also possible that singing birds pause only because of interruptions, rather than some Markov process. For present purposes, the realism of the model does not depend upon the underlying mechanisms. What matters is whether the resulting phenomenology, the bout structure, brackets the range of variation of naturally-occurring bouts.

There was, however, an intentional bias in this objective. The purpose of the study was to investigate the impact of clusters of songs interspersed with periods of silence on survey results. Accordingly, CONTPROB was varied from 0 to 0.9. Negative continuation probabilities were not modeled, because they would have resulted in hyper-dispersion of songs, which would increase availability and enhance detection. Availability ceases to be a serious problem in that region of parameter space, which was therefore not studied.

Few empirical data exist for validating the range of bout structure produced by this simulation model. A Least Bell's Vireo produced 4-14 runs per hour during the morning hours on 7 days distributed throughout the breeding season (Scott et al. unpublished). These values are intermediate in the range of values produced by the simulation model. The same vireo averaged 6 to 13 songs per minute, with a maximum of 24, also within the range simulated. These comparisons, though sparse, suggest that the model was successful at bracketing natural values.

THE MODEL

The expected value of total songs was constant across CONTPROB, because (1) the initial behavior of each series in a bout was generated independently with the same probability PROBSING, regardless of the value of CONTPROB, and (2) CONTPROB was the same for both states of the SING state variable. But it was not clear in advance that the duration of the simulated soundscapes (one hour) was adequate for these expected values to be realized in simulations with high values of continuation probability. As it turned out, one hour provided sufficient time for expected values to be realized, as shown by the results of the ANOVA.

The CONTPROB parameter was used to generate bout structure. The higher the value of this continuation probability, the more highly-structured became the hour of behavior. The resulting clumping of songs made for longer periods of singing and silence, and these periods of silence often exceeded 1 min in duration. Consequently, AVAILABILITY was well below 1 under many combinations of PROBSING and CONTPROB. The model therefore easily generated a wide range of values of AVAILABILITY under a realistic range of values for the generating parameters.

With such simple input, why is the relationship between singing rate and AVAILABILITY (Fig. 1) not linear? It is a power function, not because of the input, but because of the way AVAILABILITY is defined, which in turn depends upon the rules for counting an individual in point counts. In standard point-count practice, one or more songs in a minute (or other count period) have the same effect, namely making AVAILABILITY 1 for that period. The steep portion of each curve is where s is less than one per count period. The flat portion is the incremental increase in AVAILABILITY with additional songs, filling in for stochastic misses.

Question 1. Short-term Singing Rate Predicts Availability

A better predictive model of AVAILABILITY on singing rate could perhaps be achieved by modeling the power function, but that would require an estimate of CONTPROB from sample data. R / C_1 is such an estimator, but it is a bad one. In regressions of CONTPROB on estimates of this parameter, based on sample sizes of 1, 2, 5, 10, 25, 50, and 60 2-min sequences, the coefficient of determination varied from 0.01 to 0.05. Why is the fit so bad? CONTPROB operates at the level of the song, and with 30 songs possible in a minute, its effect apparently is buffered at this level. Minute-to-minute autocorrelation is measured by the number of runs per hour, which is complexly related to both CONTPROB and PROBSING. Runs per hour increases with CONTPROB at high probabilities of singing, but decreases with CONTPROB at low probabilities of singing (Fig. 4). Estimating CONTPROB from sample data will therefore probably require estimates of the inter-song interval, which would be too tedious and time-consuming for point-count observers. It appears that a simple method for estimating availability from singing rate may be difficult to devise.

Question 2. Short-term Presence/Absence of Song Predicts Availability

The fact that a small number of very brief samples does not produce an adequate estimate of availability is not particularly damaging to the removal model, because it, like other recent models, does not attempt to estimate N for individual point count stops. In these models, individual point counts are pooled, leading to a single estimate of P, resulting in a single estimate of *N* or *D* for the collection of sites. This is in contrast to spot-mapping, area search, and other intensive methods that produce estimates of *D* for each intensively studied area.

Even if two 1-min count periods were used with the Farnsworth et al. (2002) model, a sample of 25 counts would yield a very serviceable estimate of availability, according to the present study. Using more than two periods, of more than 1-min duration, as proposed (Farnsworth et al. 2002), should improve the estimate, although this assumption should be tested. Moreover, lower mean continuation probabilities in nature would raise the upper curve in Figure 3.

Availability is a continuous variable that can assume any value from 0 through 1. In this simulation, 0 and 1 are the only possible counts in a single sample, so they are also the only possible values of R_{11}. The number of possible values for this estimate of AVAILABILITY is of course $n + 1$, where n is the number of samples taken. In the simulation, values of 0 and 1 were modal up through $n = 25$, which was the point at which r^2 reached 0.9. In the removal method (Farnsworth et al. 2002), the count of individuals can be > 1, so the same precision in the estimate of availability should be achievable at a smaller sample size.

RELEVANCE TO BBS

The results of this simulation study, coupled with the recent release of the removal model (Farnsworth et al. 2002), offer hope that detection probability, or at least a major component of it (availability), can be estimated from data collected on the BBS, without major expense or inconvenience to the observers. The observer would simply divide the 3-min count period into two consecutive time segments, then tally all individuals detected in the first segment, and classify each individual detected in the second segment as either “new” or “recapture.” Estimation of availability would become part of the data analysis conducted by the BBS office at Patuxent.

Estimates of availability obtained with the removal method, or other methods based on mark-recapture logic, will be immune to observer variability, because observer-specific sources of bias will affect the counts in both segments of each stop, and therefore will cancel out. This feature is one of the main practical benefits of separating detection probability into availability, which is independent of the observer, and detectability, which (as defined above) is a function of observer aptitude.

Availability does vary with environmental conditions (e.g., wind, temperature), of course, but these are likely to be highly correlated throughout the route. Because availability will be estimated for each route, another major source of bias, the effect of environmental conditions, will also be estimable. The BBS has always collected data on wind conditions, but heretofore it has been possible only to correlate counts with Beaufort state. Raw counts, however, are confounded by abundance, which is not “caused” by wind conditions, while availability to an important degree probably is. Regressing availability on wind conditions will be much more informative than regressing estimates of abundance on wind conditions.

Recently, Rosenberg et al. (2002) estimated the change in availability with time of day for each species well-surveyed by the BBS. This enabled them to calculate a species-specific correction factor for each stop on a BBS-route. These corrections were relative to the stops with the highest counts, typically those early in the counts (i.e., early in the morning), but in some cases (e.g., buteos) toward the end of the route. An availability of 1 was implicitly assumed for these stops. The mean availability calculated from the temporal corrections was taken to be detection probability (i.e., including detectability as well as availability), and abundance of each species on the BBS, and in the BBS sample area, was estimated. Several years of route-specific availability estimates would make it possible to increase the accuracy of these estimates of abundance, because the true maximal availability would be estimated more accurately. This would result in either no change or upward adjustments of the estimates of abundance. Conversely, these BBS-wide estimates of temporal changes in availability will be useful in correcting the route-specific estimates for temporal effects. More work will be needed to determine the optimal integration of the two estimates.

COUNTS VS. PRESENCE/ABSENCE

For the half-century or so in which quantitative land bird surveys have been conducted, the parameter of interest has been abundance, N, or density, $D = N / A$, where A is the area in which the count was taken. During this time, presence/absence data have also been collected, often less formally, and these data have typically been treated with less confidence. The movement to reform land bird survey techniques by giving serious attention to estimating detection probability, P, (Nichols et al. 2000, Buckland et al. 2001, Bart and Earnst 2002, Rosenstock et al. 2002, Farnsworth et al. 2002, and especially Thompson 2002), was first focused on methods that estimate N. Recently, estimation of detection probabilities for presence/absence data has also been addressed (MacKenzie et al 2002, Royle and Nichols 2003). For these latter models, the parameter of interest is Proportion of Area Occupied (PAO).

These results of this study are relevant to PAO models, also, as to any method that includes a means, explicit or implict, of estimating availability. I chose to model availability of a single bird in this study, because the concept of availability is largely unexplored, and it seemed prudent to take the exploration one step at a time. This restriction to a single bird does not mean that this is a presence-absence model. The bird is always present in this model, it is only its behavior and availability that vary. In presence-absence models, birds are either present or absent, and determining whether apparent absence is a false or true negative is complicated by their availability and detectability.

This study has shown once again the usefulness of simulation modeling for exploring the intricacies of detection probability (see Royle and Nichols 2003). The results were counterintuitive, at least for the author, which shows that no question should go uninvestigated. The encouraging results of this study should lead in two directions. First, additional simulation modeling should explore the effects of multiple time periods, and multiple individuals, on the estimation of AVAILABILITY with R_{11}. In other words, the simulation model should be as general at the situations for which the removal model (Farnsworth et al. 2002) was designed. Second, the conclusions reached here should now be validated with real data, because the results reported here suggest that the usefulness of such data will justify the expense of collecting and processing them.

ACKNOWLEDGMENTS

This study was funded by contract 02ERSA0327 from USGS, BRD, for the Patuxent Wildlife & Research Center, Laurel, Maryland. I appreciate the cooperation and encouragement of Keith Pardieck and Bruce Peterjohn in the planning and execution of this project.

LITERATURE CITED

Bart, J., and S. Earnst. 2002. Double sampling to estimate density and population trends in birds. Auk 119:36-45.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: Estimating abundance of biological populations. Oxford Univ. Press, Oxford, UK.

Hailman, J. P., M. S. Ficken, and R. W. Ficken. 1985. The 'chick-a-dee' calls of *Parus atricapillus*: A recombinant system of animal communication compared with written English. Semiotica 56:191-224.

Huff, M. H., K. A. Bettinger, H. L. Ferguson, M. J. Brown, and B. Altman. 2000. A habitat-based point-count protocol for terrestrial birds, emphasizing Washington and Oregon. Gen. Tech. Rep. PNW-GTR-501. Portland, OR; U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 39 pp.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm.. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248-2255.

McCallum, D. A. In press. A conceptual guide to detection probability for point counts and other count-based survey methods. Proceedings of Partners in Flight International Meeting, March 2002. U.S. Department of Agriculture, Forest Service General Technical Report PNW-xxx.

Nichols, J. T., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117:393-408.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777-790.

Ralph, C. J., J. R. Sauer, and S. Droege. 1995. Monitoring bird populations by point counts. U.S. Department of Agriculture, Forest Service General Technical Report PSW-149.

Rosenberg, K. V., and P. J. Blancher. 2002. Estimating population size and setting conservation targets for North American land birds. Poster, North American Ornithological Conference, New Orleans, LA, September 2002.

Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. Land bird counting techniques: Current practices and an alternative. Auk 119:46-53.

Scott, T. A., P.-Y. Li, G. Greene, and D. A. McCallum. Ms. Singing rate and detection probability: An example from the Least Bell's Vireo (*Vireo belli pusillus*).

Thompson, W. T. 2002. Toward reliable bird surveys: Accounting for individuals present but not detected. Auk 119:18-25.

FIGURE LEGENDS

Figure 1. Effect of probability of singing (PROBSING) on AVAILABILITY. Each curve is for a different value of continuation probability (CONTPROB). Each data point is the mean of 10 replicates for the specified combination of CONTPROB and PROBSING. AVAILABILITY increases rapidly with PROBSING (and also with songs per minute (SPM), which is perfectly correlated with PROBSING), but this increase is slowed increasingly by increasing values of CONTPROB. The tendency to continue (which produces bout structure is this model) therefore decreases AVAILABILITY. See Figure 2 for more detail.

Figure 2. Effect of continuation probability (CONTPROB) on AVAILABILITY. Each curve is for a different value of PROBSING. Each data point is the mean of 10 replicates for the specified combination of CONTPROB and PROBSING. For each value of PROBSING, increasing CONTPROB leasds to monotonic decrease in AVAILABILITY. AVAILABILITY is $<< 1$ for PROBSING < 0.1. The expected value of AVAILABILITY is 1 for PROBSING $= 0.05$ (see tex), so the 20% reduction in AVAILABILITY at CONTPROB $= 0$ is the effect of stochasticity. Therefore, according to this graph, PROBSING $= .15$, which has an expected singing rate of 4.5 songs per minute, is needed to counteract stochasticity when singing is not organized into bouts.

Figure 3. Coefficient of determination (r^2) for the prediction of AVAILBILITY from brief estimates of "recapture rate" (R_{11}, the proportion of minutes containing a song that were preceded by a minute containing a song) and singing rate (SONGS), based on different sample sizes. The estimate of R11 was based on two consecutive 1-min time segments; the estimate of SONGS was based on 1-min samples. The coefficient of determination quantifies the proportion of the variation in AVAILABILITY explained by the predictor, and is based in each case on 10 reps of all combinations of CONTPROB and PROBSING, so these results are general. Although brief estimates of singing rate do explain some 40% of the variation in AVAILABILITY, this inadequate fit cannot be improved by increasing the number of such samples. Pooling brief estimates of recapture rate, on the other hand, can give a very good estimate of AVAILABILITY.

Figure 4. Runs per hour as a function of CONTPROB. In this case a run is a string of minutes of silence or of singing (at least one song in each minute), so the maximum number of runs is 60. Runs measure the amount of nonrandomness in sequences of two character states; a low number indicates clumping, while a high number means the data are hyper-dispersed. For the purposes of the present study, the most important point illustrated by this graph is that runs per hour does not bear a simple relationship to CONTPROB, and therefore runs, though relatively easy to measure in the field, will not serve as a useful estimator of CONTPROB. This graph also shows that minute-to-minute singing was most nearly random (high runs-per-hour) with low probabilities of singing and continuing or high values of both parameters. AVAILABILITY tended toward 1 with high PROBSING, regardless of CONTPROB (Fig. 1), and runs were constrained to very low values in these situations.

FIGURES

FIGURE 1

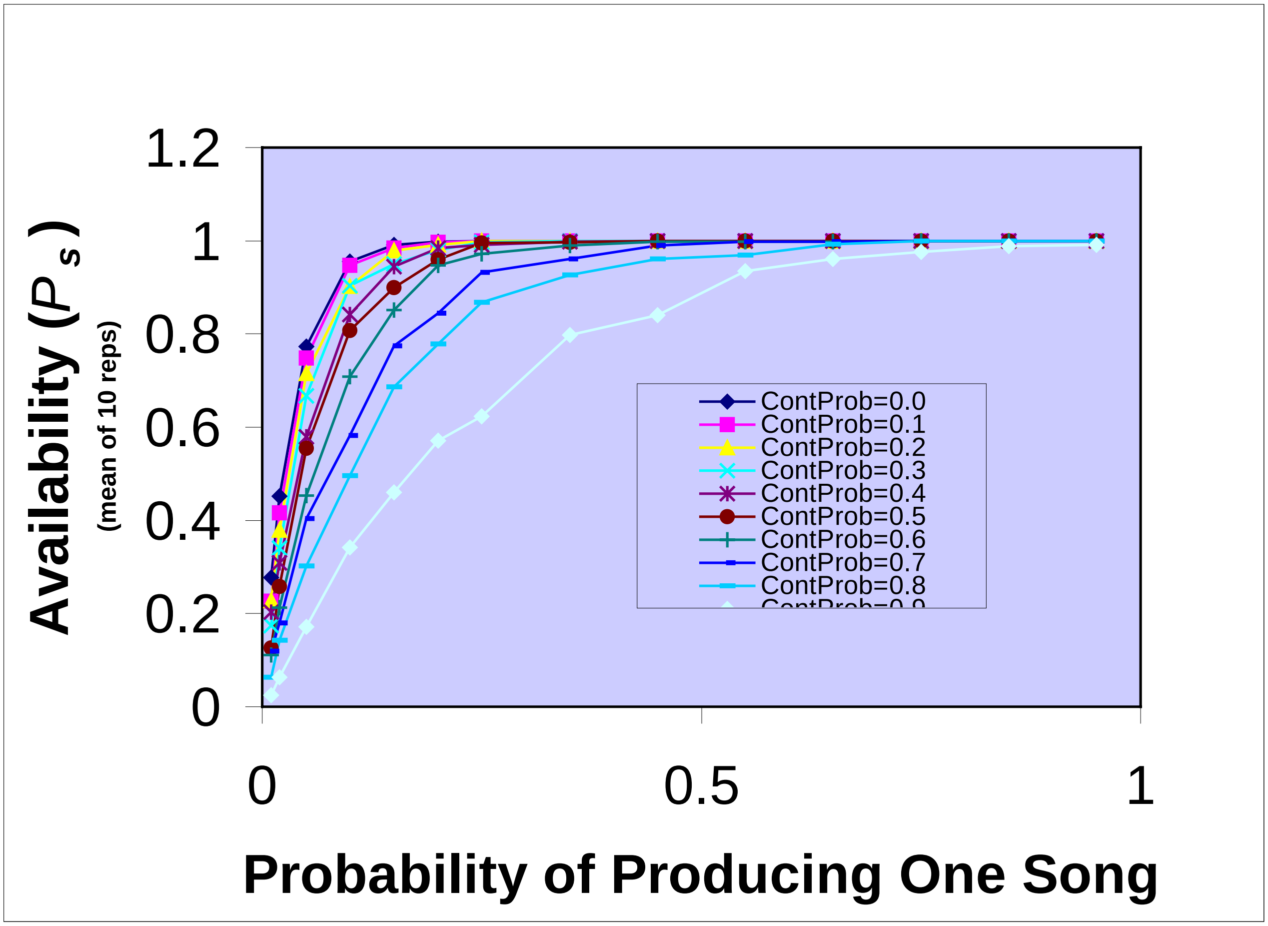

Figure 1. Effect of probability of singing (PROBSING) on AVAILABILITY. Each curve is for a different value of continuation probability (CONTPROB). Each data point is the mean of 10 replicates for the specified combination of CONTPROB and PROBSING. AVAILABILITY increases rapidly with PROBSING (and also with songs per minute (SPM), which is perfectly correlated with PROBSING), but this increase is slowed increasingly by increasing values of CONTPROB. The tendency to continue (which produces bout structure is this model) therefore decreases AVAILABILITY. See Figure 2 for more detail.

FIGURE 2

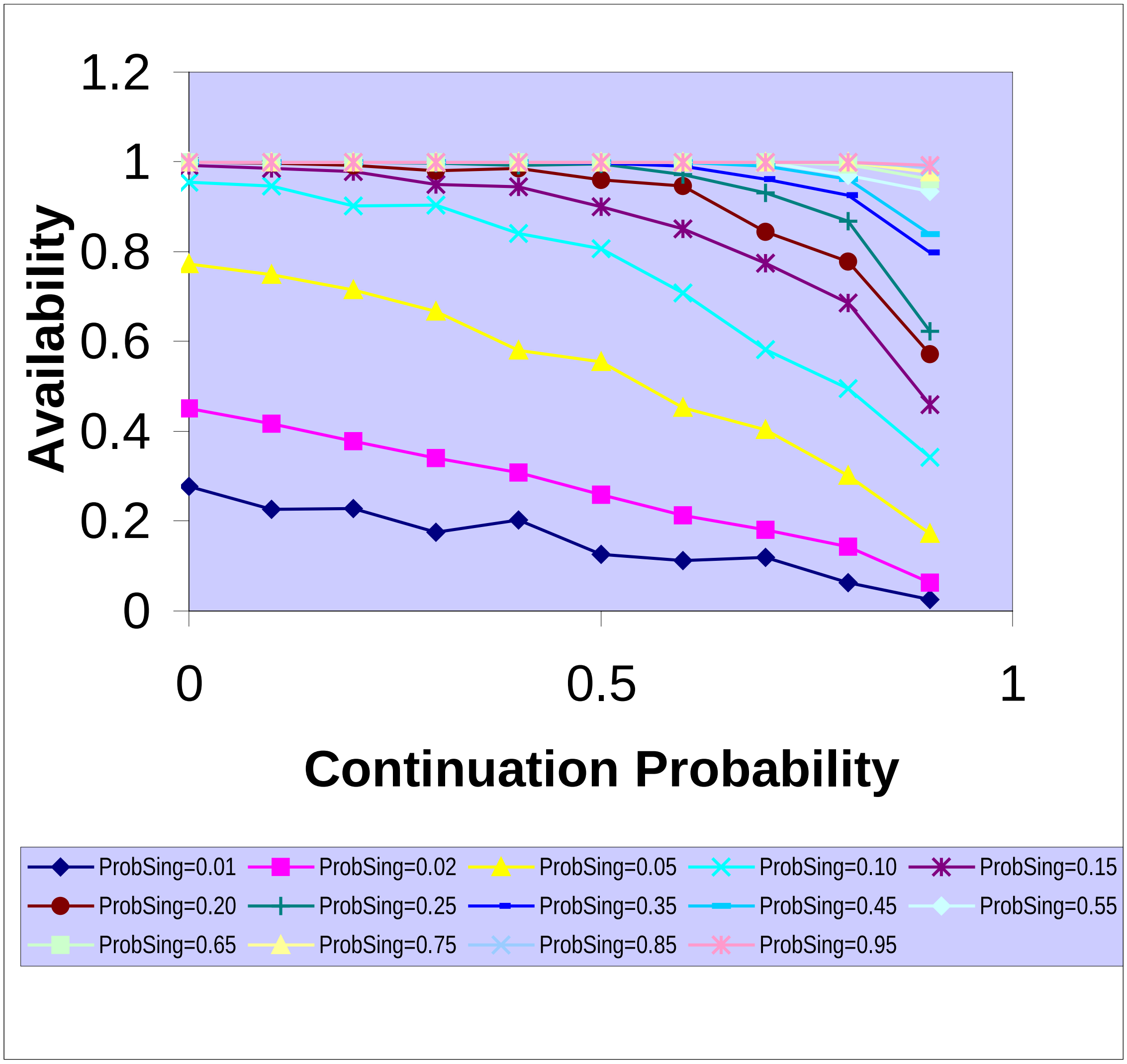

Figure 2. Effect of continuation probability (CONTPROB) on AVAILABILITY. Each curve is for a different value of PROBSING. Each data point is the mean of 10 replicates for the specified combination of CONTPROB and PROBSING. For each value of PROBSING, increasing CONTPROB leasds to monotonic decrease in AVAILABILITY. AVAILABILITY is << 1 for PROBSING < 0.1. The expected value of AVAILABILITY is 1 for PROBSING = 0.05 (see tex), so the 20% reduction in AVAILABILITY at CONTPROB = 0 is the effect of stochasticity. Therefore, according to this graph, PROBSING = .15, which has an expected singing rate of 4.5 songs per minute, is needed to counteract stochasticity when singing is not organized into bouts.

FIGURE 3

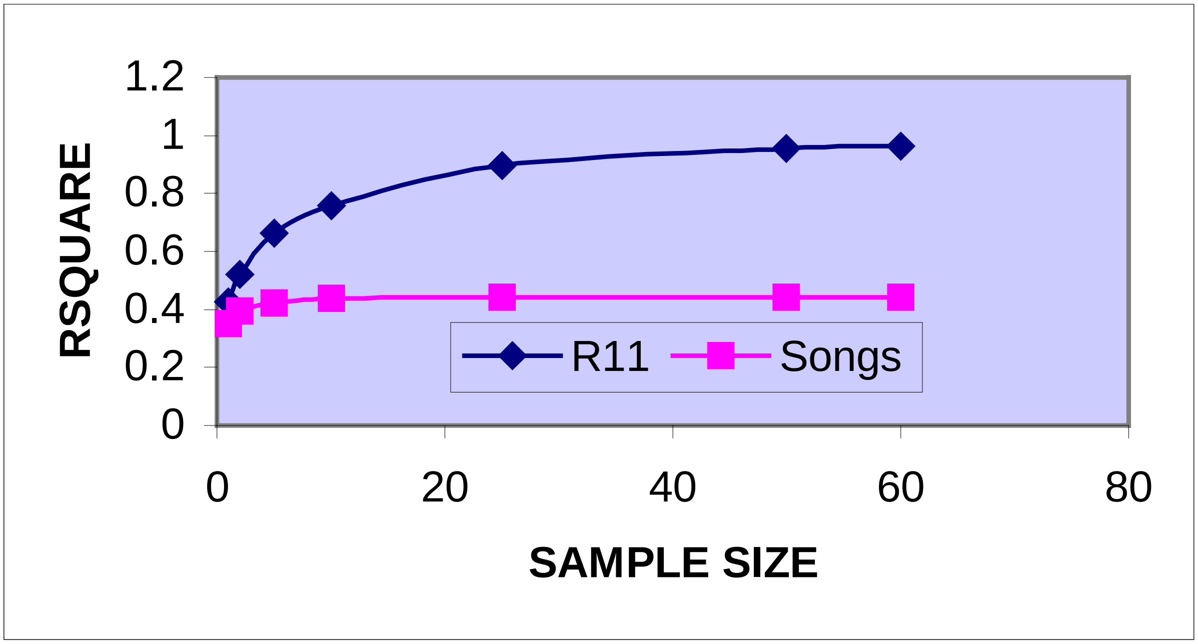

Figure 3. Coefficient of determination (*r*^{2}) for the prediction of AVAILBILITY from brief estimates of "recapture rate" (R_{11}, the proportion of minutes containing a song that were preceded by a minute containing a song) and singing rate (SONGS), based on different sample sizes. The estimate of R11 was based on two consecutive 1-min time segments; the estimate of SONGS was based on 1-min samples. The coefficient of determination quantifies the proportion of the variation in AVAILABILITY explained by the predictor, and is based in each case on 10 reps of all combinations of CONTPROB and PROBSING, so these results are general. Although brief estimates of singing rate do explain some 40% of the variation in AVAILABILITY, this inadequate fit cannot be improved by increasing the number of such samples. Pooling brief estimates of recapture rate, on the other hand, can give a very good estimate of AVAILABILITY.

FIGURE 4

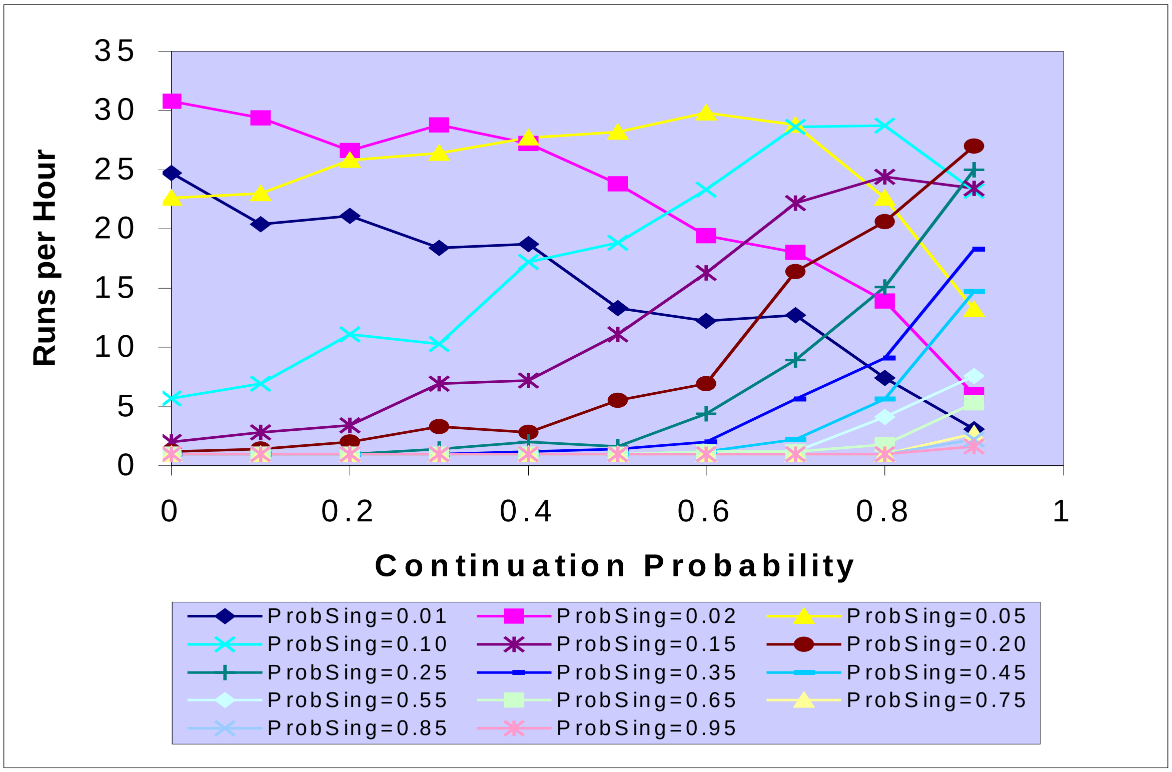

Figure 4. Runs per hour as a function of CONTPROB. In this case a run is a string of minutes of silence or of singing (at least one song in each minute), so the maximum number of runs is 60. Runs measure the amount of nonrandomness in sequences of two character states; a low number indicates clumping, while a high number means the data are hyper-dispersed. For the purposes of the present study, the most important point illustrated by this graph is that runs per hour does not bear a simple relationship to CONTPROB, and therefore runs, though relatively easy to measure in the field, will not serve as a useful estimator of CONTPROB. This graph also shows that minute-to-minute singing was most nearly random (high runs-per-hour) with low probabilities of singing and continuing or high values of both parameters. AVAILABILITY tended toward 1 with high PROBSING, regardless of CONTPROB (Fig. 1), and runs were constrained to very low values in these situations.