COENOSES 10(2-3):81-87, 1995.

ESTIMATING STATEWIDE SPECIES RICHNESS OF BREEDING BIRDS IN PENNSYLVANIA[1]

Glen D. Johnson and Ganapati P. Patil

Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University, University Park, PA., U.S.A.

Keywords: Biodiversity, Breeding birds, Covariate-directed sampling, Species richness estimation, Statewide species area curve.

Abstract. Our motivation is to design a sampling approach for estimating the species richness (total number of species) of breeding birds in Pennsylvania. The sampling frame is a tesselation of Pennsylvania by hexagons, each of size 635 km^2, that are a subset of the Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) hexagons that cover the United States. Using the concept of a species-area relationship, our objective is to select hexagons for successive aggregation into larger areas in such a way that maximally accelerates the species-area curve, thus minimizing the number of hexagons that require species enumeration. Using a known population of breeding birds, we conducted a retrospective analysis for the sake of designing future estimates of species richness when statewide survey data are not available.

Our results show that using tree species richness as a covariate for determining the order of selecting and aggregating hexagons causes the bird species-area curve to accelerate faster than either directional sampling (spatially contiguous aggregation) or random sampling. Since tree species richness is actually only weakly correlated with bird richness, as stronger covariates are identified, the performance of covariate-directed sampling is expected to perform even better.

Introduction

With such a strong and legitimate concern for the alarming loss of biodiversity on our planet (Wilson 1986, Stevens 1995), monitoring methods are essential for quantifying the biodiversity of large geographic regions. Such methods would allow the monitoring of trends in biodiversity over time in a given region, so that comparisons can be made with other factors like land use patterns. Results of such analysis can be used to indicate where biodiversity is decreasing, thus requiring more intensive conservation efforts, or where it may even be increasing, thus indicating what helps support biodiversity.

The question then arises as to what is an appropriate measure of biodiversity. With all the shortcomings of "indices", the single measure that seems to be the least controversial is species richness (the number of different species). For a relatively large geographic area like the state of Pennsylvania, biodiversity is best expressed as a 2-dimensional distribution of species richness that shows the heterogeneity across the landscape; however, data that is extensive enough to create such "maps" may only rarely be available. When updated full survey data are not available, we must rely on a sample based *estimate* of species richness. Meanwhile, biodiversity researchers recognize that reliable methods of estimation still require development (Yoon 1995).

Using breeding birds as an indicator of biodiversity, our objective is to develop a sampling approach to obtain a statewide estimate of breeding bird species richness. Since we are dealing with a non-additive variable, this is not a simple matter of estimating the total by multiplying a sample mean by the number of population units in the sampling frame. For this reason, we turn to species-area curves, which are used by ecologists for several reasons (Kilburn 1966), including the prediction of species richness in larger areas than those sampled (Evans, Clark and Brand 1955).

The species-area relationship basically states that as the area within a homogeneous habitat increases, the number of different species encountered will also increase until "a point of no return", after which increasing the area does not further increase the number of different species encountered. Our situation presents quite a different application of the species-area curve because our region of interest is the whole state of Pennsylvania, which is a montage of very diverse habitat

1 Prepared with partial support from the Statistical Analysis and Computing Branch, Environmental Statistics and Information Division, Office of Policy, Planning, and Evaluation, United States Environmental Protection Agency, Washington, DC, under a Cooperative Agreement Number CR-821531. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

types. Our sample units are, however, very large, as discussed in the next section, and therefore serve to smooth the smaller scale habitat heterogeneity.

We seek to design a sampling plan that maximizes the acceleration of a species-area curve towards it's plateau in order to minimize the area that needs to be sampled. This is especially critical when sampling from a very large geographic area like the state of Pennsylvania since this can rapidly become a very expensive exercise. In order to meet this objective, we want to exploit any readily available covariate information that can be used to direct sampling in a manner that increases the chance of encountering the maximum number of new species with each new sample unit.

Database

In support of the Pennsylvania Gap Analysis program, the Nature Conservancy (TNC) was contracted by the Environmental Protection Agency (EPA) to compile species lists within Environmental Monitoring and Assessment Program (EMAP) hexagons. Each hexagon is 635 km^2 in area. These data were provided to Penn State University in the form of DBASE IV^{TM} database files. Information was provided for the major vertebrate groups, along with some invertebrate groups and trees. The most accurate group is breeding birds, since they were compiled from the Pennsylvania Breeding Bird Atlas (cf. Brauning, 1992) which is a statewide census of almost 5000 blocks, where each block is equivalent to one sixth of a 7.5 minute U.S.G.S. quadrangle map. Other species groups were compiled from records that were assesed from institutions such as museums, colleges and state agencies.

Species were codified according to the degree of subjective probability that a given species occured, based on available evidence. Focusing on breeding birds for our analysis, we considered a species to be present in an EMAP hexagon if it was listed as confirmed, meaning that there is a 95% subjective probability that the species occurs.

Prior to constructing any species-area curves, all the EMAP hexagons that were not at least fifty percent within the state boundary were eliminated, twenty-five in total, resulting in a sample frame of 186 hexagons.

In order to group the hexagons so that the directional species-richness curves could be computed, each hexagon was assigned an x and y coordinate pair, resuting in two axes that form an obtuse angle.

Characteristics of the Breeding Bird Population

The distribution of species richness for breeding birds with respect to EMAP hexagons is displayed in Figure 1 in the form of a greyscale thematic map.

In order to characterize the statewide species-area relationship, we constructed four species-area curves for the entire state, where the smallest area corresponds to one

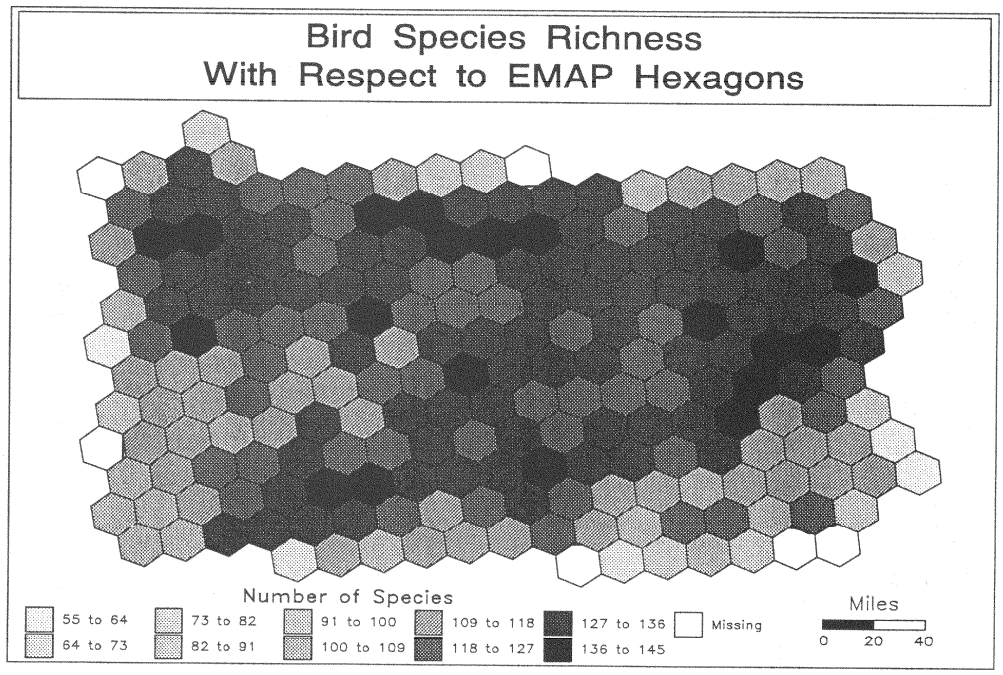

Figure 1. Bird species richness with respect to EMAP hexagons.

hexagon and the largest area corresponds to the whole state. Each of these curves was constructed by starting in one corner of the state and aggregating hexagons by expanding outward. The four directions were northeast, southeast, southwest and northwest.

The resulting curves are displayed on a common scale in Figure 2, where the shapes are seen to be very similar. Furthermore, all of the curves can be described by the conventional species-area relationship for sampling over homogeneous habitat (Usher, 1985), defined as:

$$S=kA^z \quad (1)$$

where S is the species richness and A is the area, while z and k are population specific parameters.

These results indicate that species-area curves can provide a very legitimate approach to estimating the total number of species when sampling over a large geographic area that constitutes a wide variety of habitats. This probably holds true in large part because the primary sample units are also very large in this application (635 km^2), thus smoothing out the effects of smaller scale habitat heterogeneity.

Sampling Strategies

The objective of a cost efficient sampling design is to minimize the sample size required for estimating the overall species richness of a larger area. This objective will in turn be satisfied by a sample that maximizes acceleration of the species-area curve so that the curve's plateau is attained sooner. In other words, we want to encounter all species, or almost all species, in a region of interest with as few a number of sample units as possible.

When constructing a species-area curve from successive aggregation of discrete sample units, the usual approaches are to either combine sample units in a continuous fashion or to combine units that are obtained at random from throughout the region of interest. The directional curves reported in Figure 2 are examples of continuous aggregation.

When habitat is diverse across the region, random aggregation may result in a steeper curve than is obtained from continuous aggregation because spatially discontinuous sample units may encounter more diverse habitats, therefore increasing the chance of encountering different species. Out

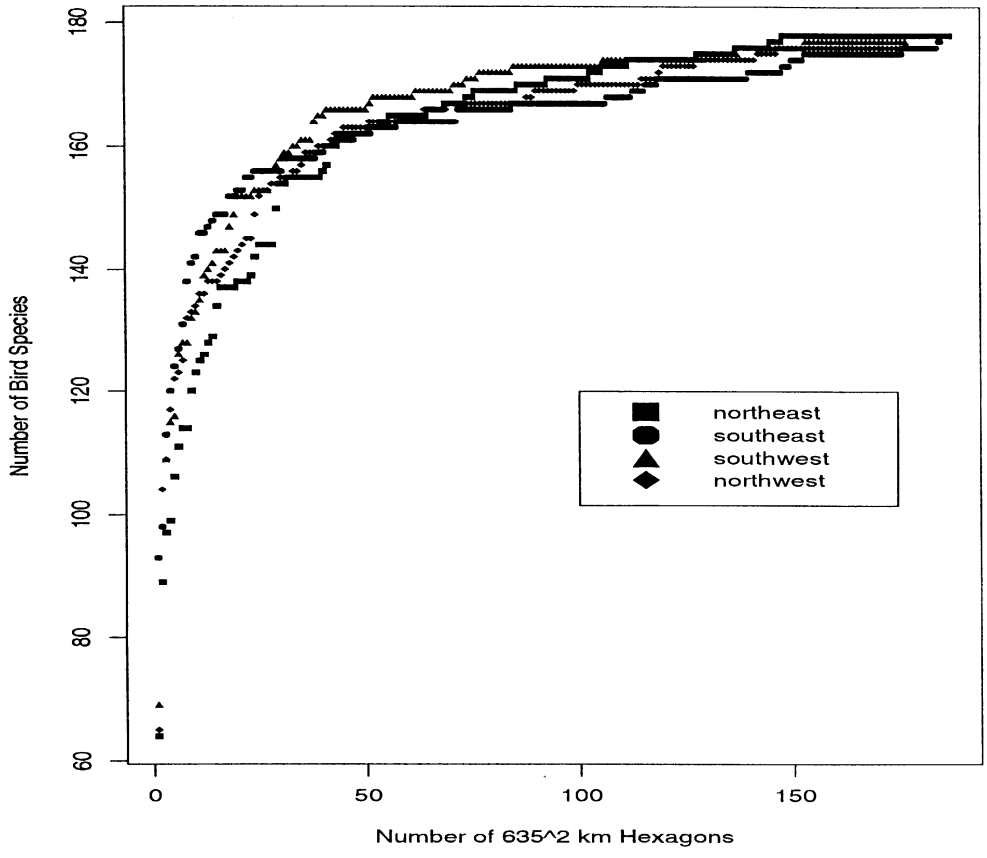

Figure 2. Bird species-area curves from aggregating hexagons in each of four designated directions (see legend).

of the $n!$ ways to randomly choose and aggregate n sample units, the expected number of species is

$$E[S_n] = \sum_{i=1}^{s} \left[1 - \frac{\binom{N-A_i}{n}}{\binom{N}{n}} \right] \quad (2)$$

where $E[S_n]$ is the expected number of species encountered in n sample units of equal size obtained at random, A_i is the number of sample units, out of N total units, that contain species i and s is the total number of species amongst the N sample units. This expression is also found elsewhere (Kobayashi, 1974; Engen, 1976) and is the same expression for "rarefaction" curves when N and n are numbers of individuals (organisms), instead of numbers of sample units (James and Rathbun, 1981).

Directed Sampling using a Covariate

An alternative approach to continuous aggregation or random sampling is to perform directed sampling based on values of some covariate that are readily available for the sample units. Such a covariate may be available through information contained in a geographic information system (GIS), either directly or from modeling. The desired property of a covariate would be to direct sampling in a manner that accelerates the species-area curve faster than would be observed with spatially continuous or random sampling.

Of the information available in our database, tree species presents the most promising covariate for choosing an optimal hexagon ordering for ultimately measuring bird species richness. We basically hypothesize that differences in bird species are likely to be associated with differences in tree species; therefore, if hexagons are chosen in an order that corresponds to maximum acceleration of the tree species richness curve, using this same ordering will accelerate the bird richness curve.

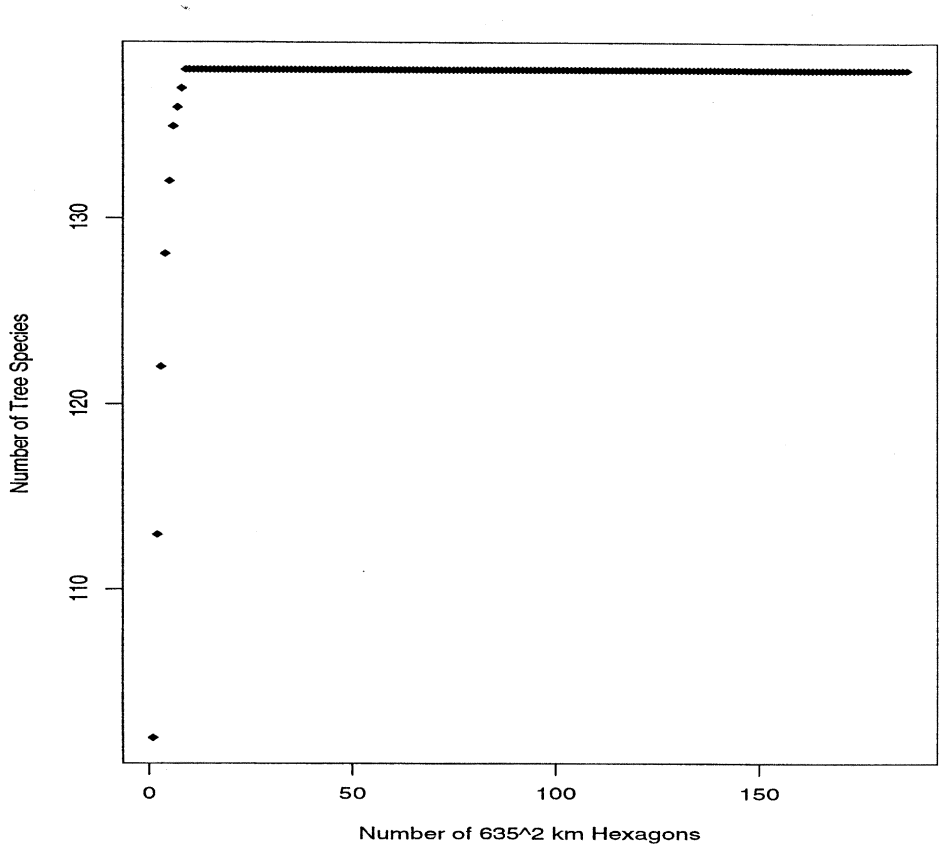

Figure 3. Tree species-area curve obtained from optimal aggregation of hexagons, using known species lists per hexagon for trees. Note that all tree species are accounted for within nine hexagons.

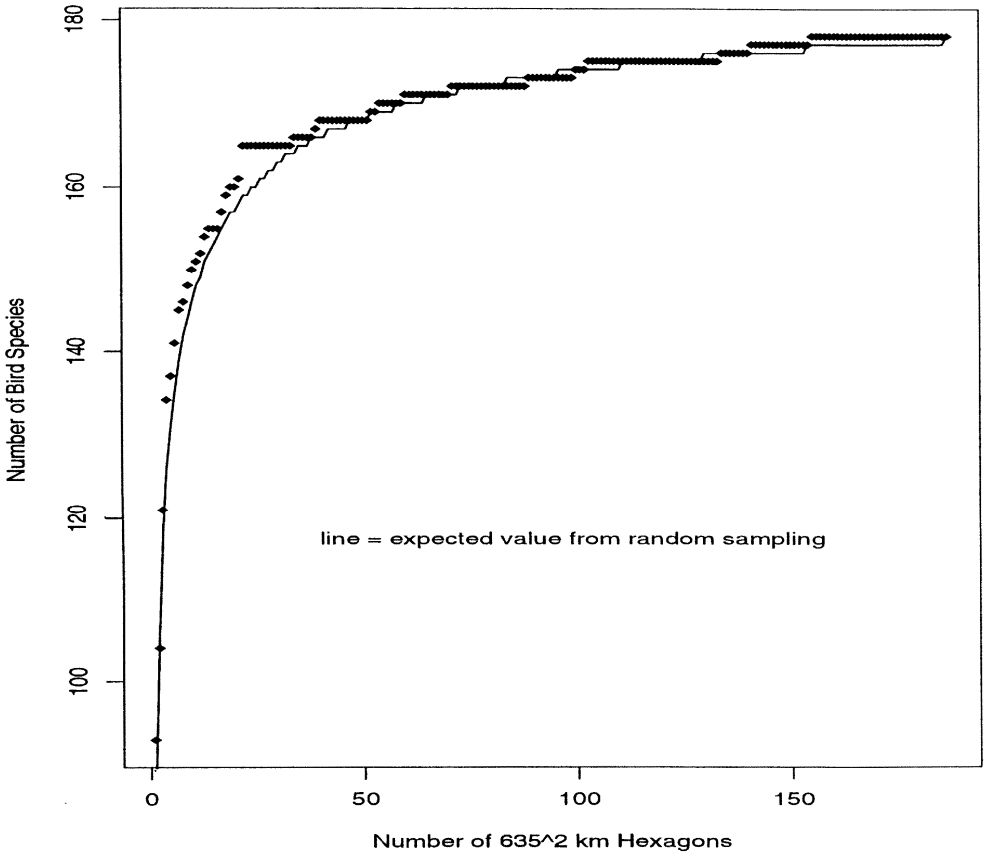

Figure 4. Bird species-area curve, based on selecting the first nine hexagons from optimal tree richness ordering, followed by random sampling/aggregation, as shown by points. The expected values from completely random sampling for n = 1 to 186 hexagons are indicated by the line.

Constructing the optimal tree species richness curve was performed by the following protocol:

1. Choose the first hexagon as the one containing the highest tree species richness.
2. After noting which species were in the first hexagon, delete all members of these species from the remaining hexagons.
3. Choose the second hexagon as the one now containing the highest tree richness.
4. Repeat steps 1 and 2 until all the tree species have been accounted for.
5. The remaining hexagons are then aggregated at random.

The resulting tree species richness curve from this protocol is seen in Figure 3, where we see that all tree species were accounted for within the first nine hexagons sampled.

The bird species richness curve that was constructed using the same ordering of the first nine "tree-directed" hexagons, followed by random sampling/aggregation, is reported in Figure 4. Under this protocol, 148 out of 178 total bird species were accounted for within the first nine "tree-directed" hexagons, which is higher than was obtained within nine hexagons for any of the continuously sampled/aggregated curves in Figure 2 (range=120-141).

Estimating Total Species Richness

Whether we are using a covariate-directed sample, a spatially continuous sample or a random sample, we can estimate the total species richness for a larger area that contains the sample units either by *(i)* fitting a model to the sample species-area curve, followed by extrapolation to the maximum area of the whole region of interest, or *(ii)* by the final cumulative species richness.

Table 1. Statewide bird richness estimates from modeling a sample which consists of the first nine hexagons obtained from optimum tree richness directed sampling, followed by random sampling.

	parameter estimates predicted			
final sample size	*k*	*z*	value	bias*
29	36.87	0.157	230	+52
39	42.76	0.139	216	+38
59	51.42	0.117	202	+24

*bias = predicted statewide value minus known value

Table 2. Statewide bird richness estimates from direct accumulation of species using the same samples described in the previous table. For comparison, the expected values from purely random sampling are provided.

final sample size	estimate (s.e.)*	bias**	expected value from random sampling
29	166 (0.27)	-12	163
39	168 (0.17)	-10	166
59	171 (0.19)	-7	170

* estimate = mean of 100 iterations; s.e. = standard error of the mean

**bias = predicted statewide value minus known value

note: median of 100 iterations equaled the mean for all cases

For our application of statewide bird species richness in Pennsylvania, Equation 1 presents an appropriate general model; however, when the parameters of this model are fit by a sample of hexagons that represent only the initial part of the true statewide curve, there will always be an upward bias when extrapolating to the whole state. A valuable area of research is to investigate more appropriate models or appropriate corrections for the general purpose model of Equation 1 for obtaining unbiased estimates (predictions) of species richness for a larger area that contains the sample units.

The simple method of using the final accumulated species richness as an estimate of overall species richness will be negatively biased, although it is possible to capture all the species of the larger area of interest.

Using the optimal "tree-directed" protocol for choosing the first nine hexagons, followed by random sampling with an additional 20, 30 and 50 hexagons, we obtained statewide bird species richness estimates that are reported in Tables 1 and 2 for both the model and direct accumulation estimates, respectively. For the direct accumulation estimate, random sampling was reiterated for each of the sample sizes one hundred times in order to estimate the standard error of this estimate. For comparison to purely random sampling/aggregation of hexagons, the expected species richness for each final sample size, considering all possible samples of size *n*, was calculated using Equation 2 and is provided in Table 2.

The direct accumulation approach is much less biased than the modeling approach and the bias errs towards conservative estimates of species richness. Furthermore, the standard error of the direct accumulation estimates indicate a high degree of precision. The gain in species encountered, with subsequent reduction in bias, appears to be very little for substantial increases in the number of hexagons included in the random component of the sampling.

Using Covariate-Directed Selection of All Sample Units

Using tree richness as a covariate was further investigated by ordering every hexagon in the population through repetition of the protocol defined earlier. After the first nine hexagons that accounted for all the tree species were identified, these hexagons were deleted from the database and the protocol was repeated for the remaining hexagons. After repeating this cycle until all hexagons were ordered, the bird species-area curve was then developed based on this ordering.

The resulting curve is seen in Figure 5, where the bird species richness is indeed consistently higher than the expected richness obtained through random sampling. This indicates that tree richness is correlated with bird richness at the measurement scale of EMAP hexagons (635 km^2), and that when tree species lists are available for each hexagon, this information can be used to reduce the number of hexagons that require analysis of bird species when constructing a sample-based statewide estimate.

Summary

Although species-area curves were developed for analyzing communities within fairly homogeneous habitat, we show that a classical hyperbolic response is obtained for breeding birds across the whole state of Pennsylvania. This observation may be largely due to the size of sample units used, each being a 635 km^2 hexagon, which serves to smooth out the effects of local habitat heterogeneity that occurs on smaller spatial scales.

Our primary objective is then to choose a sample of hexagons that describes enough of the statewide species-area curve to include the beginning of the "plateau" region of the curve. This would allow an estimate of the statewide species richness when the species enumeration is affordable for only a sample of hexagons. In order to obtain an economically efficient sample, we want to minimize the sample size (number of hexagons requiring species enumeration).

To this end, we found that tree species richness was a valuable covariate for ordering hexagons to construct a bird species area curve. Specifically, the hexagon order was chosen which optimally accelerates the tree species-area curve; then using this same ordering for enumerating bird species resulted in a bird species-area curve that accelerates to the plateau faster than is expected with random selection and aggregation of hexagons, which in turn appears to accelerate faster than a curve based on spatially continuous aggregation.

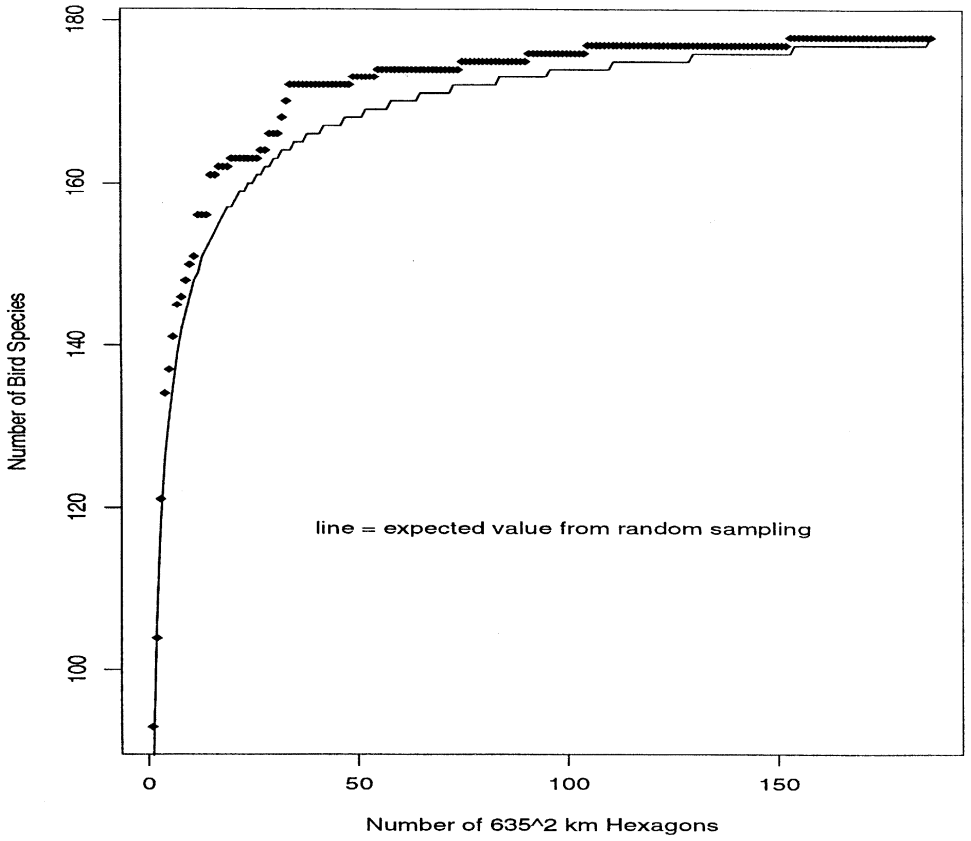

Figure 5. Bird species-area curve based on selecting all hexagons from optimal tree richness ordering, as shown by points. The expected values from completely random sampling for n = 1 to 186 hexagons are indicated by the line.

We feel that as habitat information becomes available in a GIS format, other covariates can be obtained that may further accelerate the bird species-area curve. Meanwhile, this approach of using GIS-based covariate information to choose an ordering of sample units for increasing the species-area curve acceleration for a particular species appears to reduce the number of sample units required for estimating the overall species richness of a region.

Although we may come very close to the true species richness of a region with a sample of only a fraction of that region, we still need truly unbiased estimators of true richness based on such samples.

Acknowledgements. We are grateful to Veronica Poscik for performing most of the computational work while participating in the National Science Foundation's "Research Experience for Undergraduates" program at Penn State University, Department of Statistics.

References

Brauning, D.W. 1992. Atlas of Breeding Birds in Pennsylvania. University of Pittsburgh Press, Pittsburgh. 484 pp.

Engen, S. 1976. A note on the estimation of the species-area curve. J. Cons. int. Explor. Mer, 36(3):286-288.

Evans, F.C., Clark, P.J. and Brand, R.H. 1955. Estimation of the number of species present on a given area. Ecology, 36:342-343.

James, F.C. and Rathbun, S. 1981. Rarefaction, relative abundance, and diversity of avian communities. The Auk, 98:785-800.

Kilburn, P.D. 1966. Analysis of the species-area relation. Ecology, 47:831-843.

Kobayashi, S. 1974. The species-area relation I. A model for discrete sampling. Res. Popul. Ecol., 15:223-237.

Stevens, W.K. 1995. How many species are being lost? Scientists try new yardstick. in The New York Times, p. C4, July 25, 1995.

Usher, M.B. 1985. Implications of species-area relationships for wildlife conservation. J. Environ. Mngt., 21:181-191.

Wilson, E.O. 1988. Biodiversity. National Academy Press, Washington, D.C., 521 pp.

Yoon, C.K. 1995. Monumental inventory of Costa Rican forest's insects under way. in The New York Times, p. C4, July 11, 1995.