PENNSTATE

1855

Center for Statistical Ecology
and Environmental Statistics

MULTISCALE ANALYSIS OF THE SPATIAL DISTRIBUTION OF BREEDING BIRD SPECIES RICHNESS USING THE ECHELON APPROACH

by Glen Johnson[1], Wayne Myers[2], Ganapati Patil[1], and David Walrath[2]

[1]Center for Statistical Ecology and Environmental Statistics
Department of Statistics
[2]School of Forest Resources and
Environmental Resources Research Institute

The Pennsylvania State University
University Park, PA 16802

NSF Project Officer: Penelope Firth
EPA Project Officer: Lawrence H. Cox, Barry Nussbaum, and Anthony R. Olsen
NBS Project Officer: Reid Goforth

Prepared with partial support from the National Science Foundation Cooperative Agreement Number DEB-9524722, U. S. Dept. of Interior, United States Environmental Protection Agency Cooperative Agreement Numbers CR-821783 and CR-821531, and National Biological Service Cooperative Agreement Number 14-16-0009-1548. 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.

Invited plenary paper presented at the Conference on Assessment of Biodiversity for Improved Forest Planning, Monte Verita, Switzerland, October 1996.
Paper appears in *Assessment of Biodiversity for Improved Forest Planning*, P. Bachmann, M. Kohl, and R. Paivinen, eds. Kluwer Academic Pubishers. Pp. 135—150.

Technical Report Number 96-1101
TECHNICAL REPORTS AND REPRINTS SERIES
November 1996

Department of Statistics
The Pennsylvania State University
University Park, PA 16802

G. P. Patil
Distinguished Professor and Director
Tel: (814)865-9442 Fax: (814)865-1278
Email: gpp@stat.psu.edu
http://www.stat.psu.edu/~gpp

MULTISCALE ANALYSIS OF THE SPATIAL DISTRIBUTION OF BREEDING BIRD SPECIES RICHNESS USING THE ECHELON APPROACH

GLEN JOHNSON[1], **WAYNE MYERS**[2],
GANAPATI PATIL[1], **and DAVID WALRATH**[2]
[1]*Center for Statistical Ecology and Environmental Statistics,*
Department of Statistics
[2]*School of Forest Resources*
and Environmental Resources Research Institute
Penn State University, University Park, PA 16802 USA

Abstract.—As an important component of biodiversity monitoring, species richness is a response variable of considerable interest to conservation biologists. Monitoring over a large spatial extent (such as statewide) requires an objective method for characterizing the spatial distribution of species richness. Objectivity is needed for delineation of areas that are relatively species rich or poor, and to also provide a way to compare other spatial distributions, as with temporal monitoring for change detection. The echelon approach introduced by Myers, Patil and Taillie (1995) provides a way to objectively characterize spatial structure of response variables that are at least ordinal. We apply this method to the characterization of breeding bird species richness across the state of Pennsylvania, using US-EPA EMAP hexagons which constitute a very coarse grain (635 km^2). After identifying relative species rich areas at the coarse measurement scale, the hexagons covering such areas are analyzed in much finer detail, using the same echelon protocol but with a measurement scale (grain) equal to approximately 24 km^2 rectangles corresponding to one sixth of a 7.5 minute USGS topographic quadrangle.

---

1 Introduction

In order to minimize expansion of the endangered species list, preventative medicine is required. This includes protection of critical habitat areas, as identified from a landscape to regional perspective. In the United States, the national GAP Analysis program responds to this need by aiming at the determination of "gaps" in protection of critical habitat areas (Scott, et al. 1993). As part of the Pennsylvania Gap program, species richness mapping has been performed statewide using data that was compiled at the scale of 635 km^2 hexagons that constitute a subset of the USEPA-EMAP national tessellation (Myers, Brooks, Storm and Bishop, 1995).

Following the protocol introduced by Myers, Patil and Taillie (1995), and expanded in greater detail by Myers, Patil and Joly (1996), the statewide species richness for various vertebrate groups was spatially organized into hierarchically related echelons using the EMAP hexagons as the primary mapping units. Of these different groups, bird species were compiled from the Pennsylvania Breeding Bird Atlas (cf. Brauning, 1992) database which is based on mapping units that equal one sixth of a 7.5 minute USGS quadrangle (approximately 24 km^2).

In this paper, we use the coarse-grained (EMAP hexagon) statewide echelon map of breeding birds to identify areas of the state that exhibit regionally high species richness from a statewide perspective. We then apply echelon mapping to a selection of these areas using the much finer-grained original data from the Breeding Bird Atlas (Brauning and Gill, 1983-1989).

Results are presented with some assessment of the spatial structure of species richness using the finer resolution mapping units within those areas of regionally high species richness. A primary research objective is to decipher what echelon patterns at the finer scale may lead to different conclusions between the two measurement scales about locally species-rich areas within a regionally rich area.

2 The echelon concept

Consider a geographic area that is either systematically tessellated into equal size and shape sectors or is partitioned into irregular size sectors, such that a variable of interest has a realized value within each sector. Often, a primary interest is to identify "hot spots", which are areas of concern due to either high or low response values.

For certain applications, a benchmark for hot spot determination may be provided by a numerical criterion such as a chemical action level or cleanup standard, or perhaps by a percentile of the data (Johnson, Nussbaum, Patil and Ross, 1995). In the absence of a numerical criterion, hot spots may be defined as areas having high or low responses "relative to their surroundings". Indeed, a hazardous waste site manager may want to identify areas of relatively high contaminant concentrations, regardless of a numerical criterion.

In other applications, such as identification of areas across a given landscape that reveal high or low biodiversity, there simply may not exist objective criteria for defining an area of concern. As an example, consider the thematic presentation of species richness in Figure 1. While human visual perception can see a pattern of relative highs and lows, the delineation of boundaries for local to regional areas of high (or low) species richness becomes a subjective exercise. The visual cues for subjectively distinguishing relative highs are influenced by the arbitrarily chosen number of greyscale categories in this particular map.

Amongst the potential applications of the echelon approach (Myers, Patil and Joly, 1996), this concept can serve to objectively delineate areas of relative high or low values of a regionalized response variable by constructing hierarchically related spatial objects from the original data. First, local peaks and plateaus of the response variable are identified amongst all sectors (regular cells or irregular polygons) of a spatial partitioning. Then, echelon objects of the first order are identified by moving outward and downward from the local peaks and plateaus until sectors are reached that yield a change in the response variable towards increasing values. Sectors that constitute such local minima in the virtual topography serve to delineate the boundaries of first order echelons. All of the sectors within each first order echelon then become a member of that echelon. This process is similarly repeated to identify higher order echelon objects which are founders (parents) of lower order echelon objects (children). A top-view perspective can then be mapped where each sector is assigned to a unique echelon object which in turn has an order ranging from one (1) to the highest order echelon object. This spatial structure can then be represented in a similar manner as streams are organized in a drainage basin, where first order streams have no tributaries, second order streams are formed by the confluence of first order streams, etc.. By this protocol, different observers will arrive at the same echelon structure.

The hierarchical nature of echelons further provides a means of delineating high and low areas at various spatial scales. In other words, once a large regionally high area is delineated, then sub-regional to local highs can be delineated within this region. This is especially important for large scale monitoring since environmental management is done at various scales (such as Statewide→ County → Township; or Chesapeake Bay Watershed → Susquehanna→ Conestoga→ etc.). The purpose is to *objectively* delineate relative high/low areas at multiple scales, whereas simply viewing a thematic map or a response surface can result in multiple interpretations of relative high/low, depending on who the interpreter is.

3 Data

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 Pro-

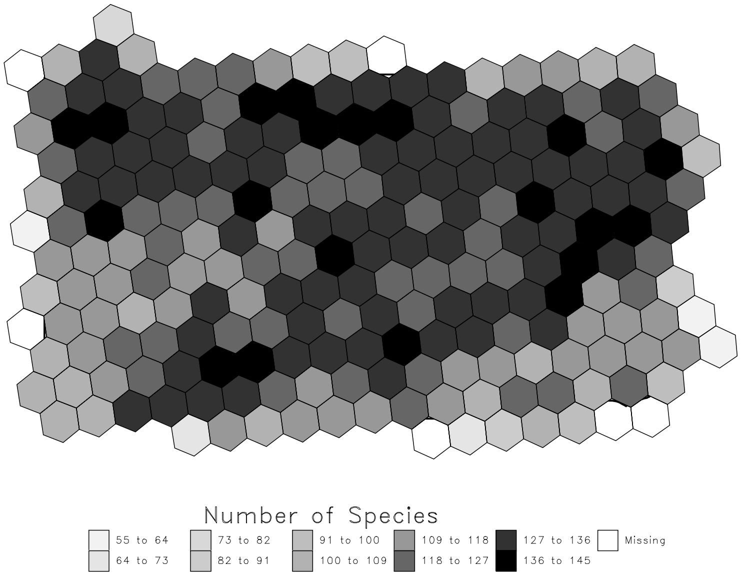


Figure 1: Bird richness in the hexagons.

gram (EMAP) hexagons. Each hexagon is 635 km^2 in area. 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.

Species were codified according to the degree of subjective probability that a given species occurred, based on available evidence. Focusing on breeding birds for our analysis, we treated a species as present in a block or hexagon if it was listed as confirmed or probable (at least an 85% chance of breeding based on evidence).

4 Statewide Mapping with EMAP Hexagons

The statewide echelon map based on the coarse grained EMAP hexagons is presented in Figure 2. Regions of high species richness, relative to their surroundings in a statewide perspective, were delineated by complexes of first and second order echelons. General correspondence can be seen with the thematic map in Figure 1; however, the boundaries of regionally high areas in Figure 1 are not clearly defined, whereas the echelon objects of Figure 2 serve to objectively define such boundaries.

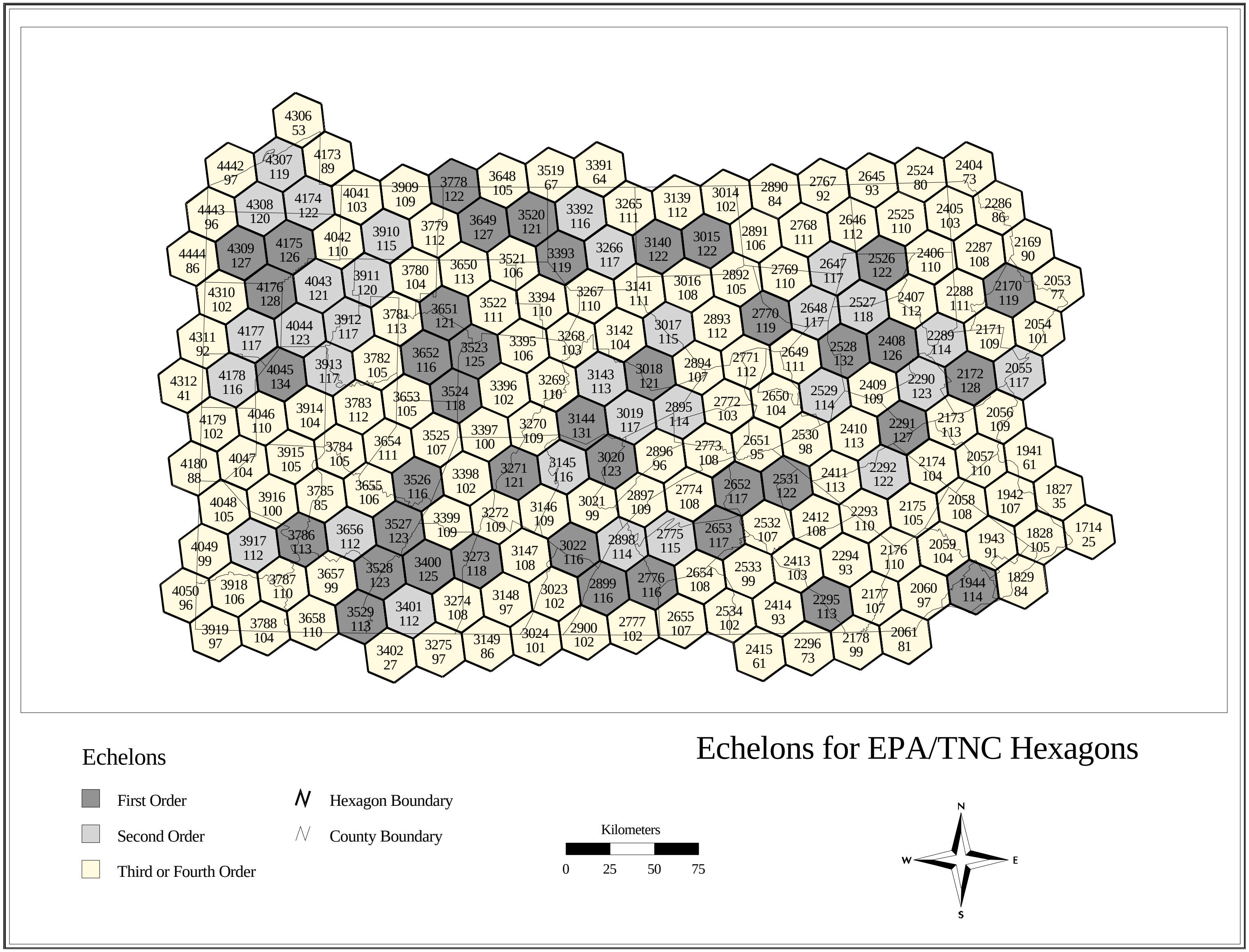
For the purpose of exploratory data analysis, we focused on the two largest complexes, one in the northwest and one in the northeast. Since we know the underlying land use is quite diverse for both of these complexes, for comparison we also chose a smaller complex that appears in the northcentral part of Pennsylvania which has more homogeneous land use, consisting largely of forest interior with some agricultural land.

5 Regional Mapping with Atlas Blocks

The regional areas that were further analyzed are presented in isolation from the remaining state, along with the finer resolution echelon maps within these areas, in Figures 3 to 8.

In all three regions, the finer scale mapping revealed a complexity of surface response that could not necessarily be expected from what was observed at the coarser scale. Therefore, while the coarse scale of EMAP hexagons may provide a way to identify regionally high areas from a statewide perspective, finer scale analysis is necessary for determining possible locations and networks of high species richness within a region.

Figure 2: Statewide echelon map based on EMAP hexagons. The 4-digit number in each hexagon is the EPA-EMAP identifier, while the number below is species richness.



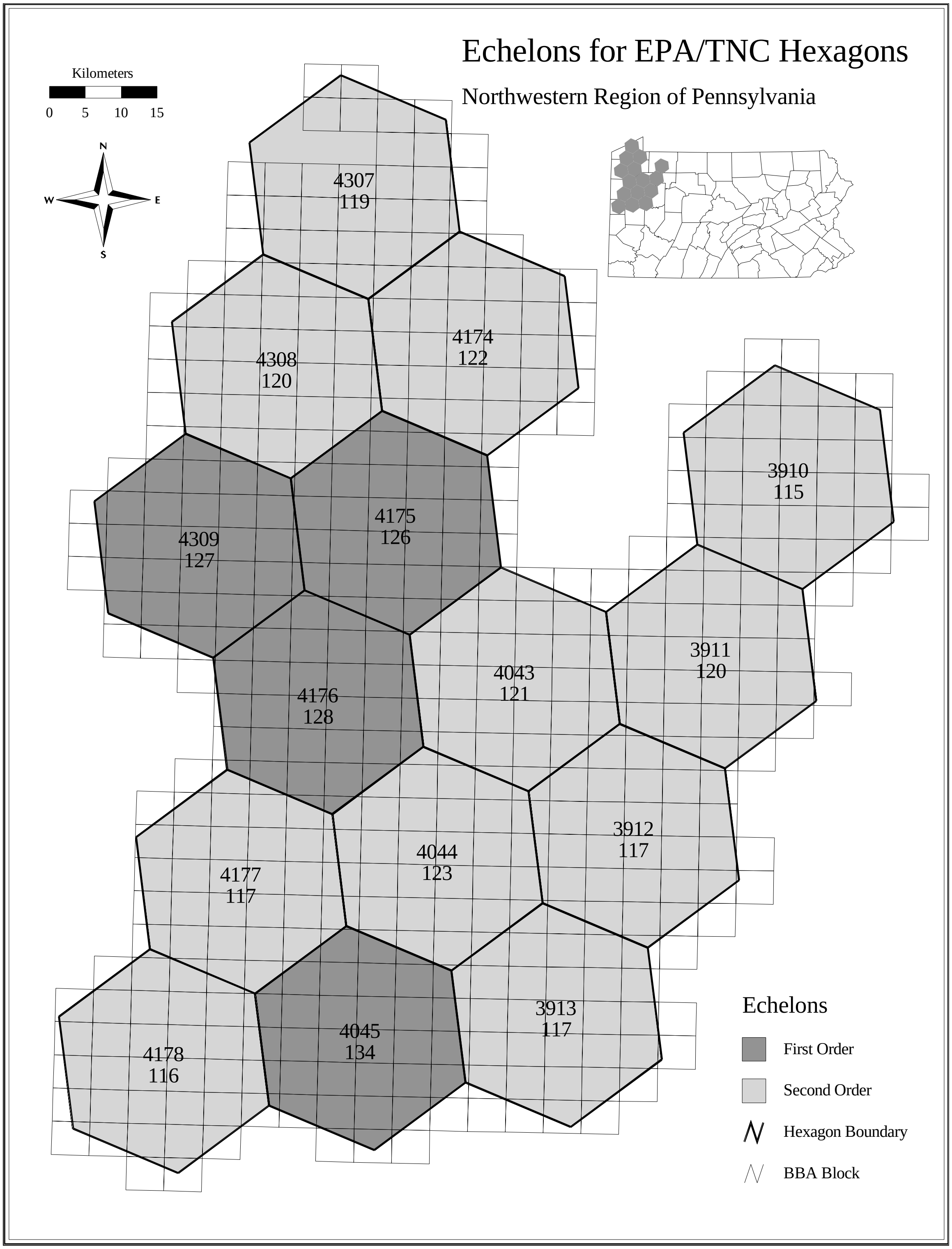


Figure 3: Northwest region EMAP hexagons. The 4-digit number in each hexagon is the EPA-EMAP identifier, while the number below is species richness.

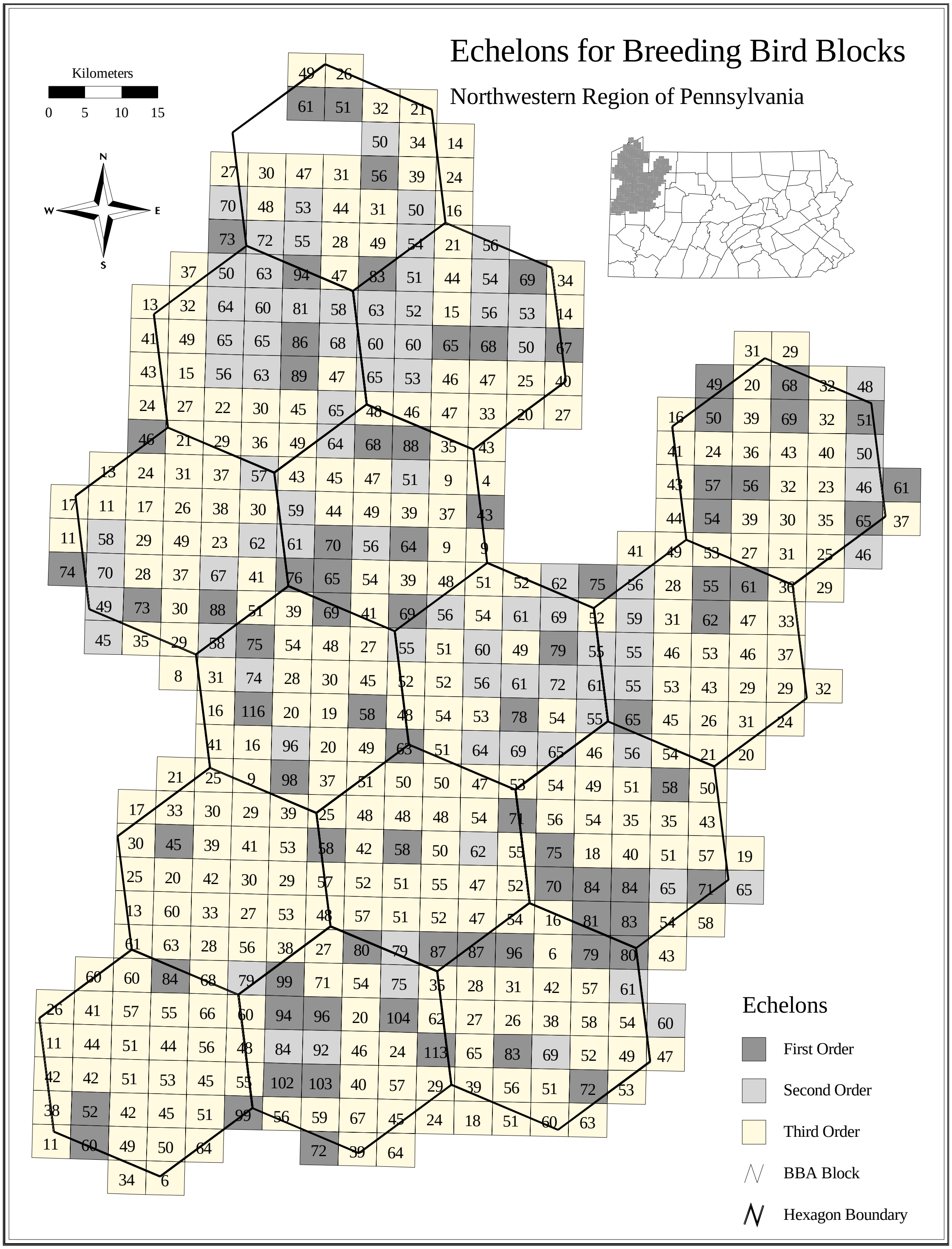


Figure 4: Northwest region Breeding Bird Atlas blocks. The number in each block is the species richness.

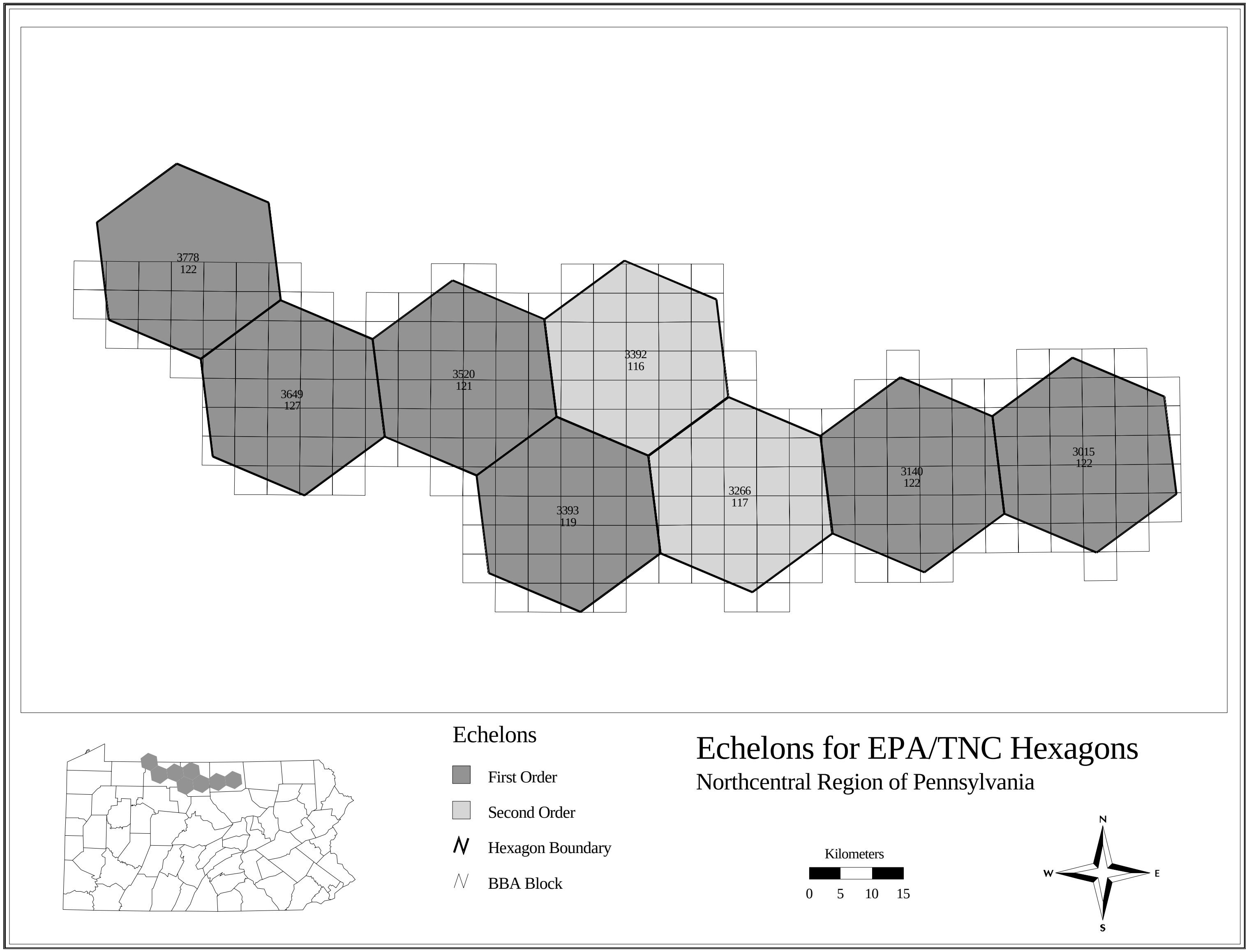


Figure 5: Northcentral region EMAP hexagons. The 4-digit number in each hexagon is the EPA-EMAP identifier, while the number below is species richness.

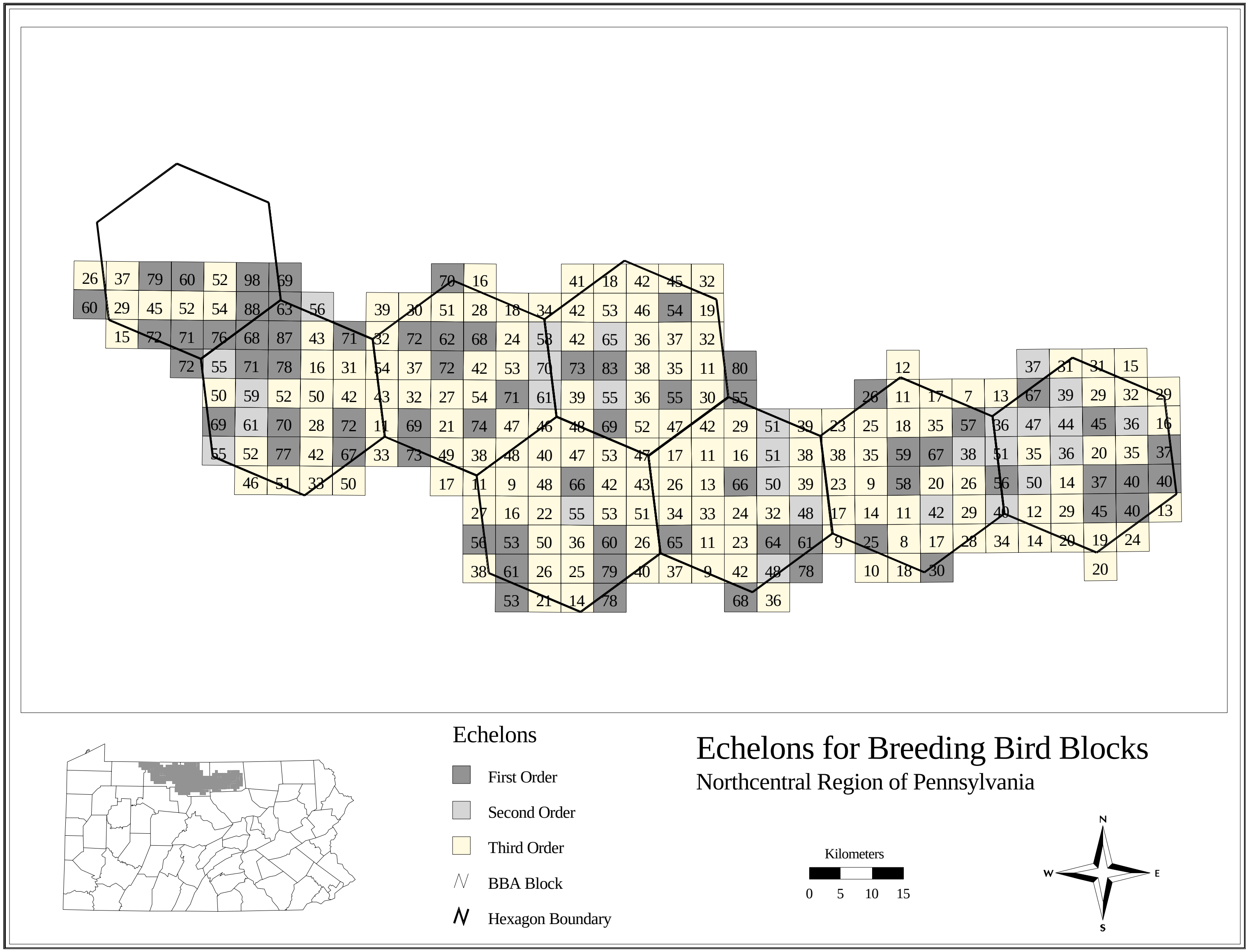


Figure 6: Northcentral region Breeding Bird Atlas blocks. The number in each block is the species richness.

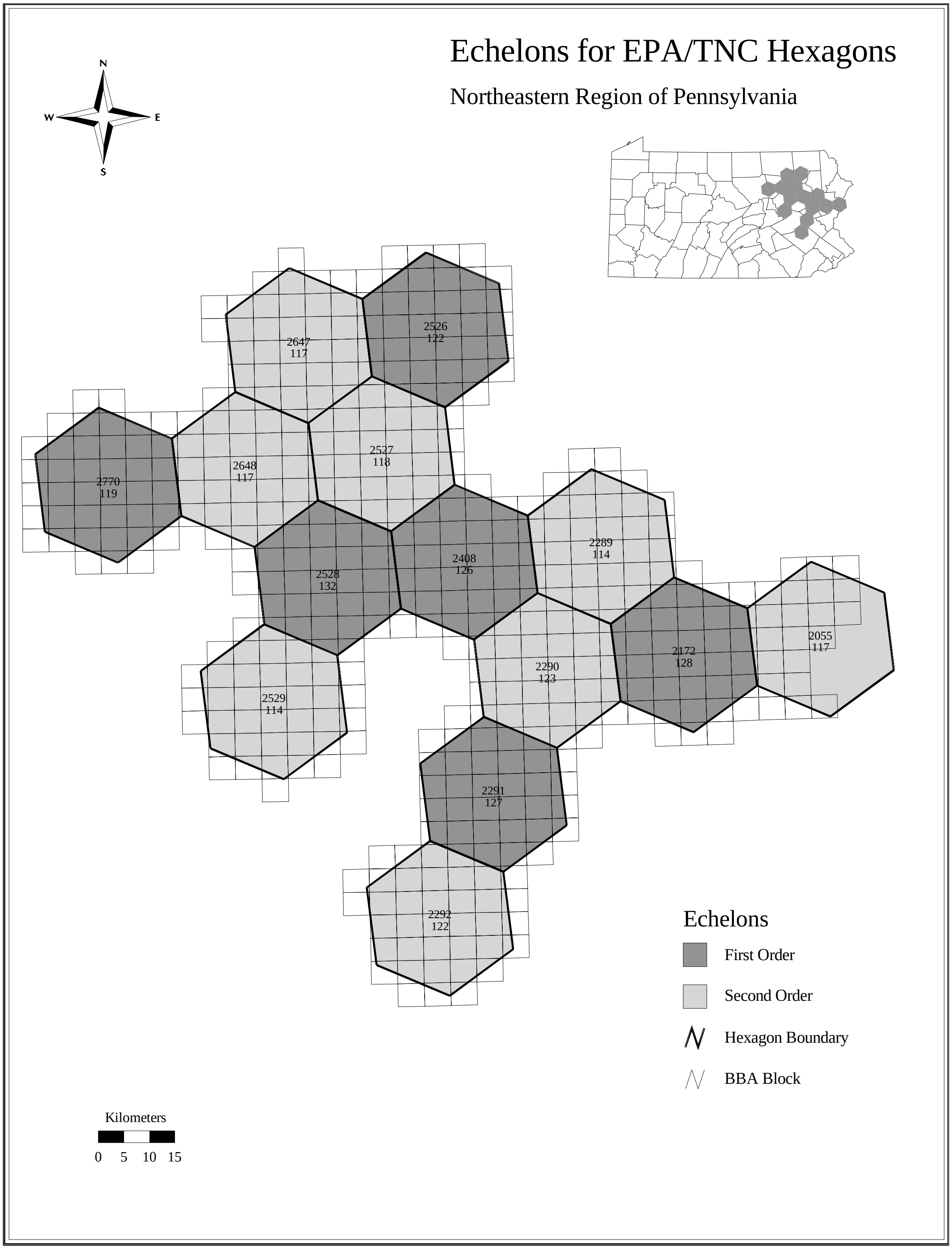


Figure 7: Northeast region EMAP hexagons. The 4-digit number in each hexagon is the EPA-EMAP identifier, while the number below is species richness.

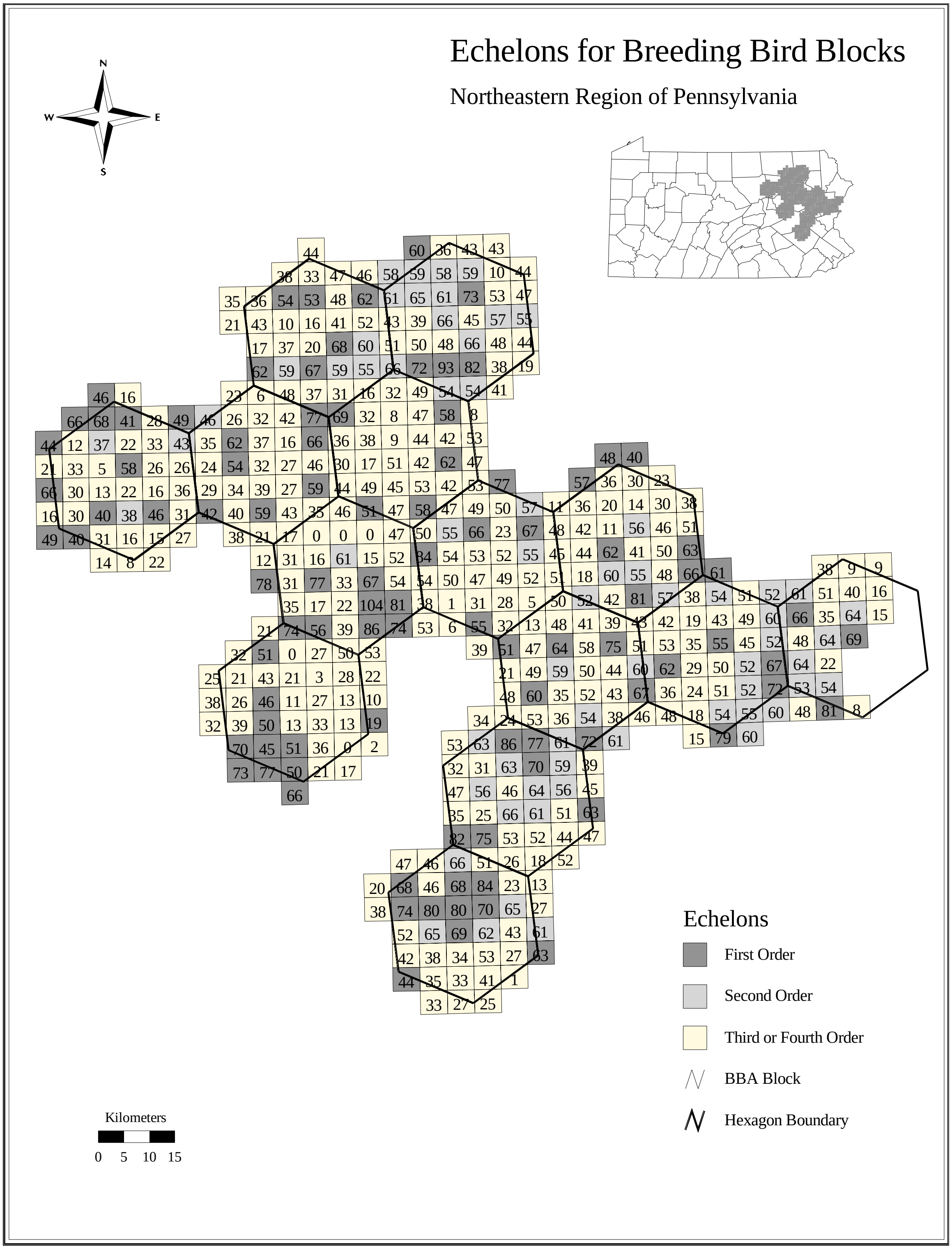


Figure 8: Northeast region Breeding Bird Atlas blocks. The number in each block is the species richness.

6 Analysis of Select Hexagons

Certain hexagons appeared particularly interesting because they fell into one of two categories. The first category contained members of first order echelons from a statewide perspective that did not appear high within their regions when mapped at the finer scale, as evidenced by containing few blocks that were of first order. We will call these "globally high/locally moderate" hexagons. The second category revealed the opposite phenomenon, where hexagons were of second order from a statewide perspective, although they appeared high within their regions when mapped at the finer scale, as evidenced by containing several blocks of first order. We will call these "globally moderate/locally high" hexagons.

Such observations were not so striking in the northcentral region where underlying land use is more homogeneous than the other two regions. In the northwest and northeast regions, hexagon numbers 4309 and 2408 were of the "globally high/locally moderate" category, while hexagon numbers 3910, 3912 and 2292 were of the "globally moderate/locally high" category. One could argue that other hexagons fall into one of these categories, but for the purposes of initial exploratory research we focus on those that seem most obvious. Properties of these particular hexagons were further investigated, as discussed next.

entropy measures

In order to assess the distribution of breeding bird atlas blocks amongst the echelon orders within a hexagon, the Shannon entropy was calculated as $-\sum_{i=1}^{k} p_i \ln(p_i)$, where p_i is the proportion of blocks in echelon order i and k is the number of orders. These results were normalized with respect to maximum entropy, $\ln(k)$, which is often termed "evenness" because it reveals the deviation from an even distribution of units amongst the k classes. An evenness of one (1) indicates maximum entropy, and as evenness decreases towards zero (0), this indicates increasing dominance of particular classes. The results are presented in Table 1, where values are also presented for the whole regions.

The "globally high/locally moderate" hexagons, numbers 4309 and 2408, revealed greater evenness of distribution amongst echelon categories, compared to their whole respective regions. Likewise, the "globally moderate/locally high" hexagons revealed lower evenness than their whole respective regions. Further analysis showed that these differences were attributed to a dominance of highest order echelons and, secondarily, first order echelons within all of the "globally moderate/locally high" hexagons; meanwhile, the "globally high/locally moderate" hexagons contained a greater number of mid-order echelons.

region/hexagon	entropy	evenness
Northwest	.903	.822
hex4309	.911	.829
hex3910	.859	.782
hex3912	.879	.800
Northcentral	.892	.812
Northeast	1.196	.863
hex2408	1.233	.890
hex2292	1.148	.828

Table 1: Spatial entropy and evenness for the three regions and select hexagons, using echelon orders as categories. The northwest and northcentral regions contained three echelon orders, while the northeast region contained 4 orders.

Apparently, when per-block species richness is fairly uniformly distributed over a hexagon, the finer scale echelon mapping may not indicate many local first order echelons, although the species add up across the hexagon to reveal a first order echelon from a statewide perspective. Such results may help provide evidence that biodiversity protection efforts may be best directed at the whole of hexagons like 4309 and 2408, while hexagons like 3910, 3912 and 2292 may contain locally species rich areas that deserve high priority for protection although they were members of second order echelons from the statewide perspective.

species area curves

In order to investigate the distributions of species, thus revealing the spatial turnover of species throughout a hexagon, we constructed species area curves in a directed manner. Starting with the block that was mostly in the middle of a hexagon, each new block was obtained by spiraling outward in a clockwise direction. The species area curve is thus constructed by adding any additional species that were encountered with each new block that was encountered. These results are presented in Figure 9, where the spiral-directed curves are compared to curves that would be expected under random sampling of blocks within a hexagon, computed as

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

where $E[S_n]$ is the expected number of species encountered in n blocks sampled at random from a total of N. A_i is the number of blocks, out of the N from which we sample, where the ith species appears and s is the species richness within the hexagon. This expression can also be found in Kobayashi (1974), Engen (1976) and Hurlbert and Archibald (1995). In the following discussion, any references to land use are based on the EPA Multiresolution Land Coverage database (cf. USEPA, 1995) which is not shown here due to loss of clarity through any greyscale renditioning and to economize space.

One would expect a species-rich hexagon that has a fairly uniform distribution of per-block species richness, to reveal a fairly smooth species area curve when blocks are accumulated in a directed manner. This is observed with hexagon 2408 which follows the smooth expected curve fairly closely. The initial "blip" is associated with the early encounter of a first order block, followed immediately by encounter of a fourth order block. Subsequent overshoots of the smooth expected curve are also associated with encountering first order blocks. The majority of hexagon 2408 that is uniformly of third and fourth order echelons is associated with continuous forest interior, spotted with small lakes. Meanwhile, the first and second order echelons in the northwest area of this hexagon are associated with both a large river and urbanization. On one hand, the river, with its associated wetlands, and the diversity of habitat offered by a small urban area surrounded by mostly interior forest may yield increased species richness. On the other hand, some observation bias may be taking place due to ready accessibility around the urban area.

Hexagon 4309 appears to start out fairly smooth, eventually plateauing, but then increases as a pocket of new species are encountered. This second increase is associated with the two blocks in the southern part of hexagon 4309 that are single-block sized first order echelons (see Figure 4). Investigation of land use in this hexagon (not shown here) shows that while habitat is fairly uniformly distributed throughout the hexagon as a patchwork of open land and continuous forest, the two localized first order echelons in the southern part are associated with open water bodies. Indeed, the local second order echelons that extend from these first order echelons are associated with the same water bodies. Thus far, we are seeing how the finer mapping scale of breeding bird atlas blocks reveals echelon structure that is more closely tied with landscape habitat features.

The three hexagons that appeared "globally moderate/locally high" seem to reveal fairly jagged species area curves from spiral-directed sampling, where local increases in slope are generally associated with encountering a block that is of a first order echelon. The uniformly third order blocks in hexagons 3910 and 3912 are associated with mostly continuous forest interior habitat, while the local first order blocks are associated with breaks in such habitat by either a large river, small scale development or transitional land. Land use within hexagon 2292 is quite distinctly segregated into continuous forest

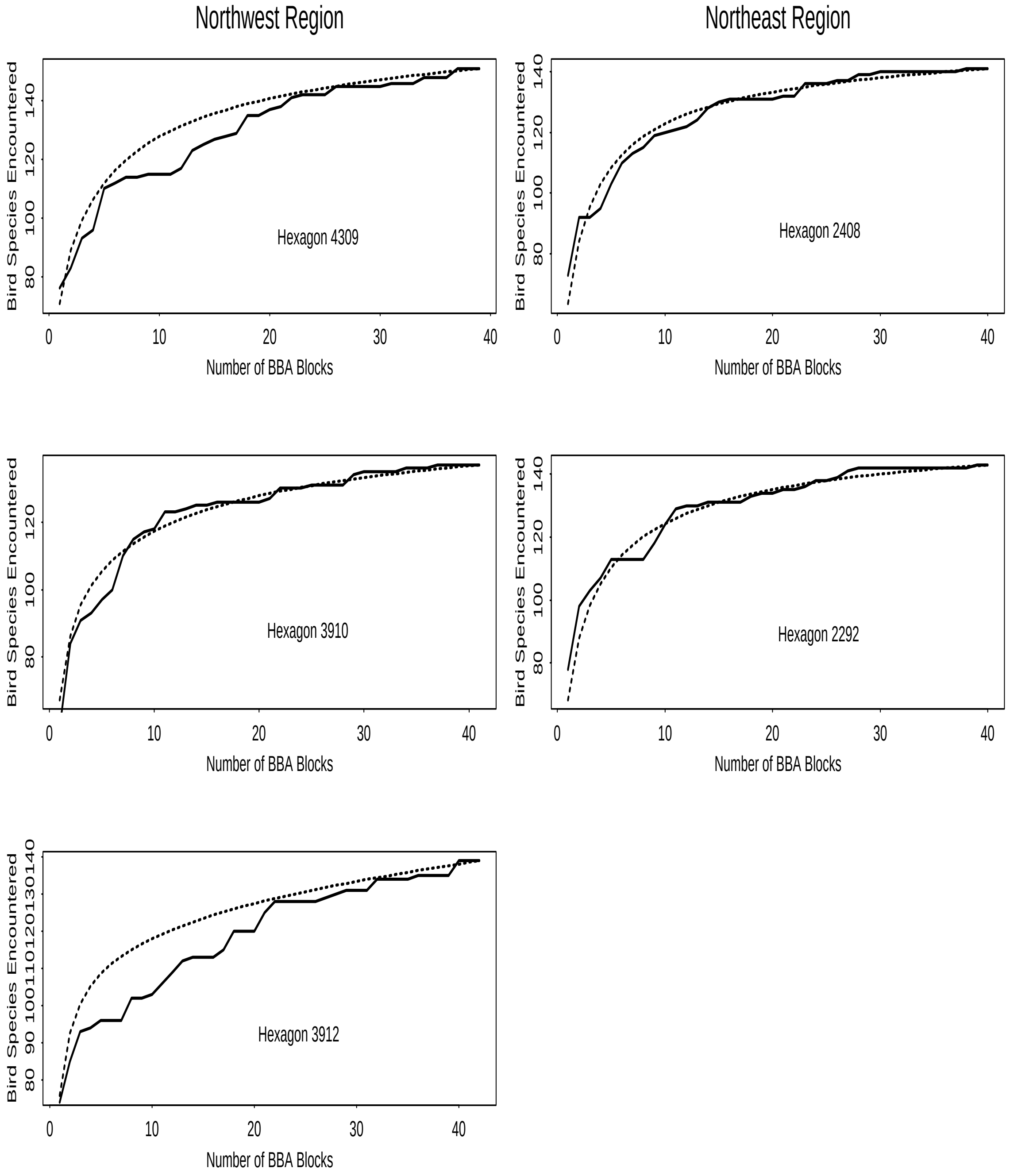


Figure 9: Species area curves for select hexagons within the northwest and northeast regions; Spiral-directed sampling (solid line) compared to expectation with random sampling (dashed line).

in the northwest, while the remainder of the hexagon is mostly agricultural land. Here we see the opposite effect that was seen in the other hexagons because the first order blocks are clustered in the area of continuous forest, while the agricultural area almost uniformly reveals blocks that are of fourth order echelons, regardless of the presence of water bodies. This observation supports the hypothesis that while some forest fragmentation may actually increase species richness from increasing habitat diversity, too much forest loss results in an overall drop in species richness (Noss, 1983).

7 Discussion

Assessing biodiversity variables such as species richness over a large geographic area requires analysis at multiple spatial scales (Turner, 1995; Turner, et al. 1991). Stoms (1994) addressed this issue by rescaling vertebrate species richness maps for two very different areas in Idaho. Although we used non-aligned grids with different shaped cells, Stoms' hierarchically nested scaling range encompassed our two scales. Visual assessment of Stoms' thematic maps indicate some agreement with our findings. Certain species-rich cells that were similar in size to an EMAP hexagon were mapped as a relatively species-poor area when the mapping units were similar in size to the Breeding bird Atlas blocks. More generally, the finer scale maps were necessary to bring out structure that was not predictable from the coarser scale map. One of Stoms' primary objectives was to see if the identification of species-rich locations was scale dependent. Since identification was based on subjective viewing of thematic maps, such an activity should benefit by the increased objectivity provided by echelon analysis.

Structuring a response surface into echelon objects opens up a greater variety of analyses. While all analyses of the original data can still be performed, properties of the echelon objects themselves can also be very informative (Myers, Patil and Joly, 1996). In our study, the combination of entropy analysis of the Breeding Bird blocks with respect to echelon order categories, along with spiral-directed species area curve construction and referencing to known land usage, together proved to be quite informative. After delineating species-rich regions from a statewide perspective, the distributions of species and per-block species richness within particular EMAP hexagons were characterized, along with rational interpretation of these observations with respect to landscape characteristics.

Acknowledments

Prepared with partial support from the United States Department of Interior, the National Biological Service (Gap Analysis Program); the United States Environmental Protection Agency (Environmental Statistics and Information Division, Office of Policy, Planning and Evaluation; and Environmental Monitoring and Assessment Program, EMAP Design and Statistics Group); and the National Science Foundation (NSF/EPA Water and Watersheds Program). The contents have not been subjected to reviews of these agencies and therefore do not necessarily reflect the views of these agencies and no official endorsement should be inferred.

We appreciate the comments from two anonymous reviewers.

References

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

Brauning, D.W. and Gill, F.B. (1983-1989) Pennsylvania Breeding Bird Atlas data. The Academy of Natural Sciences of Philadelphia, Pennsylvania Game Commission and Wild Resource Conservation Fund, Harrisburg. (as licensed for use in GAP Analysis at the Pennsylvania State University)

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

Hurlbert, S.H. and Archibald, J.D (1995) No statistical support for sudden (or gradual) extinction of dinosaurs. *Geology*, 23(10):881–884, October 1995.

Johnson, G.D., Nussbaum, B.D., Patil, G.P. and Ross, N.P. (1995) Innovative statistical mind sets and novel observational approaches to meet the challenges in the management of hazardous waste sites. *in* Lewis, R.A. and Subklew, G. (eds.), The Proceedings of a Specialty Conference: Challenges and Innovations in the Management of Hazardous Waste, VIP-52. Air and Waste Management Assoc., Pittsburgh, PA. pp. 3-32.

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

Myers, W.L., Brooks, R., Storm, G. and Bishop, J. (1996) Pennsylvania Gap Analysis Report–1995. ER9602. Environmental Resources Research Institute, Penn State University, University Park, PA.

Myers, W.L., Patil, G.P. and Joly, K. (1996) Echelon Approach to Areas of Concern in Synoptic Regional Monitoring. Technical Report 96-0601, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA. (*to appear in Environmental and Ecological Statistics*)

Myers, W.L., Patil, G.P. and Taillie, C. (1995) Comparative Paradigms for Biodiversity Assessment. *in* Boyle, T.J.B. and Boontawee, B. (eds.), Measuring and Monitoring Biodiversity in Tropical and Temperate Forests. Center for International Forestry Research, Bogor, Indonesia, pp. 67-85.

Noss, R.F. (1983) A regional landscape approach to maintain diversity. Bioscience, 33(11):700-706.

Scott, J.M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., Anderson, H., Caicco, S., D'Erchia, F., Edwards, Jr., T.C., Ullimann, J. and Wright, R.G. (1993) Gap analysis: A geographic approach to protection of biological diversity. Wildlife Monographs, 57:1-41.

Stoms, D.M. (1994) Scale dependence of species richness maps. Professional Geographer, 46(3):346-358.

Turner, S.J. (1995) Scale, observation and measurement: critical choices for biodiversity research. *in* Boyle, T.J.B. and Boontawee, B. (eds.), Measuring and Monitoring Biodiversity in Tropical and Temperate Forests, pp. 97-111. Proceedings of a IUFRO Symposium held at Chiang Mai, Thailand 1994.

Turner, S.J., O'Neill, R.V., Conley, W., Conley, M.R. and Humphries, H.C. (1991) Pattern and scale: statistics for landscape ecology. *in* Turner, M.G. and Gardner, R.H. (eds.), Quantitative Methods in Landscape Ecology. Springer-Verlag, New York. pp. 17-49.

USEPA (1995) Mid Atlantic Landscape Indicators Project Plan. EPA/620/R-95-003. 37 pp.