A Comparison of Sampling Design Options for EMAP-Agroecosystems

Steve L. Peck, John O. Rawlings and Alva L. Finkner, NCSU
Steve Peck, NCSU 1509 Varsity Dr. Raleigh NC 27606

Key Words: Agricultural Sampling, EMAP

Introduction:

The Environmental Monitoring and Assessment Program (EMAP) was conceived as a framework to monitor the status and trends of the nation's ecological resources. It is designed to track environmental change over large geographical areas and over long periods of time (Norton & Slonecker, 1990). It differs from many other environmental monitoring programs in that it will be a probability based sample allowing regional estimates of status and change.

To facilitate interagency cooperation and the use of scientific expertise EMAP has been organized into several broad resource groups; each of which is responsible for a particular ecological resource (Heck, et.al. 1991). These resource groups are: agroecosystems, arid lands, estuaries, forests, the Great Lakes, inland surface waters and wetlands. This paper is concerned with the Agroecosystem component of EMAP and the sampling design issues encountered in sampling agricultural systems for the purpose of monitoring extent and changes.

Agroecosystem Research Group

The science of agroecology has its roots in the realization that cropping systems are driven by ecological processes. (Altieri, 1987). Recently there has been much work in developing a conceptual and theoretical framework from which to view agriculture systems ecologically. (See for example, Carroll, et.al. 1990). Agroecosystems are more complex than natural ecosystems because humans manipulate normal ecosystem structure and function (Gliessman, 1990).

In addition to being driven by a mixture of management and ecosystem processes, agroecosystems have a major influence on other ecosystems. Because they make up approximately 30% of the land area in the United States (Coleman and Hendrix, 1988) and because of the influence of human inputs, e.g. pesticides and fertilizers, their effect on other systems such as forests, wetlands, surface waters, and ground waters may be profound.

For EMAP, agroecosystems are defined as land used for crops, pasture and livestock; the adjacent uncultivated land that supports other vegetation (hedgerows, woodlots, etc) and wildlife; and the associated atmosphere, underlying soils, ground water, and drainage networks (Heck, et. al.,1991). It is the task of the Agroecosystem Research Group (ARG) to identify, measure and quantify important environmental indicators of the status and trends of the nation's agroecosystems from an ecological perspective. The ARG is also charged with the task of implementing the sampling program for Agro-EMAP.

The EMAP statistical design team has developed an overall sampling strategy for the EMAP program which will be described briefly in the next section. Consistent with EMAP directives, the Agro-EMAP design group has tried to capitalize on the extensive agricultural sampling that is already being done by various government agencies such as the National Agricultural Statistics Service (NASS).

Two sampling plans have been proposed by the Agroecosystem Resource Group for sampling the nation's agriculture systems. Both involve NASS, one also uses the sampling design proposed by the EMAP statistical design team.

EMAP Sampling Design

The overall EMAP design calls for an area sample of the nation's ecological resources via a triangular point grid with approximately 27 km between points. A 40 km^2 hexagon surrounding the point will be sampled for ecological indicators. Randomization of the grid will be achieved by a random shift in the plane of the entire grid of points. There is no stratification at this point.

It is anticipated that each resource group will sample approximately 3200 hexagons nationwide. More information on indicator variability, cost and the desired precision may change this estimate. Where the need to sample rare or localized resources exists, such as the Redwood Forests in California, the grid can be enhanced to ensure adequate coverage.

National Agriculture Statistics Service Area Frame

The NASS area frame is constructed state by state and provides complete coverage of the conterminous United States. The efficiency of each State's frame is reviewed annually and is renewed as needed or at least every 10 to 15 years.

NASS uses stratification, based primarily on the proportion of area devoted to agriculture, to improve the precision of its survey estimates. The land is divided into 6 to 8 land use categories. For example in North Carolina the strata were:

- -Greater than 50% cultivated,
- -15%-50% cultivated ,
- -Agro-Urban: Greater than 20 homes per square mile,
- -Commercial: Greater than 20 homes per square mile,
- -Resort: Greater than 20 homes per square mile,
- -Less than 15% cultivated, and
- -Non-agricultural (NASS, 1990).

The strata are required to have well defined boundaries and remote sensing is used to delineate the stratum boundaries. All strata (including non-agricultural) are sampled.

The strata are divided into primary sampling units (PSUs) designed to be from 6-8 square miles in highly agricultural strata and from .5 to 2 square miles in less intensive agriculture strata. In all cases PSUs are designed to contain 6-8 sample segments which, like the PSUs and strata, have fixed permanent boundaries (Cotter & Nealon, 1987).

Agroecosystem Sampling Plan

The Agroecosystem Resource Group has identified two options for ecosystem sampling of indicator data. Both use NASS operationally as well as the NASS area frame. The first, called the Hexagon Plan, uses the centroids of the sample EMAP hexagons to identify the NASS PSU and segment for inclusion in the Agroecosystem sample. Once a segment is selected it will remain in the program *in perpetuum*.

In the Hexagon Plan the sample (approximately 3,200 sampling units) would be divided to provide four independent interpenetrating replicates (of approximately 800 each). One set of 800 would be enumerated each year. The sample will not be stratified. As a result, only a subset of the 800 sample segments will contain agricultural lands and only on those can indicator data be taken.

The cost of the Hexagon Plan can be broken into two parts. The first is the cost of sample location identification. This includes the cost of locating the centroids, identifying the PSUs and delineating the segments on aerial photos and highway maps. The second is the cost of data collection.

The second plan, called the Rotational Panel Plan, uses approximately 20% of the area frame sample used by NASS for their June Enumerative Survey (JES). This sample is chosen without regard to the EMAP hexagons. Twenty Percent of the segments would then, following the NASS design, be replaced with a new sample each year; after five years the sample will have been completely replaced.

NASS stratification ensures that the sampled segments are concentrated in areas that are agricultural. For this plan, the sample would be a subset of the NASS sample, and therefore would involve only the incremental costs associated with the additional data gathered from current NASS surveys and additional indicator data taken when NASS is not normally in the field.

The key difference between the two plans is that the Hexagon Plan, a longitudinal design, keeps the sample segment for the duration of the program. In the Rotational Panel Plan the individual samples are replaced with a new sample segment after five years.

Advantages and Disadvantages of the Plans.

Many of the statistical advantages and disadvantages of the two plans are being investigated in a Ph.D dissertation at the University of North Carolina at Chapel Hill by Virginia Lesser. (Lesser, Work in Progress).

Other advantages and disadvantages of these plans may be characterized as follows:

* In the Hexagon Plan if resources of two or more resource groups are collocated it may be possible to observe some measures of association between data collected on variables from different resource groups. The fact that each group will attempt to maximize the coverage of variables which are of most interest may militate against having a meaningful sample size in joint distributions.

* Apart from ratio estimates, when they are used, estimates of variance from the Rotational Panel Plan are unbiased. Variance estimates from the Hexagon Plan must be approximated, since a systematic sample has a large number of zero pairwise probabilities.

* The cost of the Rotational Panel Plan will be much less than the Hexagon Plan because the former will be a subsample of the JES and will be conducted in close cooperation with that survey. On the other hand, the cost of constructing the sample for the Hexagon Plan will be a one-time cost since the sample units are never replaced. This cost, however, will be sizable since the complete cost of the enumeration would have to be borne for each survey of the Hexagon segments.

* Stratification, in the Rotational Panel Plan, coupled with near optimum allocation, increases the precision of NASS estimates. Neither of these advantages are available to the Hexagon Plan.

* Estimates of land use and changes in land use can be made annually using the JES in conjunction with the Rotational Panel Plan.

* The Rotational Panel Plan may lose precision vis-a-vis the Hexagon Plan (on a per unit basis) in the estimation of time trends because each sample unit is measured only twice in a four-year EMAP cycle. However, it is conceivable that this disadvantage would be offset by a larger sample size for a given cost.

* Both plans are subject to response bias but in different ways. Conditioning, because contact with the farmer may influence management practices, is potentially present in both plans. Under the Rotational Panel Plan, the farmer will be contacted every year for five years and then be rotated out of the sample. In the Hexagon sample, the farmer (or his successor) will be contacted on a four-year cycle but would continue to be contacted for the duration of the program.

In 1992 a pilot project is planned that will compare aspects of these two sampling plans.

Other Concerns That May Influence Choice of Plans

In addition to the advantages and disadvantages listed above there are several questions that may influence the decision of which of the plans to implement. These questions are more difficult to answer.

The objectives of EMAP are somewhat different from those of NASS. EMAP in interested in monitoring status and trends in the health of the nation's agricultural systems- as ecosystems through time. NASS is primarily interested in estimating agricultural production and in describing management practices at a given point in time. Does the stratification that NASS uses for its sample provide efficiency in measuring ecosystem health temporally? Is it more efficient than the EMAP design which uses no stratification?

Initial investigations have suggested that although the purpose of the sampling is different, inferences are desired for the same population group. Since the NASS stratification stratifies on intensity of agriculture, and agroecosystems are defined primarily on the basis of agricultural land, it is believed that the NASS stratification will increase the EMAP precision by concentrating points in the areas in which we are most interested.

The ARG is also interested in lands surrounding agroecosystems such as woodlots and shelter belts. These are not taken into account in the NASS stratification. These entities, however, are in close association with agricultural systems and it is believed that the stratification will also be effective for samples from these populations.

Several areas of research continue in the exploration of which of the two design plans to employ for the Agroecosystem component of EMAP. Currently, methods are being explored for creating surface maps of spatially sampled data in order to integrate indicator information across ecosystems. Several methods using techniques from spatial statistics look hopeful. These will also be useful in incorporating other sources of data, such as climate data, that are unlikely to be available at the specific sampling points in either design. We are also exploring issues relating to field sampling, such as how best to choose the number of samples per NASS sample for each of our indicators. In addition to the formal statistical issues there are also numerous practical and logistic constraints that need to be identified and resolved.

Another area of exploration is the investigation of adding some repeated sampling to the Rotational Panel Design using a multistage Keyfitz procedure (See Keyfitz, 1951, Drummond, 1980). This would allow the verification of assumptions made on the correlation structure, over time, of the NASS based design.

In conclusion, the issues involved with the initiation and choice of a design plan to survey the status and trends of the nation's agroecosystems are complex and multifaceted. As the issues relating to these designs, such as cost and logistic feasibility, are clarified through continued exploration and the 1992 pilot it is hoped that an ecologically sound and fiscally prudent decision can be implemented.

Acknowledgements

We would like to acknowledge all of the members of the Agroecosystem Resource Group for their help on this project; especially George Hess for his helpful insights and suggestions on earlier drafts of this paper. We would also like to acknowledge Tim O'Brien for reviewing the final draft and making it a more clear and readable paper.

Bibliography

Altieri, M. 1987. *Agroecology: The Scientific Basis of Alternative Agriculture.* Westview Press. Boulder, CO.

Carroll, R. C., J. H. Vandermeer, P. Rosset. 1990. *Agroecology.* McGraw-Hill, New York, NY.

Coleman, D. C. and Hendrix P. F. 1988. "Agroecosystem processes." *In Concepts of Ecosystem Ecology: A Comparative View.* ed. Pomeroy, L.R. and J.J. Alberts. Springer-Verlag, New York, pp. 149-170.

Cotter, J. and J. Nealon. 1987. *Area Frame Design For Agricultural Surveys.* Publication of the National Agricultural Statistics Service. U.S. Department of Agriculture. Washington, DC.

Drummond, D. J. 1980. *Multistage Keyfitz Updating.* Statistics Research: Research Triangle Institute. Reprint Series No. S-157. Reprinted from the 1980 Section on Survey Research Methods Proceedings of the American Statistical Association.

Gliessman, S. R. (Ed.). 1990.

Agroecology: Researching the Ecological Basis for Sustainable Agriculture. Ecological Studies, Vol. 78. Springer-Verlag. New York.

Heck, W. W., C. L. Campbell, R. P. Breckenridge, G. E. Byers, A. L. Finkner, G. R. Hess, J. R. Meyer, T. J. Moser, S. L. Peck, J. O. Rawlings. 1991. *Environmental Monitoring and Assessment Program (EMAP)-Agroecosystem Monitoring and Research Strategy.* EPA/600/4-91/013. U.S. Environmental Protection Agency, Washington, D.C.

Keyfitz, N. 1951. *Sampling With Probabilities Proportional to Size: Adjustment of Changes in the Probabilities.* JASA. 46 pp. 105-109.

Lesser, V. M. Work in Progress. *A comparison of Periodic Survey Designs Employing Multi-State Sampling.* Doctoral Dissertation for Department of Biostatistics, University of North Carolina at Chapel Hill.

Norton, D. J. and E. T. Slonecker. 1990. "The Ecological Geography of EMAP". *Geo Info Systems* 1(0). pp. 32-24.

NASS Area Frame Section. 1990. *Area Frame Design Information.* National Agriculture Statistical Service Publication.

U.S. EPA Office of Research and Development. 1991. *An Overview of the Environmental Monitoring and Assessment Program. In* EMAP Monitor. EPA-600/M-90/022. Washington, D.C.