EPA/620/R-94/014
May 1994

Environmental Monitoring and Assessment Program

Agroecosystem Pilot Field Program Report - 1992

by

C. Lee Campbell, Technical Director
Jeff M. Bay
Anne S. Hellkamp
George R. Hess
Michael J. Munster
Karen E. Nauman
Deborah A. Neher
Gail L. Olson
Steven L. Peck
Brian A. Schumacher
Kurex Sidik
Mark B. Tooley
David W. Turner

This study was conducted in cooperation with
U.S. Department of Agriculture
Agricultural Research Service
Raleigh, NC 27711

U.S. Environmental Protection Agency
Office of Research and Development
Washington, D.C. 20460

Environmental Monitoring Systems Laboratory
Las Vegas, NV 89193

Printed on Recycled Paper

Notice

This research has been funded by the United States Environmental Protection Agency (EPA) through its Office of Research and Development (ORD) under Interagency Agreements #DW12934170 with the U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS) and #DW12934747 with the USDA National Agricultural Statistics Service (USDA NASS) and by the USDA ARS. It was conducted with our research partners under the management of the Environmental Monitoring Systems Laboratory - Las Vegas in support of the Environmental Monitoring and Assessment Program (EMAP). Neither U.S. EPA nor USDA ARS endorses or recommends any trade name or commercial product mentioned in this document to the exclusion of others. They are mentioned solely for the purpose of description or clarification.

Proper citation of this document is:

Campbell, C. L., J. M. Bay, A. S. Hellkamp, G. R. Hess, M. J. Munster, K. E. Nauman, D. A. Neher, G. L. Olson, S. L. Peck, B. A. Schumacher, K. Sidik, M. B. Tooley, and D. M. Turner. 1994. Environmental Monitoring and Assessment Program - Agroecosystem Pilot Field Program Report - 1992. EPA/620/R-94/014. U.S. Environmental Protection Agency, Washington, D.C.

Table of Contents

List of Figures

List of Tables

Glossary of Acronyms

AHHC	annually harvested herbaceous crop
AIC	Agroecosytem Information Center
ARG	Agroecosystem Resource Group
ARS	Agricultural Research Service
cdf	cumulative distribution function
cv	coefficient of variation
EMAP	Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency
ERL	Environmental Research Laboratory
GIS	geographic information systems
IM	information management
JES	June Enumerative Survey
LAN	local area network
MCL	maximum contaminant level
MI	maturity index (for free-living nematodes)
MOU	Memorandum of Understanding
NASDA	National Association of State Departments of Agriculture
NASS	National Agricultural Statistic Service
NCSU	North Carolina State University
OM	organic matter
PPI	maturity index (for Plant-Parasitic Nematodes)
PSU	Primary Sampling Unit
QAO	quality assurance officer
QA/QC	quality assurance/ quality control
SHAN	Shannon index of trophic diversity
SRPG	soil rating for plant growth
USDA	United States Department of Agriculture

Acknowledgements

The Agroecosystem Resource Group (ARG) would like to thank all those who contributed to the planning, execution, and analysis of the 1992 Pilot in North Carolina and all those who shared their expertise with us.

The involvement of the USDA National Agricultural Statistics Service (NASS) has become integral to the ARG. We especially thank our past and present administrative contacts: Robert Bass, Ray Halley, Steve Manheimer, and Jim Gibson in Washington, and Craig Hayes in Raleigh. Special thanks also to those who helped us "sweat the details" of data collection and analysis: Sarah Hoffman, Becky Cross, Tom Sabel, and David Luckenbach. Of course, the data would never have been collected without the conscientious efforts of a group of 26 enumerators with the National Association of State Departments of Agriculture, working through the North Carolina NASS office.

The statistical expertise, and guidance of Len Stefanski, Department of Statistics, North Carolina State University, kept us on track during the preparation of this report. The programming and assistance of Cathy Barrows, Department of Statistics, allowed us to produce our cumulative distribution functions and is gratefully acknowledged.

Our increasing cooperation with the USDA Soil Conservation Service (SCS) is also acknowledged, with thanks to our SCS liasons, Bob Smith and Bill Roth.

Thanks also to various U.S. Environmental Protection Agency (EPA) personnel including Ann Pitchford and Sue Franson at EMSL-Las Vegas for their excellent management support, and Charlie Smith, Bill Payne, and Andy Paeng at the Athens-ERL, who assisted in the development of the water sampling protocols and analyzed the water samples.

Gratitude is also expressed to those who have helped in the production of this project report: Tracey Diggs for getting the drafts into shape, Carla Tutor for photocopying and mailing, Phyllis Garris in the library, and Charles Harper for general technical support. Also, our thanks to Clara Edwards for all the things she does to make our program possible.

Deep admiration and gratitude are due to four emeritus members of the Agroecosystem Resource Group. Dr. Walt Heck (USDA-ARS) was Technical Director of the ARG during the planning and field phases of this pilot field program. His capable and visionary leadership continues in his role as Associate Director for the EMAP terrestrial groups. Dr. John Rawlings, Department of Statistics, North Carolina State University, and Dr. Alva Finker, Emeritus Statistician, helped lay a foundation for the Agroecosystem program in both statistics and common sense. Finally, Dr. Julie Meyer, now of the University of Wisconsin, was a pioneering member of the ARG, who brought a special knowledge and goodwill to the initiation of soil quality indicators, development of the survey questionnaire, and birth of the program.

1. INTRODUCTION

1.1 MISSION AND GOALS OF THE EMAP-AGROECOSYSTEMS PROGRAM

The mission of the Agroecosystem Resource Group (ARG) of the Environmental Monitoring and Assessment Program (EMAP) is to develop and implement a program that will, in the long term, monitor and assess the condition and extent of the nation's agroecological resources from an ecological perspective through an interagency process. The specific objectives of the agroecosystem program are to:

- Estimate the status, trends, and changes in selected indicators of the condition of the nation's agroecological resources on a regional basis with known confidence.
- Estimate the geographic coverage and extent of the nation's agroecological resources with known confidence.
- Seek associations between selected indicators of natural and anthropogenic stresses and indicators of the condition of agroecological resources.
- Provide annual statistical summaries and periodic assessments of the nation's agroecological resources.

1.2 CONCEPTUAL APPROACH

The EMAP-Agroecosystems monitoring effort is based on assessment questions related to three societal values, which in turn are related to sustainability. Biotic and abiotic condition indicators are being developed to address each assessment question.

1.2.1 Sustainability

The sustainability of agroecosystems is of primary importance to the people of the United States and the world. Although there are several aspects of sustainability, the ARG is interested in the ecological sustainability of agroecosystems.

> An agroecosystem is ***ecologically sustainable*** if it maintains or enhances its own long-term productivity and biodiversity, the biodiversity of surrounding ecosystems, and the quality of air, water, and soil.

Two other facets of sustainability, economic and social, are addressed by other federal agencies such as the U.S. Department of Agriculture (USDA) Economics Research Service and the USDA National Agricultural Statistics Service (NASS).

> A farm is *economically sustainable* if it is economically viable over the long term.

An agricultural system is *socially sustainable* if it meets the basic food and fiber needs of society and maintains or enhances the quality of life for farmers and rural communities.

1.2.2 Societal values

The three societal values for agroecosystems are the components of ecological sustainability: ***quality of air, water, and soil; productivity; and biodiversity.***

Agroecosystem performance depends on the *quality of the air, water, and soil* entering and within the agroecosystem. In addition, agricultural practices can affect the air, water, and soil of the agroecosystem and surrounding ecosystems.

In the traditional ecological sense, *productivity* is the rate at which a given trophic level captures energy; net primary productivity is the net accumulation of plant biomass per unit area per unit time. Society has a compelling interest in the biomass that is harvested for food and fiber. However, in a monitoring context, other aspects of productivity are of interest, for instance, biomass production in pastures and in ecotonal areas such as windbreaks, and the efficiency with which resources are used in the agroecosystem.

Biodiversity in a field and in the surrounding landscape affects agroecosystem function and is affected by agricultural practices. Abundance and diversity of some species, including pollinators and insect predators, can positively affect plant productivity; diversity of others, such as parasites, can have negative effects on both plants and animals. Genetic diversity is important as the raw material for adaptation and in the prevention of devastating epidemics.

1.2.3 Assessment Questions and Indicators

Assessment questions are general, enduring, biologically oriented questions that drive the program. Assessment questions are answered relative to some specified value with information from indicators. Indicators are measures that reflect the condition of an ecological resource or its exposure to stress. Condition is judged as acceptable (nominal), marginal or unacceptable (submnominal) relative to standards established for the specified value.

Although assessment questions are not actually answered in the 1992 Pilot, they are the driving force behind it. Examples of the kinds of questions that EMAP-Agroecosystems will answer in the future are in Table 1.1. Indicators currently under consideration by the ARG are listed below; those in italics were explored in the Pilot and are discussed in Chapters 2 and 3.

Crop Productivity	*Soil Quality: physical/chemical*
Soil Biotic Diversity (nematode indices)	*Water Quality: ponds and existing wells*
Land Use and Cover	*Pest Management (including agrichemical use)*
Landscape Structure	Biological Ozone Indicator (clones of white clover)
Insect Diversity	Habitat Quality for Wildlife
Symptoms of Foliar Injury	Genetic Diversity
Livestock Productivity	

Table 1.1 Societal values and related assessment questions.

Societal Value	Assessment Questions
Quality of Air, Water, and Soil	○What proportion of agroecosystems has soil quality sufficient to sustain both crop and non-crop productivity? ○What proportion of agroecosystems has acceptable diversity of soil microbes and invertebrates?
Productivity	○What proportion of agroecosystems is attaining their productive capacity (as determined by soil map unit, climate, historic levels of production, etc.)?
Biodiversity	○What proportion of agroecosystems has acceptable diversity of insects? ○What percent of agroecosystems has associated noncrop areas with habitat suitable for wildlife species of interest? ○In what proportion of agroecosystems is diversity of wildlife declining?

1.2.4 Conceptual Model

The ARG has developed a conceptual model of agroecosystems to assist in formulating assessment questions, identifying appropriate measures for each indicator, and identifying possible relationships among measurements in the development of indices. It is also a valuable aid in communicating our ideas to a wide audience.

An **agroecosystem** (Fig. 1.1) is a dynamic association of crops, pasture, livestock, other plants and animals, atmosphere, soils, and water. The agroecosystem includes not only the field, but also the associated border areas such as windbreaks, fence rows, ditchbanks, and farm ponds. The agroecosystem boundary depends on, and varies with, the process being considered.

Agroecosystems interact with larger **landscapes** (Fig. 1.2), which include uncultivated land, drainage networks, human communities, and wildlife. The landscape is the area that directly affects the ecology of the agroecosystem and is

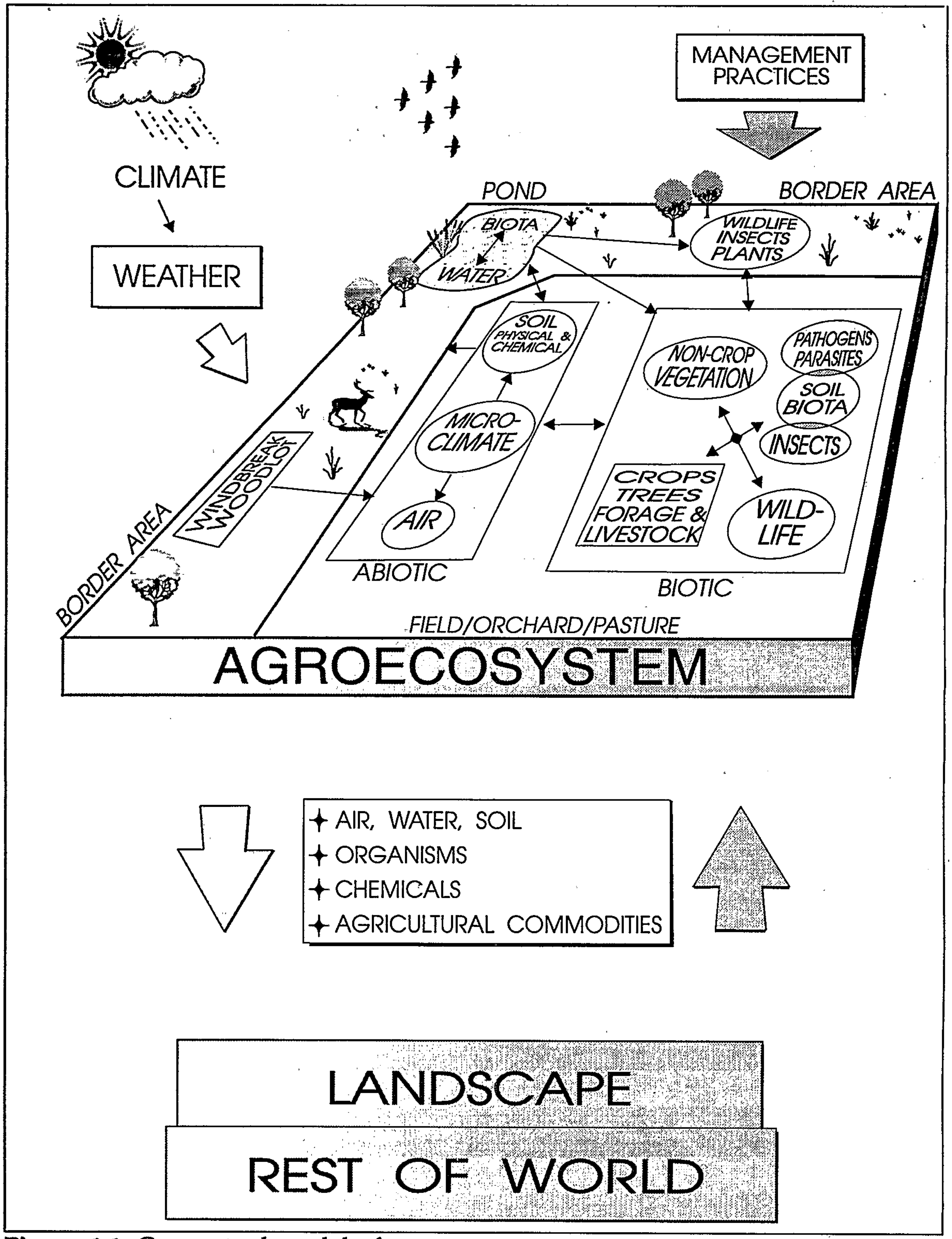


Figure 1.1. Conceptual model of an agroecosystem.

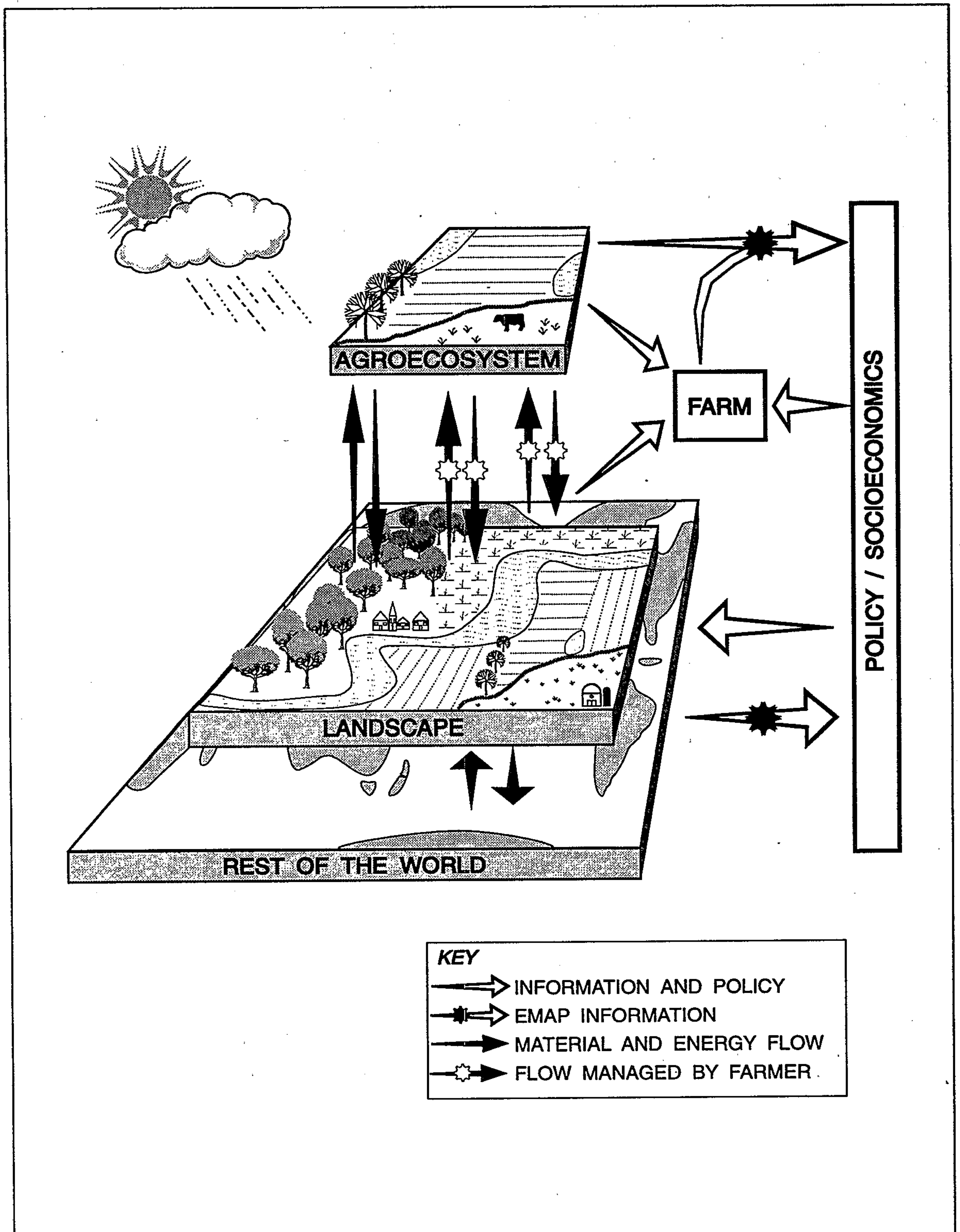


Figure 1.2. An agroecosystem in its surroundings.

directly affected by agroecosystem processes. The landscape boundary depends on, and varies with, the process being considered.

Materials and energy are brought to the agroecosystem from afar and exported to distant points (**rest of the world**, Fig. 1.2).

Agroecosystems also belong to **farms** (Fig. 1.2). A farm is an economic entity in which an operator uses information about market conditions and available resources to manage agroecosystems for economic gain.

Government policy, programs, and regulations and socioeconomics influence decisions made at all levels. In particular, policies are intended to change people's actions to realize the interests of society.

1.3 PURPOSE AND OBJECTIVES

The 1992 Pilot was designed to evaluate aspects of the EMAP Agroecosystem monitoring program critical to the implementation of a regional and national program. It was designed to address these program aspects over an area large enough to provide reliable answers to questions concerning the operation of the monitoring program but small enough to be physically and fiscally manageable. There were three major objectives for the 1992 Pilot:

1. Compare the relative efficiency, in terms of cost and precision, of two sampling frames.

2. Evaluate an initial suite of indicators to:
   - Assess the ability of each indicator to address the assessment questions of interest;
   - Establish an initial range of values for each indicator across the diverse physiographic regions in the state;
   - Assess spatial variability of indicator values within and among sample units;
   - Identify the usefulness and sensitivity of each indicator in determining ecological condition; and
   - Determine cost-effectiveness for each indicator.

3. Develop and refine plans for:
   - Sampling;
   - Logistics;
   - Quality Assurance;
   - Data analysis, summarization, and reporting;
   - Information management; and
   - Ecological health indices and their interpretation.

The 1992 Pilot was also conducted to establish a cooperative, working relationship with USDA-NASS at both the state and national level. NASS has a well established, nationwide network of enumerators and administrators experienced in conducting national surveys. Also, growers throughout the U.S. are familiar with NASS personnel and have confidence in NASS because of NASS's data confidentiality requirements.

North Carolina was selected for the 1992 Pilot because:

1. The physiographic diversity of the state is representative of the ecoregions of the Southeastern region of the United States (Fig.1.3).

2. NASS is organized by state. Limiting the Pilot to one state simplified problem resolution.

3 Most of the ARG staff is located in Raleigh, NC, which facilitated logistic activities.

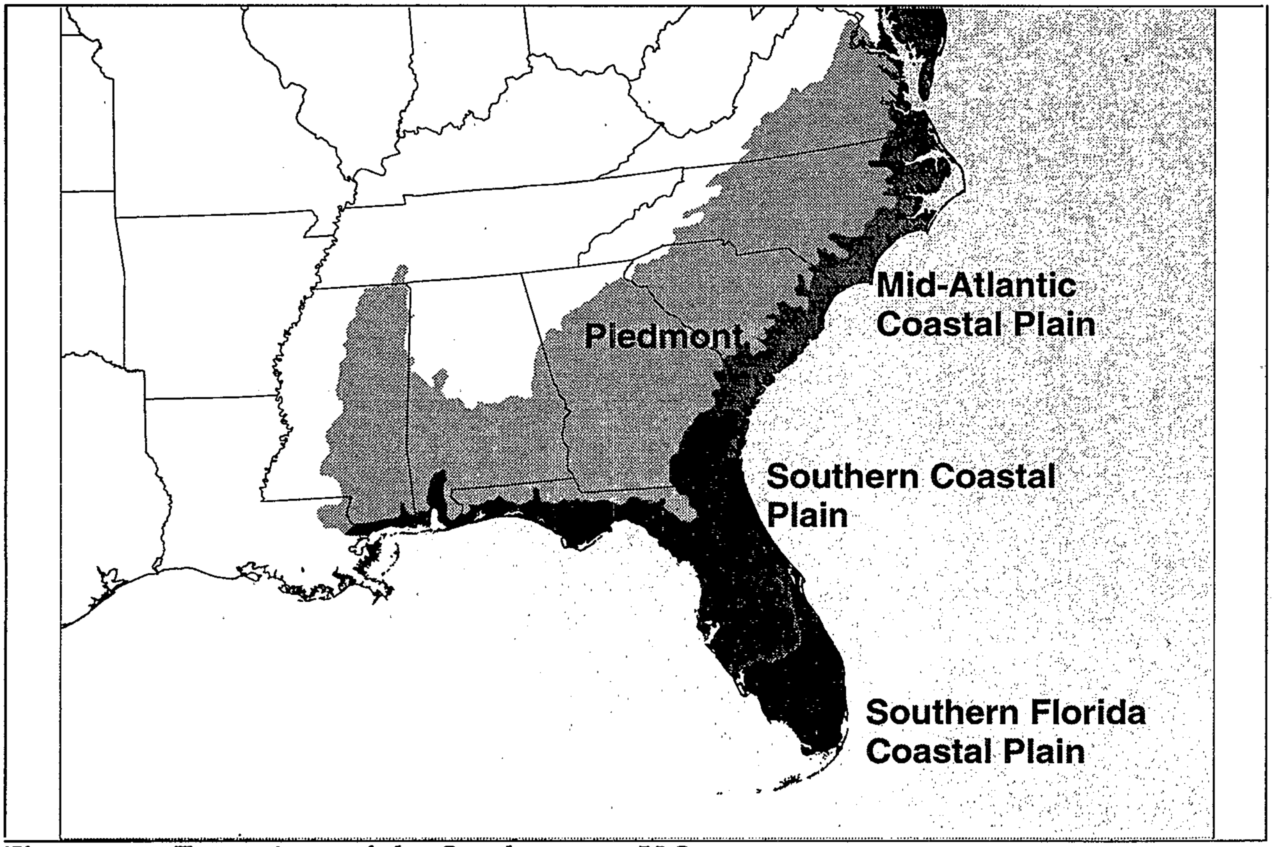


Figure 1.3 Ecoregions of the Southeastern U.S.

1.4 PLANNING AND IMPLEMENTATION

A brief summary of activities and methods associated with the planning and implementation of the 1992 Pilot are presented here. The Agroecosystem 1992 Pilot Project Plan (Heck et al., 1992) should be consulted for a full description of activities and methods.

1.4.1 Design and Statistical Considerations

The ARG considered two sampling plans in the 1992 Pilot: the NASS Rotational Panel Plan and the EMAP Hexagon Plan. Each used the NASS Area Frame segments as the basic sampling units. NASS Area Frame segments were defined by stratifying the state of North Carolina based on intensity of agriculture, dividing each stratum into Primary Sampling Units (PSUs) and then dividing a random sample of PSUs into six to eight sample segments. Segment size depends on strata, but is approximately 1 square mile (2.6 km^2) for agricultural strata.

The 203 hexagons (40 km^2 each) with their centroids in North Carolina were divided into four subsamples (Overton et al., 1991). Of the 54 hexagons in the one randomly selected subsample, three were over water. Thus, 51 hexagons distributed over 49 counties were used (Fig. 1.4).

1.4.2 Hexagon Sampling Plan

NASS identified the PSU that contained the centroid of each selected hexagon and divided the PSU into segments according to standard NASS criteria. The segment containing the centroid was identified and included as a sample segment. Area, cultivated acreage, and number of fields were estimated within each segment.

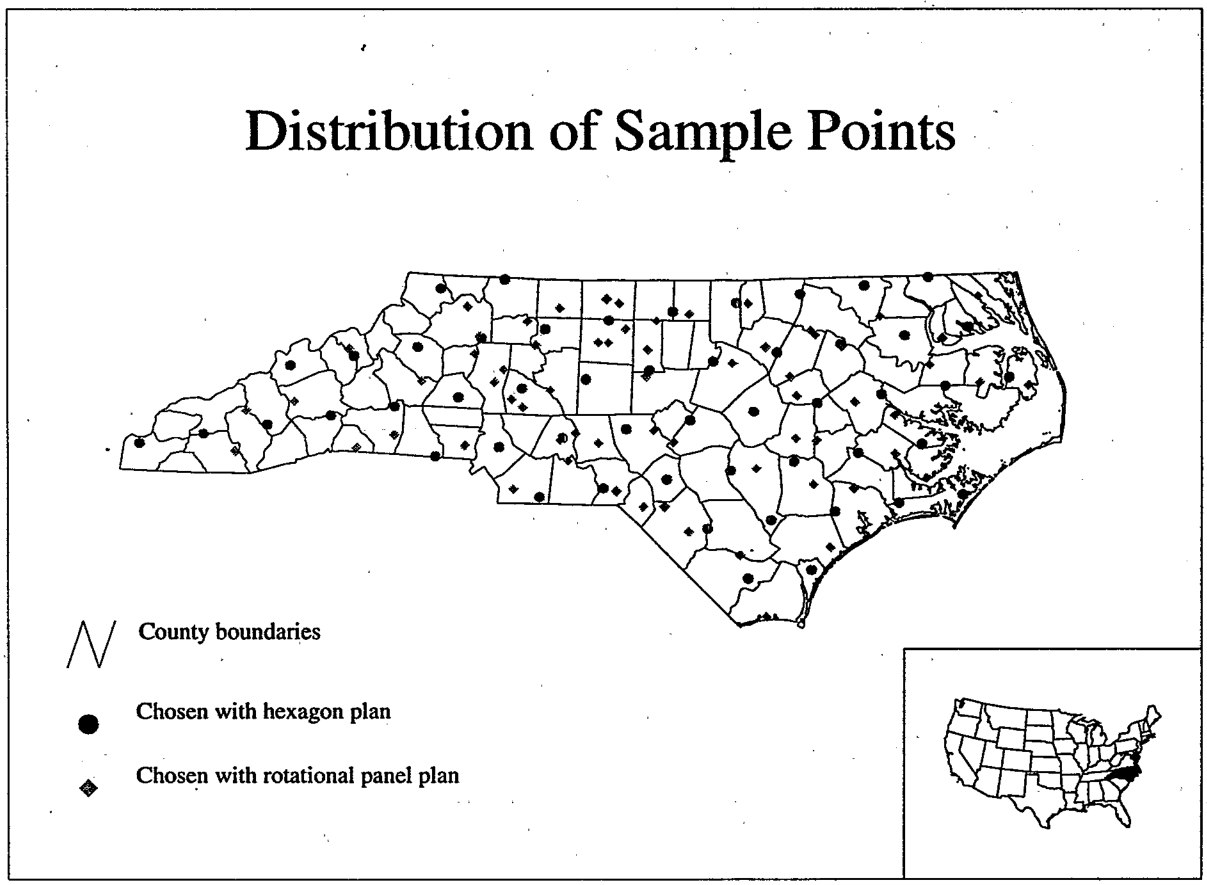


Figure 1.4 Approximate sample locations for the 1992 Pilot. Rotational panel sampling points are stratified by intensity of agricultural land use (Table 1.2). Hexagon sampling points occur on a regular grid with random initial placement of the grid.

1.4.3 Rotational Panel Sampling Plan

Table 1.2 Number of sample segments in each of the NASS strata.

Strata	Sample Points
>50% cultivated	6
15-50% cultivated	28
<15% cultivated	16
Ag-Urban (>20 homes/mi^2)	10
Commercial	3
Resort (>20 homes/mi^2)	1
Non-Agricultural	1

The Rotational Panel sample for the Pilot consisted of (approximately) a 20% subsample of the June Enumerative Survey (JES). The complete 1992 NASS sample for their JES in North Carolina had 321 segments contained in 96 (of 100) counties. Sample segments were selected from the most recent replication to enter the plan within each stratum. The 65 segments selected fell into 55 counties (Table 1.2).

1.4.4 Sample Unit Selections

During the JES, NASS enumerators obtained land use data on all areas of each sample segment. The location of each field in each sample segment was mapped on an aerial photograph and its identification number and acreage recorded. Fields eligible for the full survey were those that contained an annually harvested herbaceous crop (AHHC).

Prior to drawing the samples for the Fall Survey, field areas for both the hexagon and rotational sampling plans were expanded using NASS expansion factors in order to obtain lists of acres of AHHC in North Carolina for 1992. Multiplying an acre by its NASS expansion factor converts that acre to the number of acres in North Carolina that it represents. By expanding the field areas, each list represented an estimate of the population of AHHC acres in North Carolina. The two lists of areas were then sampled systematically to select every kth acre with random starts to identify the fields to be visited under each of the two plans for the Fall Survey. The sampling rate (k) was set to select an average of three selected acres per segment for a total of 348 selected sampling sites. Each randomly chosen acre identified a field to be visited but was not necessarily the sampled acre. A random start point for sampling site in each field was identified upon arrival at the field. The fields selected were identified and marked on the aerial photographs for use by NASS enumerators in collecting field data during the Fall Survey.

For the 65 segments of the Rotational Panel Plan, a total of 195 sample units were selected for the 51 segments of the Hexagon Plan, 153 sample units were selected (Table 1.3). (For 15 cases for the Rotational Plan and 20 cases) for the Hexagon Plan a field was selected more than once for sampling. (Table 1.4). When this occurred, the appropriate

Table 1.3 Land use or cover for the 1991-92 crop year on the sample unit in the Pilot Field Program in North Carolina. Cover crops were not included. Samples were selected using JES data, and some samples were chosen incorrectly and did not belong to the desired resource class (annually harvested herbaceous crops). Data presented are from the fall questionnaire. Questionnaire data are unavailable for some units because of farmer refusal or inaccessibility.

Crop or land use	Number of sample units, NASS design	Number of sample units, hexagon design
winter small grains (wheat, oat, barley, and rye)	8	5
small grain - soybean double crop	16	10
soybean (not double cropped)	40	29
corn for grain	34	34
corn for silage	10	6
popcorn	2	0
cotton	11	9
hay (all)	15	17
peanut	3	8
grain sorghum	1	0
sweet potato	3	0
tobacco	14	7
cucumber	1	0
pasture, clover, and grasses	4	7
set-aside or mostly set-aside	8	0
idle cropland	6	2
mixture of crops or land uses	7	3
woodland or wetland	5	1
refusal or inaccessible	7	15
TOTAL	195	153

numbers of random start points were selected for actual sampling. In some cases, a sample was not taken from the designated sample unit because it was refused by the owner/operator, inaccessible, not in the designated resource class, or there was an enumerator error (Table 1.5).

For ponds and wells, a number of ponds and wells per tract (i.e., per ownership unit that had a portion of that unit in the selected segment) were identified from the Rotational Panel plan only in the JES. Only those ponds and wells within the actual NASS segment were included in the final list for sample selection. If a segment contained 1 or 2 ponds, then each pond was included in the sample. If a segment contained more than 2 ponds, the selection probability for each pond was 2/n, where n is the number of ponds in that segment. If a segment contained 1, 2, or 3 wells, then each was included in the sample. If a segment contained more than 3 wells, the selection probability for each well was 3/m where m in the number of wells in the segment. A total of 51 ponds and 81 wells was identified for sampling. In some cases (Table 1.5) a sample was not taken from a pond or well because it was refused by the owner/operator, inaccessible, or not properly identified originally.

Table 1.4 Occurrence of multiple sample units in the same field. Includes fields that were later lost from the sample because of refusal, inaccessibility, etc.

Number of times a field was selected	Number of NASS fields	Number of hexagon fields
1	159	102
2	9	15
3	6	3
5	0	1
7	0	1

Table 1.5 Number of sample units for soil chemistry, nematodes, ponds, and wells for the 1992 Pilot in North Carolina.

	Number of Sample Units			
	NASS Rotational Panel Design		Hexagon Design	
	Sampled	Not Sampled[a]	Sampled	Not Sampled[a]
Soil chemistry	185	10	136	17
Nematodes in soil	185	10	0	0
Ponds	40	11	0	0
Wells	61	20	0	0

[a] Sample not taken because it was refused, inaccessible, or not in the designated resource class, or there was an enumerator error.

1.4.5 Sampling fields, ponds, and wells

Information on crops and land use for each field within a segment and the presence or absence of farm ponds or wells was obtained from the JES. For each field identified after the JES as part of the pilot sample, a NASS enumerator contacted the owner/operator of the field in the fall (November-December) and completed a questionnaire designed specifically for EMAP Agroecosystems. The questionnaire provided information on crop yields, land use and tillage history, fertilizer usage history, pest management practices, fuel usage, and irrigation practices.

Soil samples were taken along an approximately 90 m transect for each sample unit for the current season to determine soil physical and chemical properties and to quantify nematode community trophic structure. The transect was located in a pseudorandom manner and represented a 0.4 ha (=1 acre) portion of the field. Twenty soil cores (2.5 cm diam x 20 cm deep) were taken at equally spaced intervals along the transect. Soil was placed in a plastic bucket and mixed by hand. Portions were placed into plastic bags, one bag for physical/chemical analysis and one for nematode analysis. Bags were sealed, labelled, and shipped via Federal Express to commercial laboratories for analysis (Agrico, Agronomic Service Laboratory in Ohio for physical and chemical analysis; N&A Nematode Identification Services in California for nematode analysis). In every sixth field two independent transects were used to estimate within-field variation. In every 12th field, in addition to the additional or duplicate transect, twice as much soil was taken and the composite was divided into two samples for each of the physical/chemical and nematode analyses. The split samples were used to estimate within-sample variability.

Two methods were used to obtain water samples from ponds. In the first method, an enumerator took a boat out onto the pond. Water samples were collected from three points and three depths per point with a Kemmerer sampler and mixed to give one sample for each pond. In the second method, an enumerator used a stoppered teflon bottle at the end of a 5-m pole to obtain water samples from a shallow depth (0.3 m) from four points along the bank of each pond. Water from the four bank locations was mixed to form a composite sample. Each mixed sample was divided into two portions which were placed in separate glass bottles. Bottles were placed in styrofoam cases and shipped via Federal Express to the EPA Environmental Research Laboratory (ERL) in Athens, GA, for analysis of nitrate concentration and concentration of selected pesticides.

For well water samples, wells were purged by allowing water to flow from the water outlet. The purge time was 15 minutes if there was a tank between the well and the spigot, 5 minutes otherwise. Samples were then collected and placed into two glass bottles, one for nitrate analysis and one for pesticide analysis. Bottles were placed in styrofoam cases and shipped via Federal Express to the EPA ERL in Athens, GA.

1.5 STATISTICAL METHODS

The primary statistical tool used to summarize indicator data from the 1992 Pilot is the cumulative distribution function (CDF). The estimated CDF for a particular indicator gives the estimated proportion of the population that has values of the indicator less than or equal to any specified value -S. Ninety percent confidence bands are added to the CDFs in order to indicate the precision of the estimated CDF. Typically, in this report, the population is the area of land in North Carolina, the Piedmont region of North Carolina, or the Coastal Plain region of North Carolina cultivated with annually harvested herbaceous crops. Marginal maps indicate the region represented by the CDF. If no marginal maps are presented, the CDF represents all possible samples for that measurement or index.

As an example consider the estimated CDF for clay content of soil (0-20 cm depth) in the Coastal Plain of North Carolina (Figure 1.5).

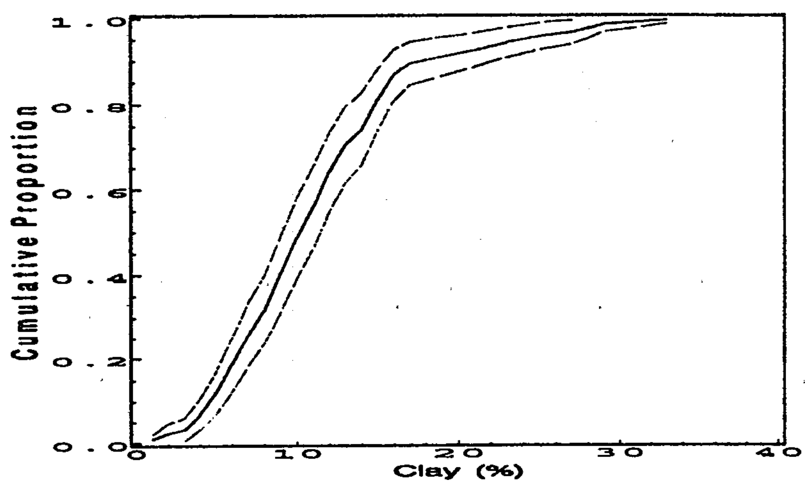


Figure 1.5 Cumulative distribution function for clay content of AP horizon at surface soil (0-20 cm deep).

The CDF shows that the proportion of land (in acres or hectares) cultivated with annually harvested herbaceous crops in the Coastal Plain with a clay content of 15% or less, for example, is estimated to be 0.80 with a 90% confidence interval of approximately (0.72, 0.88). Similarly, an estimate of the median clay content of annually harvested herbaceous cropland in the Coastal Plain is approximately 10%.

For each CDF, values of the measured or calculated quantity can be given for specified quantilies on portions of the CDF. For example, a value of 10% clay at a quantile of 50% would mean that 50% of the population of land cultivated in annually harvested herbaceous crops has 10% or less clay content in the AP horizon.

A CDF, by definition, will range from 0 to 1 on the vertical axis. An interpolation procedure is used to fit the estimated continuous CDF to the empirical step-function CDF. The procedure connects the midpoints of the steps for each observed value and is incorporated to remove bias. Hence, the estimated CDFs will not begin at 0 or end at 1. For many indicators, this is not noticeable.

In the 1992 Pilot, most data were collected using the Hexagon sampling design and the Rotational Panel sampling design (Section 1.4). Nematodes were only sampled for sample units of the Rotational Panel design. For indicators using fall questionnaire or soil sample data and both designs, CDFs computed from each design were combined to form the CDFs reported in Chapter 2. An average variance was computed for each design using the variances at the observed values for the combined dataset. The combined CDF was defined as:

$$\hat{F}(t) = \hat{w} * \hat{F}_{rp}(t) + (1-\hat{w}) * \hat{F}_{hex}(t)$$
$$\text{Where } \hat{w} = \beta/(1+\beta), \quad \beta = \overline{VAR}(\hat{F}_{hex}) / \overline{VAR}(\hat{F}_{rp})$$

The above weighting provides an approximately optimal combination of the two CDFs in terms of overall variability.

Regardless of which sampling design is ultimately selected by the Agroecosystem Resource Group, land-use and extent indicators will continue to be computed from NASS's June Enumerative Survey in order to take advantage of this resource. Thus, although land-use and extent data were collected in June using the Hexagon design, CDFs for these indicators represent estimates based only on the complete June Enumerative Survey performed using the Rotational Panel design. For June Enumerative Survey and fall survey data, variances computed from the Rotational Panel design account for the stratified nature of this design.

2. RESULTS

Results are presented for the indicator categories land use, crop productivity, soil quality-physical and chemical, soil biotic diversity, water quality and agricultural chemical use. The results for each indicator are presented in a question and answer form similar to what might be used in the future for annual statistical summaries. In a few cases, boundaries for acceptable (nominal), marginal, and unacceptable (subnominal) conditions are presented. In most cases, however, their boundaries are not proposed and will be developed for future reports.

2.1 EXTENT AND GEOGRAPHIC DISTRIBUTION OF ANNUALLY HARVESTED HERBACEOUS CROPS

Agricultural lands in North Carolina are scattered across the state in several different types of landscape, largely intermixed with forest (Figure 2.1). For this Pilot, we examined only a portion of North Carolina's agriculture – the land planted in annually harvested herbaceous crops. *Annually harvested herbaceous crops* are herbaceous plants that are harvested every year, regardless of whether the plant itself is annual or perennial (Table 2.1). These crops are planted on about 1.68 million hectares in North Carolina, covering some 13% of the total land area of the state (Table 2.2). Soybean, accounting for one third of the annually harvested herbaceous cropland, is the most common crop.

Table 2.1. Annually harvested herbaceous crops included in the Pilot.

Barley	Rye
Corn	Soybeans
Cotton	Sweet potatoes
Hay	Sorghum
Alfalfa hay	Sunflowers
Grain hay	Tobacco
Other hay	Burley
Oats	Flue-cured
Peanuts	Vegetables (all)
Potatoes	Winter wheat

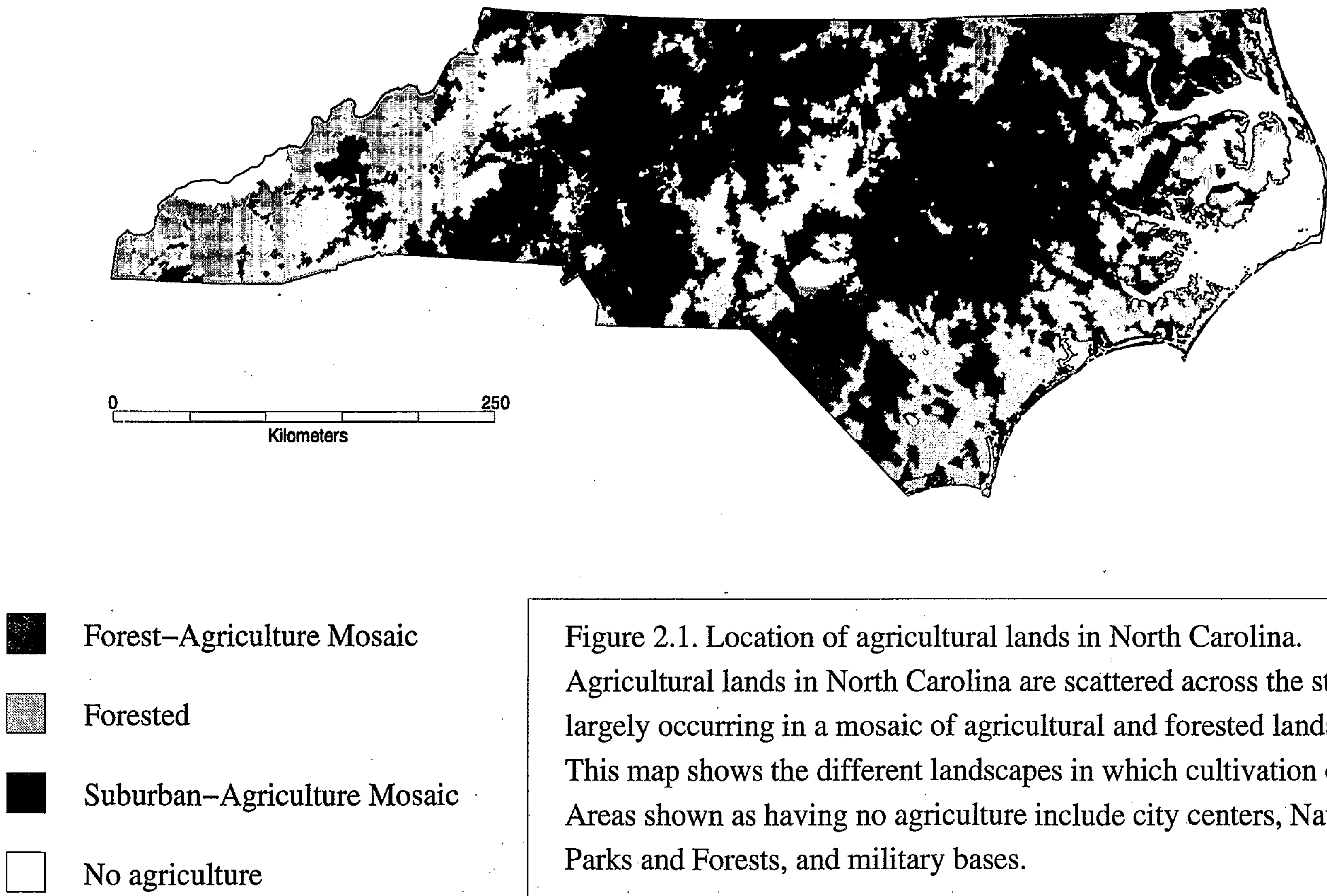


Figure 2.1. Location of agricultural lands in North Carolina. Agricultural lands in North Carolina are scattered across the state, largely occurring in a mosaic of agricultural and forested lands. This map shows the different landscapes in which cultivation occurs. Areas shown as having no agriculture include city centers, National Parks and Forests, and military bases.

Source: USDA National Agricultural Statistics Service

Table 2.2 Extent of annually harvested herbaceous cropland in North Carolina.

Landscape Type	Extent of AHHC[a] (hectares) (with 95% confidence)	Proportion of Total AHHC[a] (percent)
Forest-Agriculture Mosaic	1,401,021 ± 135,249	83
Forested	217,259 ± 87,611	13
Suburban-Agriculture Mosaic	61,816 ± 35,353	4
Non-Agricultural	0	0
State Total	1,680,096 ± 163,444	100

[a] AHHC is annually harvested herbaceous crop.

Information about the extent and distribution of annually harvested herbaceous crops was obtained by analyzing data from the complete National Agricultural Statistics Service's June Enumerative Survey. The Survey design is based on stratification of land by intensity of agricultural use (see Section 1.5). Each stratum is subdivided into *segments*, approximately 260 hectares (one square mile) in size, a sample of which is surveyed during the June Enumerative Survey. In North Carolina, most segments contain more than one field, each of which is identified during the Survey.

What follows is a description of the extent and distribution of agricultural lands, particularly land devoted to annually harvested herbaceous crops, in North Carolina. Information is presented at the scale of broad landscapes (based on the strata); smaller areas within similar landscapes (based on segments within strata); and individual fields (based on fields within segments). As you examine this information, you may notice that the sample size is not constant. Changes in sample size reflect differences in both the questions and the scale of observation. Table 2.3 summarizes the different sample sizes seen in this section of the report. No further information is provided about the suburban-agriculture mosaic because of the limited extent of annually harvested herbaceous crops in that type of landscape and the small number of samples.

You may also notice that the other indicators summarized in this section are reported for regions that differ from the landscapes shown in Figure 2.1 and Tables 2.2 and 2.3. The reasons for this inconsistency, and several potential solutions, are discussed under the heading *Reporting on Consistent Regions* in Section 3.2.5.2.

Table 2.3 Sample sizes from the June Enumerative Survey for each landscape type.

Landscape Type (Stratum)	Number of Segments	Number of Segments with AHHC[a]	Number of Fields
Forest-Agriculture Mosaic	170	160	2671
Forested	80	33	266
Suburban-Agriculture Mosaic	53	18	49
Non-Agricultural	18	0	0
State Total	321	211	2986

[a] AHHC is annually harvested herbaceous crops

How much of the land is cropped?

If you look at a random 260 hectare tract of land in North Carolina, how much of that area would you expect to be planted in annually harvested herbaceous crops? Because crops are unevenly distributed, the answer depends upon which part of the state you are in. Figure 2.2 shows the proportion of 260 hectare tracts in which *P*% or less of the area is cropped. There are separate graphs for parts of the state that are a forest-agriculture mosaic and those that are forested. As you might expect, land in the forest-agriculture mosaic landscape is more heavily cropped than land in forested landscapes.

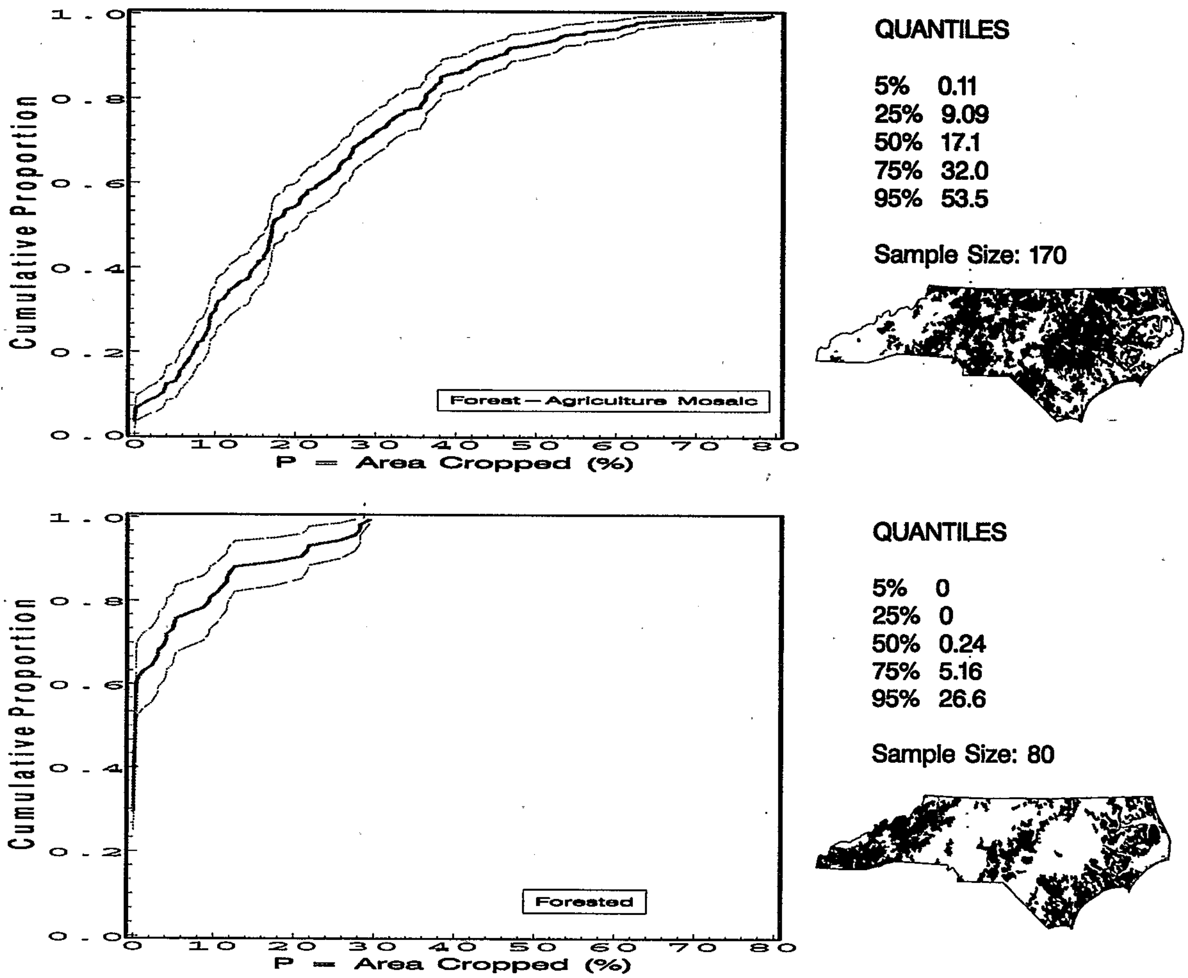


Figure 2.2 Cumulative distribution functions for percent of land area with annually harvested herbaceous crops per 260 hectare tract.

How diverse are the state's croplands?

A more diverse agriculture may be less vulnerable to disease, pest infestation, and changing economic conditions. The diversity of a collection of items – in this case, crops – is usually measured in two ways: the number of different items, and their relative abundance. When comparing the diversity of crops in several areas, each area should be of equal size; diversity can change with area in surprising ways. First, let's look at the number of different items. *In a random 260 hectare tract of land, how many different annually harvested herbaceous crops would you be likely to find?* Again, the answer depends on which part of the state you are in. Figures 2.3 shows the proportion of 260 hectare tracts containing *S* or fewer crops. As you might expect, tracts in the forest-agriculture mosaic landscape have greater diversity of crops than those in the forested landscape.

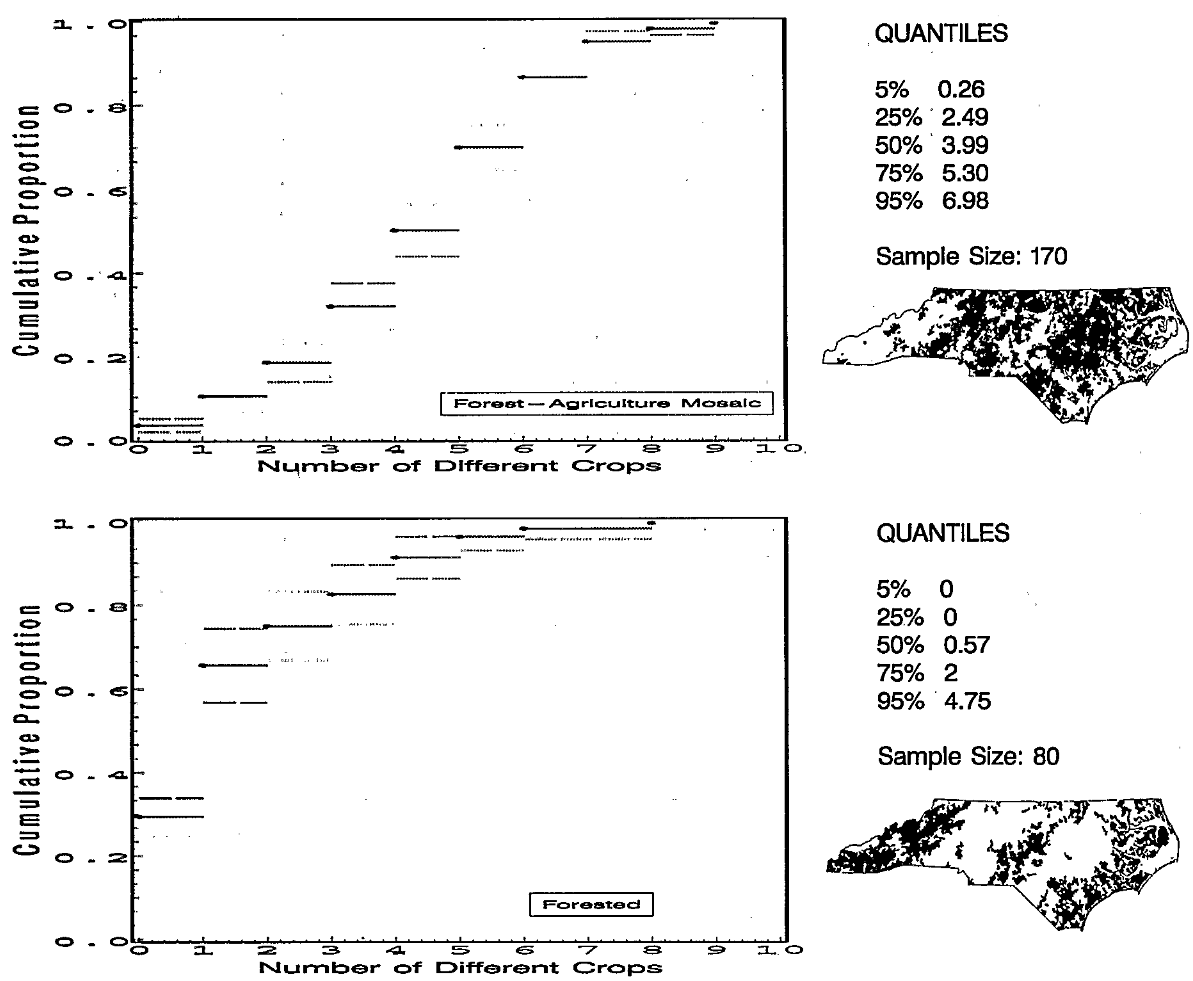


Figure 2.3 Cumulative distribution function for number of annually harvested crops per 260 hectare tract.

Do these cropped areas tend to be dominated by a single crop?

The answer to this question provides a measure of relative abundance, the second aspect of diversity. For example, two tracts of equal size may each contain corn and soybeans. In tract 1, corn may account for 90% of the cropped land, with 10% in soybeans; in tract 2, corn may account for 60% of the land. Corn is said to *dominate* tract 1; in tract 2 the crops are said to have a more *even* diversity. Large areas dominated by a single crop, particularly if crop rotation is not practiced, may be more vulnerable to disease and pest infestations. *For the 260 hectare tracts in which annually harvested herbaceous crops occur, is there a single crop that dominates the cropped area?* Figure 2.4 shows the proportion of tracts in which P% of the annually harvested herbaceous cropland is accounted for by a single crop - the one with the largest area.

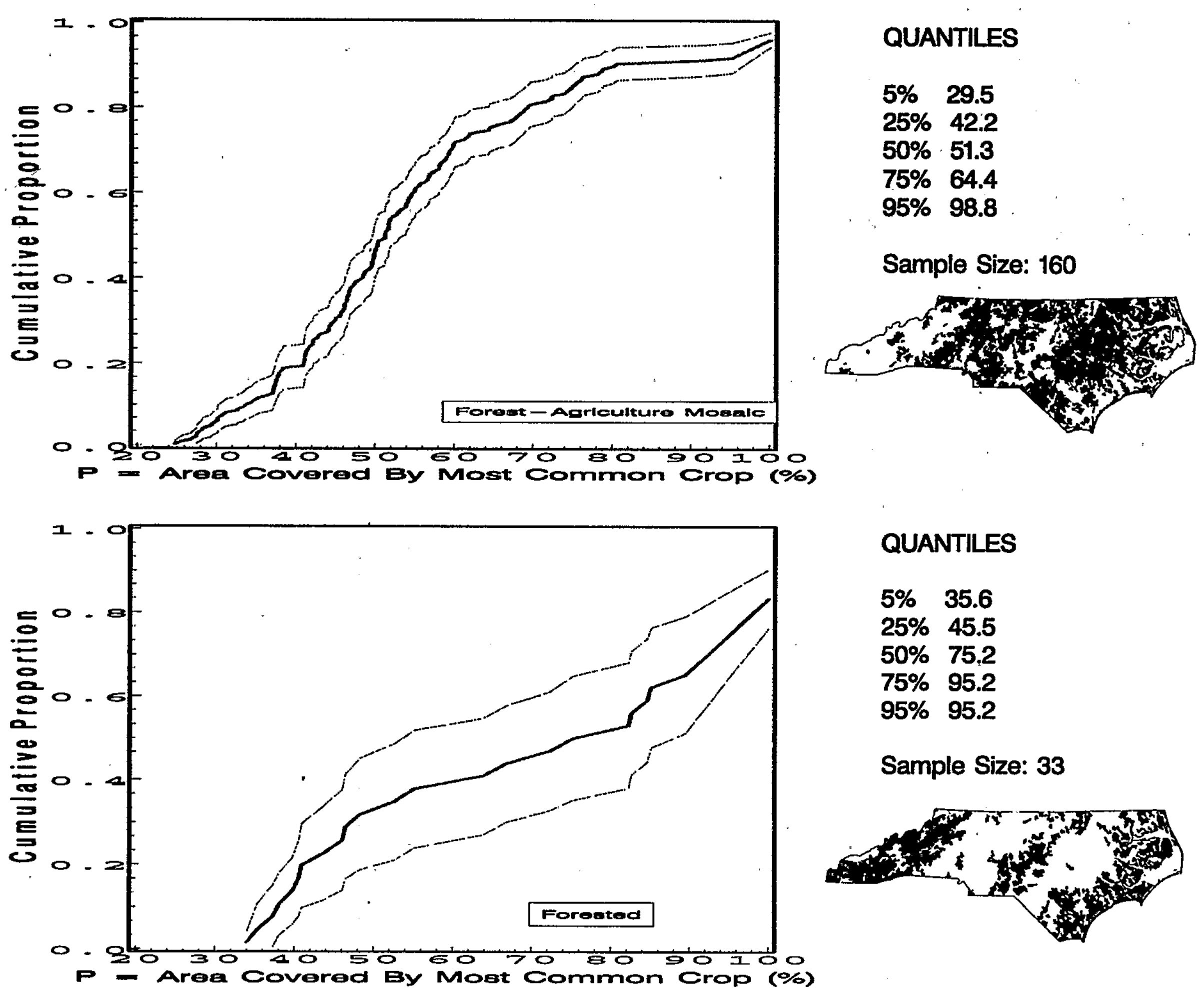


Figure 2.4 Cumulative distribution functions for percent of area covered by most common annually harvested herbaceous crop per 260 hectare tract.

How large are the crop fields?

If you were to measure the size of a random field of an annually harvested herbaceous crop how large would you expect the field to be? The distributions in Figure 2.5 show the proportion of fields of size *S* or smaller. There are separate distributions for the entire state and for the forested and forest-agriculture mosaic landscapes. North Carolina fields tend to be small, with a median field size of just over 2 hectares regardless of the type of landscape.

How do different sized fields contribute to the *total area* of annually harvested herbaceous cropland?

What proportion of the total area of annually harvested herbaceous cropland occurs in fields of a given size or smaller? The distributions in Figure 2.6 show the proportion of the total area of annually harvested herbaceous cropland accounted for by fields of size *S* or smaller. For the state, half the total area of annually harvested herbaceous cropland occurs in fields smaller than 6.09 hectares.

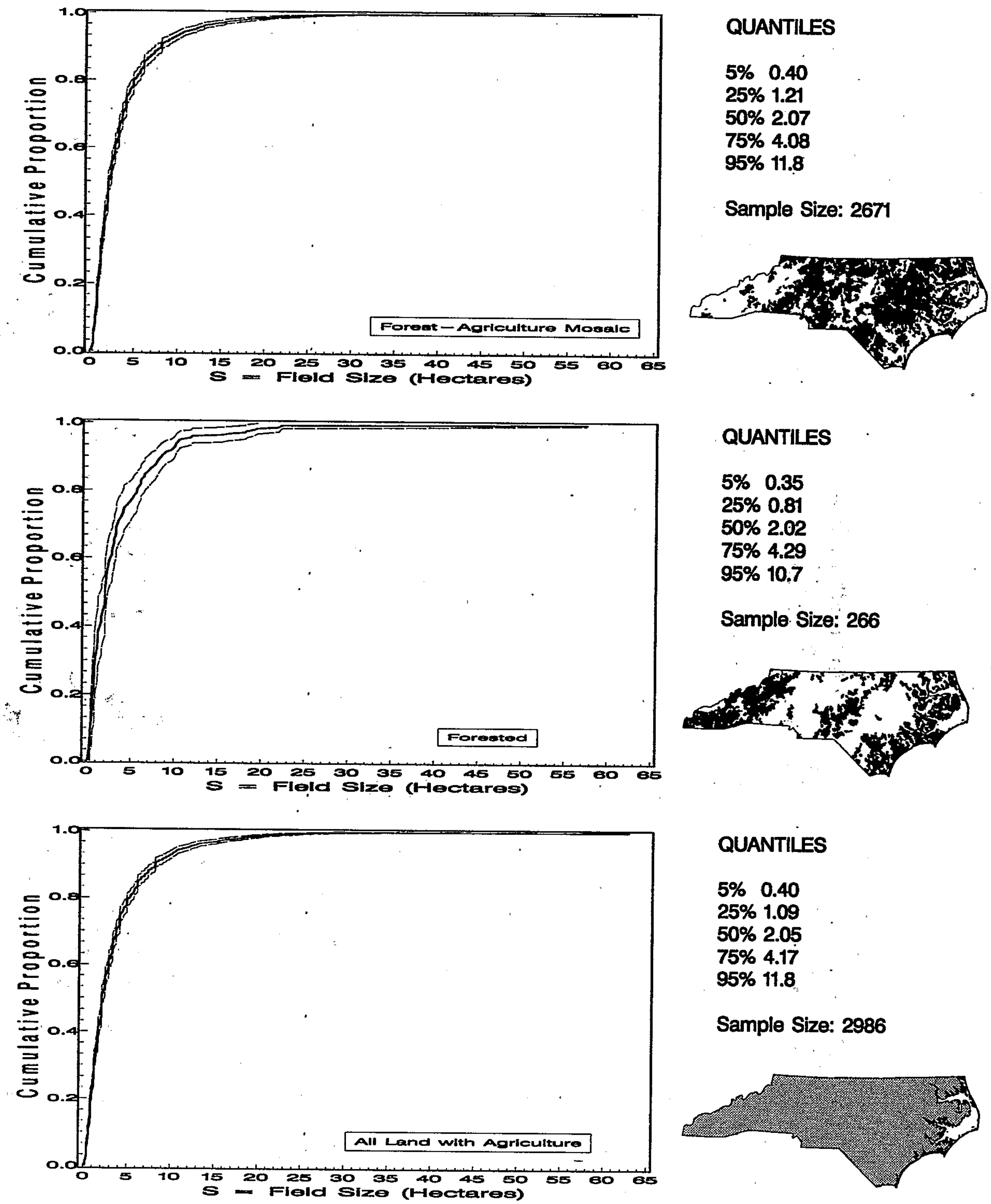


Figure 2.5 Cumulative distribution functions for the proportion of fields of size S or smaller.

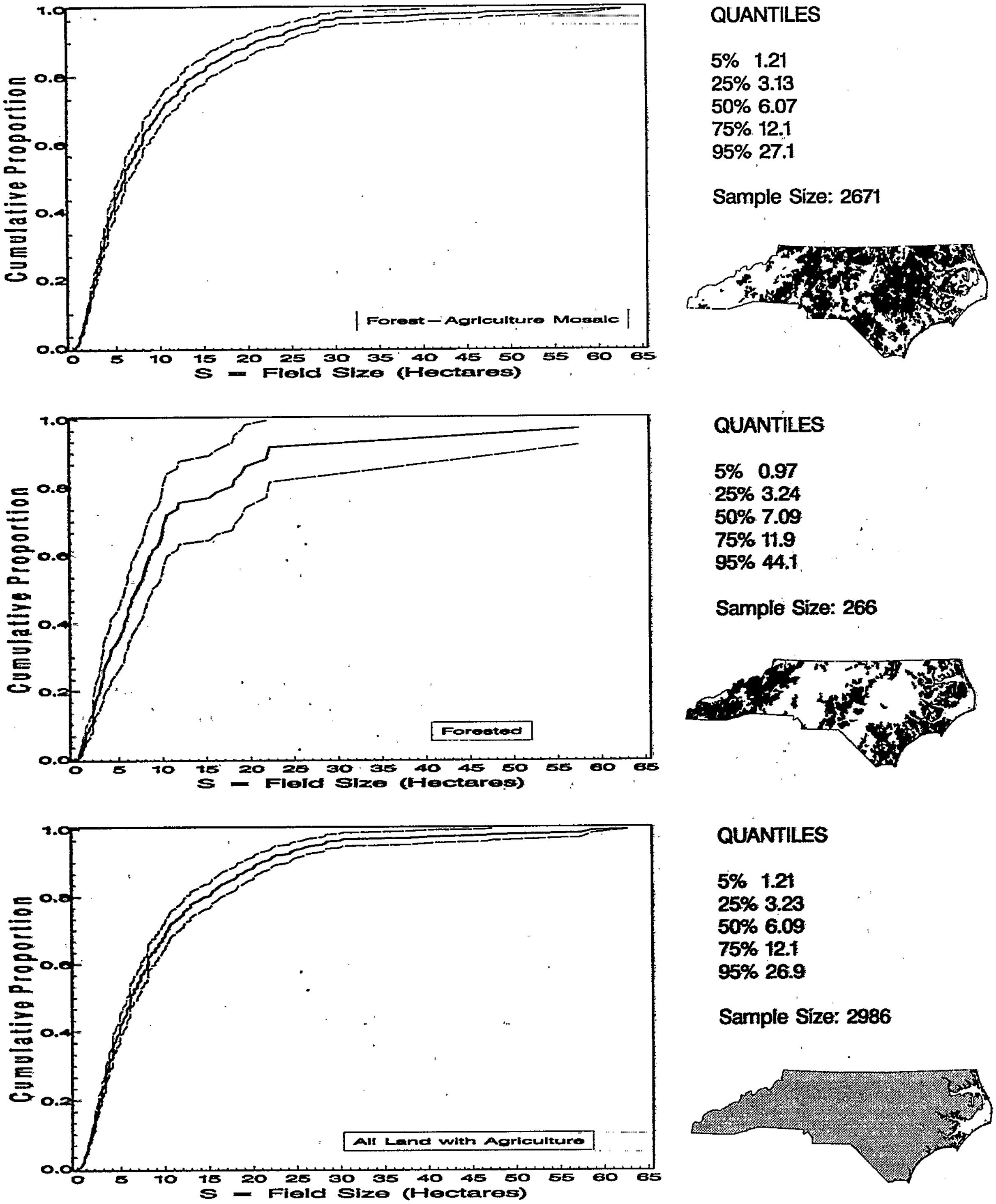


Figure 2.6 Cumulative distribution functions for the proportion of total area of annually harvested herbaceous cropland accounted for by fields of size S or smaller.

2.2 CROP PRODUCTIVITY

Productivity of agroecosystems is a broad concept with many possible definitions. Two indices of productivity are summarized here: nitrogen use efficiency and standardized yield. So far, indicator development in this area has focused on annually harvested herbaceous crops (AHHCs).

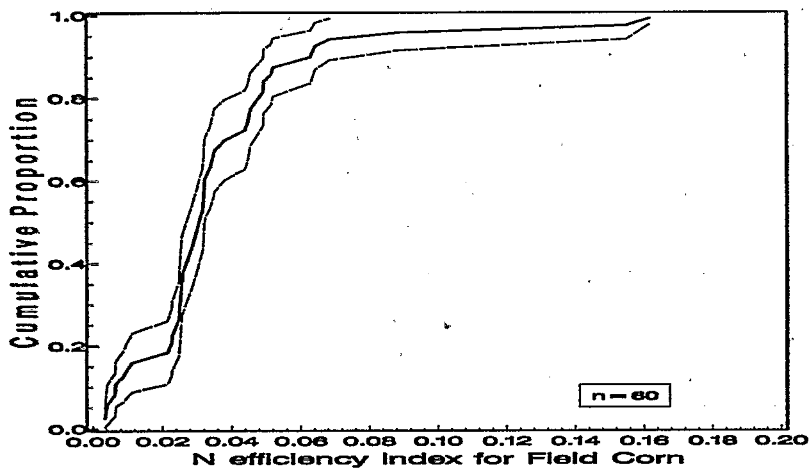
How efficiently is nitrogen being used to produce annually harvested herbaceous crops?

Nitrogen is a key plant nutrient that is often applied in large quantities to agricultural fields, either in the form of animal wastes or as commercial fertilizer. The latter requires large amounts of energy to produce (Pimentel et al., 1973, Southwell and Rothwell, 1977). If sustainability is a goal, these nitrogen subsidies should be used efficiently. Efficiency is also important because nitrogen, regardless of its source, is considered a contaminant if it reaches ground or surface water (Spalding and Exner, 1993, Angle et al., 1993).

A nitrogen efficiency index was calculated from the yield and management information obtained from the 1992 Pilot questionnaire. The index is the ratio of the amount of nitrogen applied on a field to the harvested yield from that field. In double-crop situations (fields with two harvested crops in a single season), the total weight of applied nitrogen was divided by the total weight of harvested material for both crops. All yield values used in the nitrogen efficiency index as reported here were expressed on a dry matter basis, using conversions that were obtained from university publications, extension personnel, and NASS. Both the numerator and the denominator of the index are expressed in the same units, so the ratio is dimensionless. This approach gives an indication of the efficiency of the system, not of individual plants. Also, note that smaller index values represent greater efficiency. This "reverse" interpretation was configured so that fields with no applied nitrogen could be included (otherwise division by zero excludes them). Fields with reported yields of zero were excluded from this calculation (see Section 3.2.1 for details).

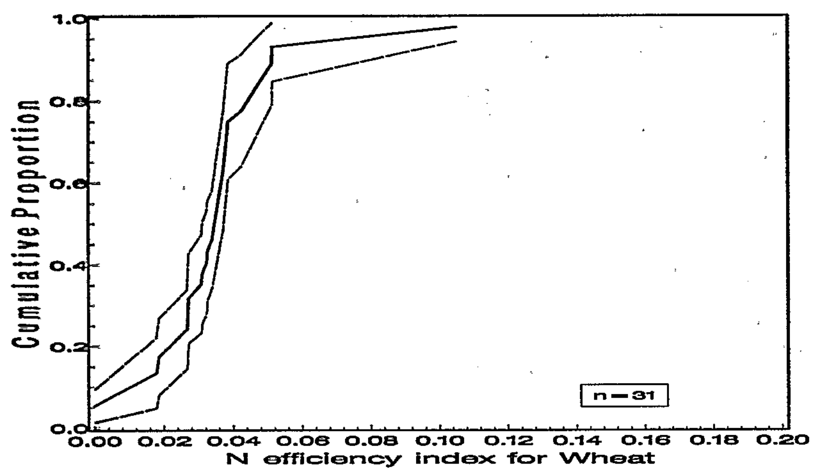
Nitrogen inputs from commercial fertilizer were provided by NASS. For estimating the nitrogen content of animal manures, conversion factors were used from *The 1991 North Carolina Agricultural Chemicals Manual* (Barker et al., 1991). Manure nitrogen could be estimated for only 19 sample units, out of a total of 33 on which manure was applied to an AHHC (see Section 3.2.1). The remaining 14 sample units were dropped from the analysis.

The cumulative distribution functions of the nitrogen efficiency index for two important cereal crops, field corn and wheat, are depicted in Figures 2.7. The two curves are quite similar. Contrast these with the graphs for soybean and cotton in Figures 2.8. The association of the soybean plant with N_2-fixing bacteria increases its efficiency (lowers the index). Cotton's low efficiencies (large index values) probably come from the relatively low weight of lint compared to grain.



QUANTILES

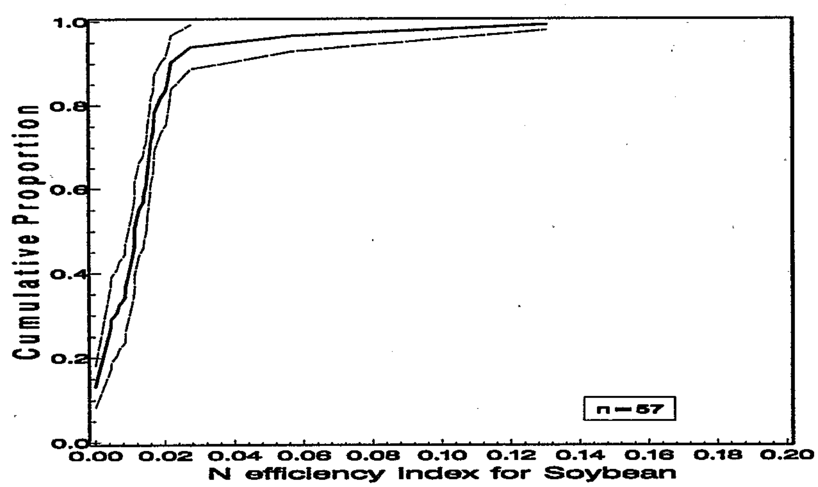
5% 0.004
25% 0.024
50% 0.031
75% 0.045
95% 0.079



QUANTILES

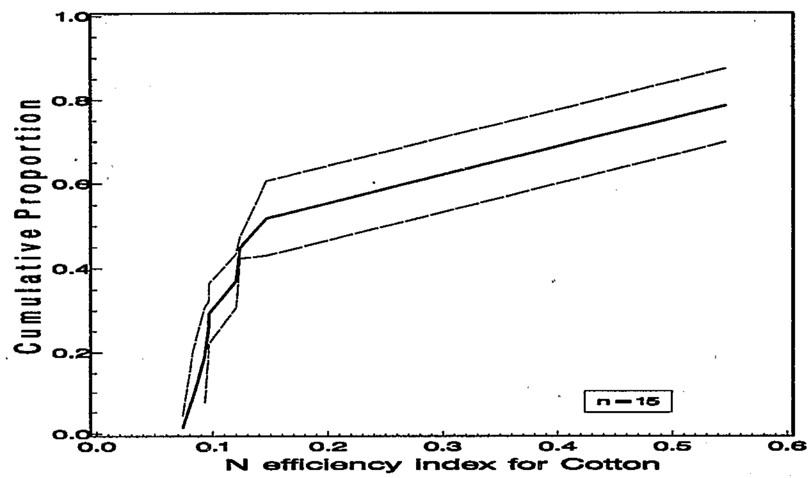
5% 0.000
25% 0.027
50% 0.035
75% 0.039
95% 0.074

Figure 2.7 Cumulative distributions functions for the nitrogen efficiency index for land in field corn (maize) and wheat. Larger index values indicate lower efficiency. The 90% confidence interval is shown by dashed lines.



QUANTILES

5% 0.000
25% 0.004
50% 0.011
75% 0.017
95% 0.040



QUANTILES

5% 0.000
25% 0.096
50% 0.141
75% 0.493
95% 0.493

Figure 2.8 Cumulative distribution functions for the nitrogen efficiency index for land in soybean and cotton. Larger values indicate lower efficiency. Dashed lines show the 90% confidence interval. Note difference in scale on the horizontal axes.

In order to show the overall condition of cropland in a region, it is our goal to present one composite index for nitrogen efficiency. It would be misleading to do so when the harvested plant parts are so different; however, it seems acceptable to pool the indices across the following *seed* crops for which we have both yield data and an approximate conversion to dry matter: wheat, oat, barley, rye, sorghum, field corn (maize), soybean, and peanut. Together, these crops represent 49% of the AHHC sample units from the fall survey. The CDFs of the nitrogen efficiency index, combined across seed crops, are summarized in Figure 2.9. Note that the distribution for the piedmont has a long tail and a large 95th percentile, indicating that the most inefficient conversion of nitrogen into harvested crop occurred there.

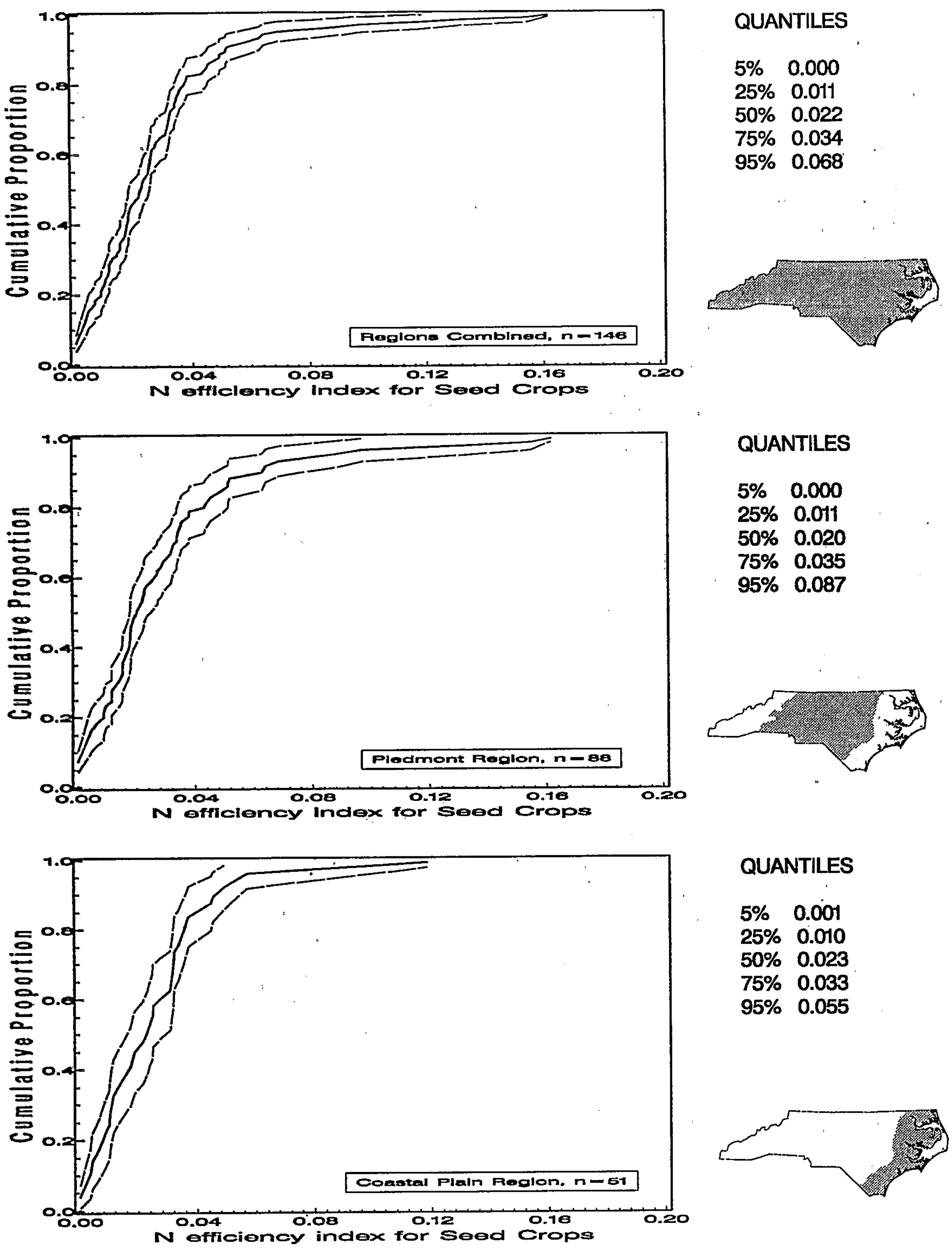


Figure 2.9 Cumulative distribution functions for the nitrogen efficiency index for land area cropped to eight seed crops. Crops included are wheat, oat, barley, rye, sorghum, field corn (maize), soybean, and peanut. Larger index values indicate lower efficiency.

Are yields meeting expectations based on soil and climate?

The Pilot Plan for 1992 proposed an index comparing each reported yield to the yield expectations for the soils on the sample field. These expected yields for soil map units within a county were not available for a fixed point in time, so the calculation of this index was postponed. In its place, an observed/expected index based on historical county average yields is presented in Section 3.2.1.1, illustrating a method for combining data across many crops.

2.3 SOIL QUALITY - PHYSICAL AND CHEMICAL

Physical and chemical properties of soils largely determine and reflect the productive potential of land. For the 1992 Pilot, we analyzed composite samples for: particle size, organic matter, cation exchange capacity, select micro and macro nutrients, and some contaminants related to sludge application. In general, geographical differences were found between the individual parameters, which are likely related to the innate characteristics of the soils. As discussed in Section 3, we are evaluating whether it is necessary to consider innate soil properties while trying to determine soil "condition."

Three cumulative distribution functions (CDFs) are presented for each parameter: all North Carolina samples combined, Coastal Plain samples, and Piedmont samples. Samples collected in the mountain region are not presented in a separate CDF because there were only eight samples; however, these data are included in the CDFs for the state summaries. For the CDFs presented, the sample size is n= 233 for the Piedmont region and n= 80 for the Coastal Plain region.

In the following discussions, we refer to proportions or percentages of samples, e.g., "About 5% of the Piedmont samples had < 35% clay." Because of the statistical design, these percentages or proportions also represent "the proportion of land area planted in annually harvested herbaceous crops." In this context, "5% of the Piedmont samples" is a simplified but analogous way of saying that "5% of the Piedmont area that is planted in annually harvested herbaceous crops".

Acceptable (nominal) and unacceptable (subnominal) ranges are proposed which may be related to whether a region has sufficient soil quality to sustain crop and non-crop productivity. The proposed limits are tentative and are presented in this report primarily as illustrations. The rationale for the limits are described in the following sections.

2.3.1 Soil Physical Condition

The physical condition of soil was determined by analyzing composited samples for particle size and organic matter content. Each of these parameters is presented separately in the first two subsections. A third subsection combines the results to provide a more integrated evaluation of soil physical condition. CDFs accompany the first two subsections but not the third.

2.3.1.1 Particle size

Do the soils have acceptable amounts of clay to sustain crop production?

Particle size analysis (percent sand, silt, and clay) reflects an innate soil property that is related primarily to geologic parent materials, transport processes, and soil age. An ideal soil for most plant growth and for soil management contains between 18-35% clay. Soils with >35% clay are difficult to till, conduct water too slowly, and are not porous enough to allow for adequate supply of oxygen to the plants. Soils with <18% clay tend to be droughty, do not retain nutrients very well, and compact easily.

CDF results for percent clay are presented in Figure 2.10. About 5% of the sampled Piedmont soils and none of the Coastal Plain samples had "too much" clay (i.e., > 35%). About 55% of the Piedmont samples and 90% of the Coastal Plain samples were subnominal because they had "too little" clay (i.e., < 18%). About 60% of the Piedmont samples had nominal amounts of clay (i.e., 18-35%), compared to about 10% of the Coastal Plain samples.

2.3.1.2 Soil organic matter in AP horizon

Do the soils have acceptable levels of organic matter in order to provide aeration to the roots and retain nutrients?

Soil organic matter is an essential component of a healthy soil. Organic matter influences: 1) the capacity of a soil to supply nutrients and trace metals to plants; 2) infiltration and retention of water; 3) aggregation and structure that affect air and water relationships; 4) cation exchange capacity; and 5) soil color, which impacts temperature relationships (Nelson and Sommers, 1982). Soils with low organic matter (<1%) content tend to be dense and less able to retain and provide nutrients to plants. In North Carolina, soils with organic matter (>20%) are organic soils that have been drained. These soils may have drainage problems or other limitations to the productivity of annually harvested herbaceous crops, although they could support non-crop, wetland vegetation. Organic matter content in the AP horizon is a more responsive monitoring parameter than soil texture, because OM content is highly dependent on management factors.

CDFs for percent organic matter are presented in Figure 2.11. Soil with less than 1% organic matter is tentatively proposoed as being in unacceptable condition for plant growth, and those with more than 1% organic matter as acceptable. None of the soil samples had more than 5% organic matter except for eight samples from the Coastal Plain. All of these samples came from one segment within the Coastal Plain. This area is mapped by the USDA Soil

Conservation Service as a conglomeration of different soil types, all of which are mucks (i.e., organic soils).

About 30% of the samples in NC had amounts of organic matter <1%. More than 30% of the Piedmont soils had 1% organic matter compared to about 20% of Coastal Plain soils. No samples in the Piedmont had > 5% of organic matter; the eight muck soils from the Coastal Plain accounted for 10% of the area with > 5% organic matter.

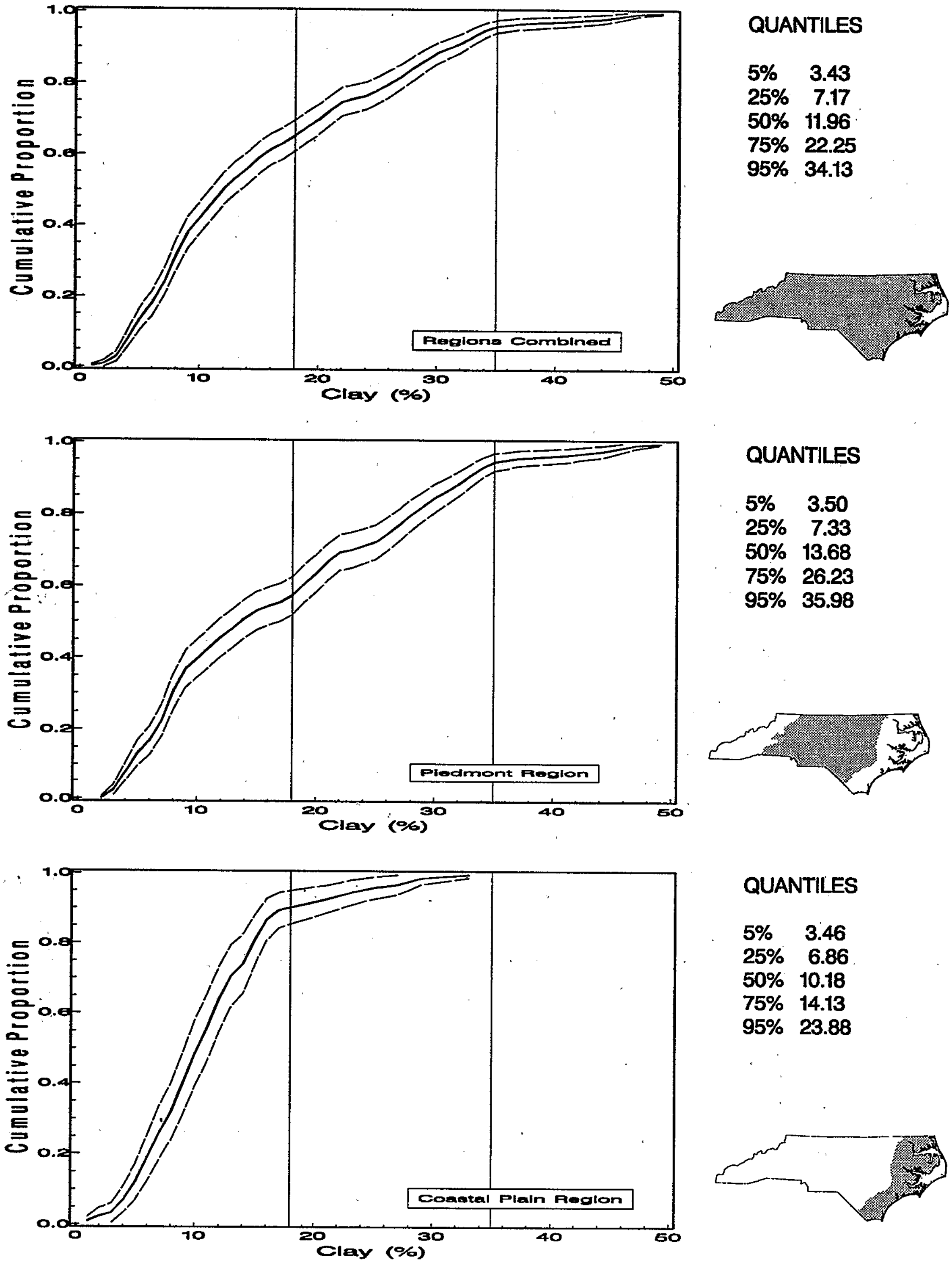


Figure 2.10 Cumulative distribution functions for percent clay in AP horizon. Cumulative proportions refer to the land area cropped with annually harvested herbaceous crops.

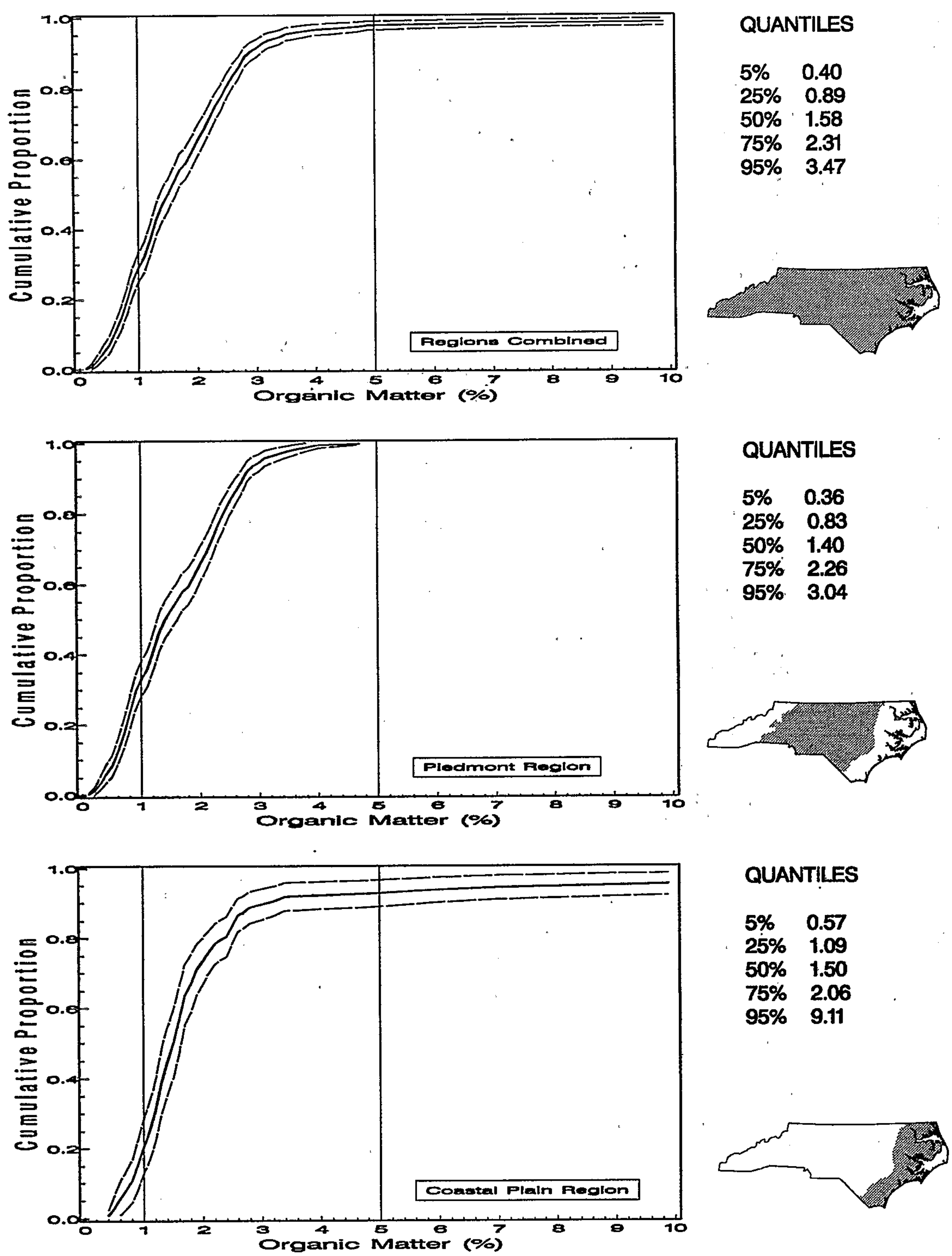


Figure 2.11 Cumulative distribution functions for percent organic matter in AP horizon. Cumulative proportions refer to the land area cropped with annually harvested herbaceous crops.

2.3.1.3 Combined assessment of organic matter and clay

Organic matter data cannot be interpreted apart from clay content data because there is a correlation between the two properties. Results from the 1992 Pilot indicate higher organic matter content in finer-textured soils (Table 2.4). The exception to this is the organic soils from the Coastal Plain, which have high organic matter content and low clay content.

Table 2.4. Organic matter content for each soil textural type. Clay percentage classes are established by the USDA Soil Conservation Service.

% Clay	Number of Samples	O.M. Content Range	Average	Median
0-18	212	0.1-18.7	1.74	1.2
18-35	98	0.6-4.7	2.17	2.2
>35	11	1.9-3.4	2.46	2.3

The results of the two properties were graphed together to provide a snapshot of the proportion of soils that have acceptable amounts of both clay and organic matter to sustain plant growth (Figure 2.12). For example, 27% of the Piedmont samples with acceptable amounts of organic matter had unacceptable subnominal amounts of clay. The information from the various sections of the graphs is summarized in Table 2.5.

2.3.2 Soil Chemical Properties

Soils with acceptable chemical properties allow nutrients to become available to plants and are not laden with toxic substances that threaten biological health of the agroecosystem. To assess soil chemical condition, we measured: pH (in water), cation exchange capacity (CEC), available phosphorus, total cadmium, and total lead. Cation exchange capacity and pH are largely responsible for the availability and supply of nutrients to the plants. Phosphorus is an essential plant nutrient. Cadmium and lead are micronutrients that are toxic to biota at certain levels.

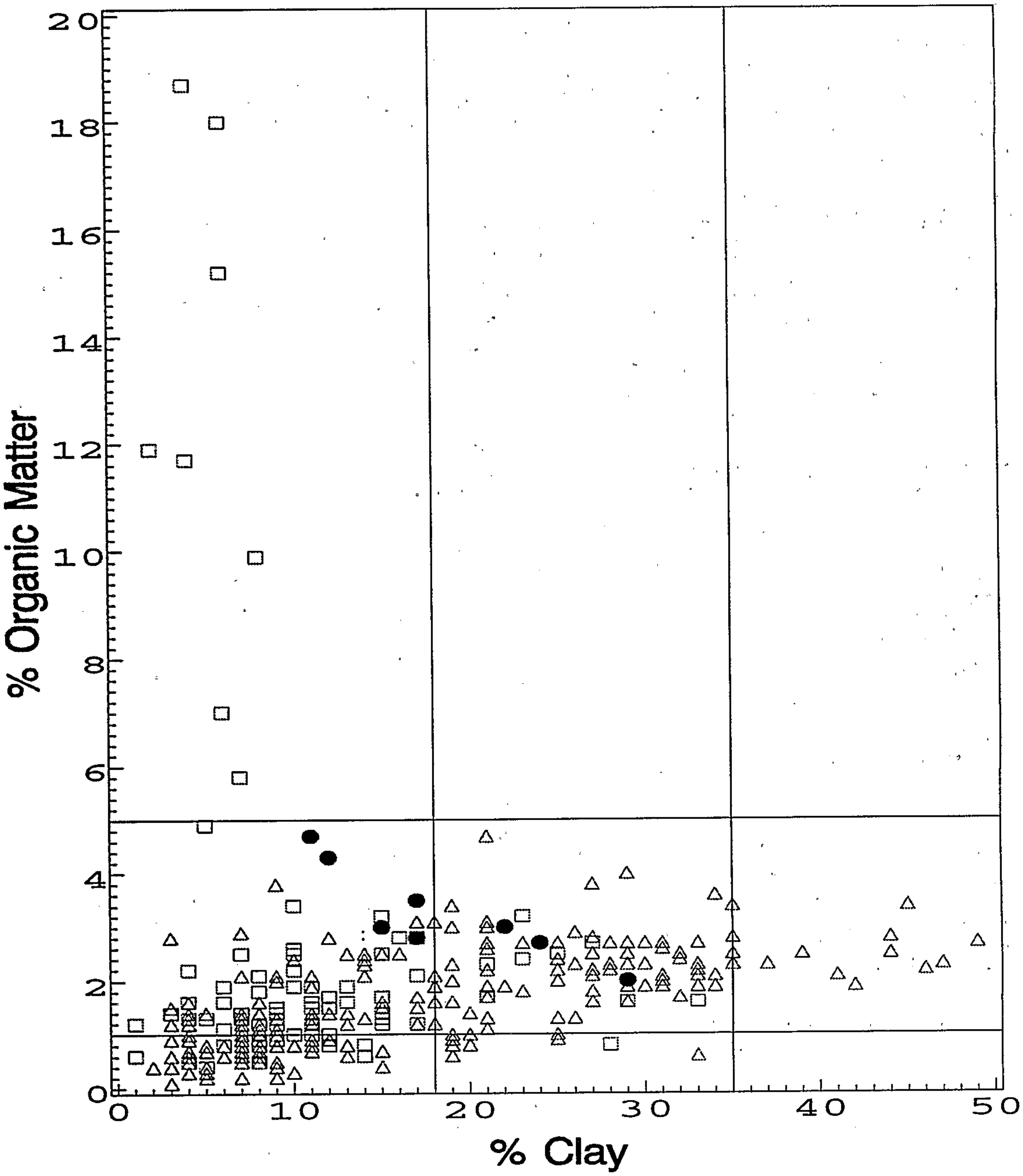


Figure 2.12 Percent clay in the AP horizon versus percent organic matter content in the AP horizon in each soil sample from North Carolina. Soils with 1% or more organic matter and between 18-35% clay have acceptable physical condition. See Table 2.5 for a summary of the percentages of samples in each compartment of this graph.

Table 2.5 Percentages of samples occupied by each compartment of the graph from organic matter and clay content in the AP horizon for each soil sample in Figure 2.12. Shaded areas show the percentages of soil samples with nominal amounts of both clay and organic matter content for the various regions and for the entire state.

Piedmont n=233	< 18% clay	18-35% clay	>35% clay
<1% O.M.	29	3	0
1-5% O.M.	27	36	5
>5% O.M.	0	0	0
Coastal Plain, n=80			
<1% O.M.	15	1	0
1-5% O.M.	64	10	0
>5% O.M.	10	0	0
Mountain, n=8			
<1% O.M.	0	0	0
1-5% O.M.	63	37	0
>5% O.M.	0	0	0
State, n=321			
<1% O.M.	25	2	0
1-5% O.M.	37	30	3
>5% O.M.	2	0	0

2.3.2.1 Soil pH

Do the soils have pH levels that facilitate the availability of essential plant nutrients?

Soil pH is an indicator of possible chemical constraints to the growth of roots and other biological communities. Chemical constraints usually associated with pH include the availability of inhibitory compounds (e.g., aluminum, salts), or a nutrient deficiency (e.g., phosphorus fixation) (Pierce et al., 1991). At a soil pH <4.5, root growth is often limited due to toxicity of soluble trace metals and H^+ ions. At pH values of 4.5-5.8, aluminum is in a soluble form, which limits growth of sensitive plants and microorganisms. Soil pH 6.5 is non-limiting to root growth. For these reasons, unacceptable pH values are considered to be <4.5, marginal values between 4.5-5.8, and acceptable values >5.8 for North Carolina soils.

None of the soil samples from the Piedmont had unacceptable pH values, and only one sample from the Coastal Plain was below pH 4.5 (Figure 2.13). This same sample had 18% organic matter and was one of eight samples from one segment.

2.3.2.2 Cation exchange capacity

Do the soils have cation exchange capacities that enable nutrient storage and supply?

Cation exchange capacity (CEC) is defined as the sum total of exchangeable cations that a soil can adsorb. Soils with low CEC are susceptible to leaching and are unable to retain nutrients and slowly release them over the growing season. Organic matter and clay content account for most of a soil's CEC. Soils with low CEC may reflect poor management practices.

As shown in the CDFs (Figure 2.14), a higher proportion of Coastal Plain samples had higher CECs than the Piedmont samples. These data need to be examined to identify if the high CECs in the Coastal region are the organic soils (i.e., the mucks), which are present in the wetland areas.

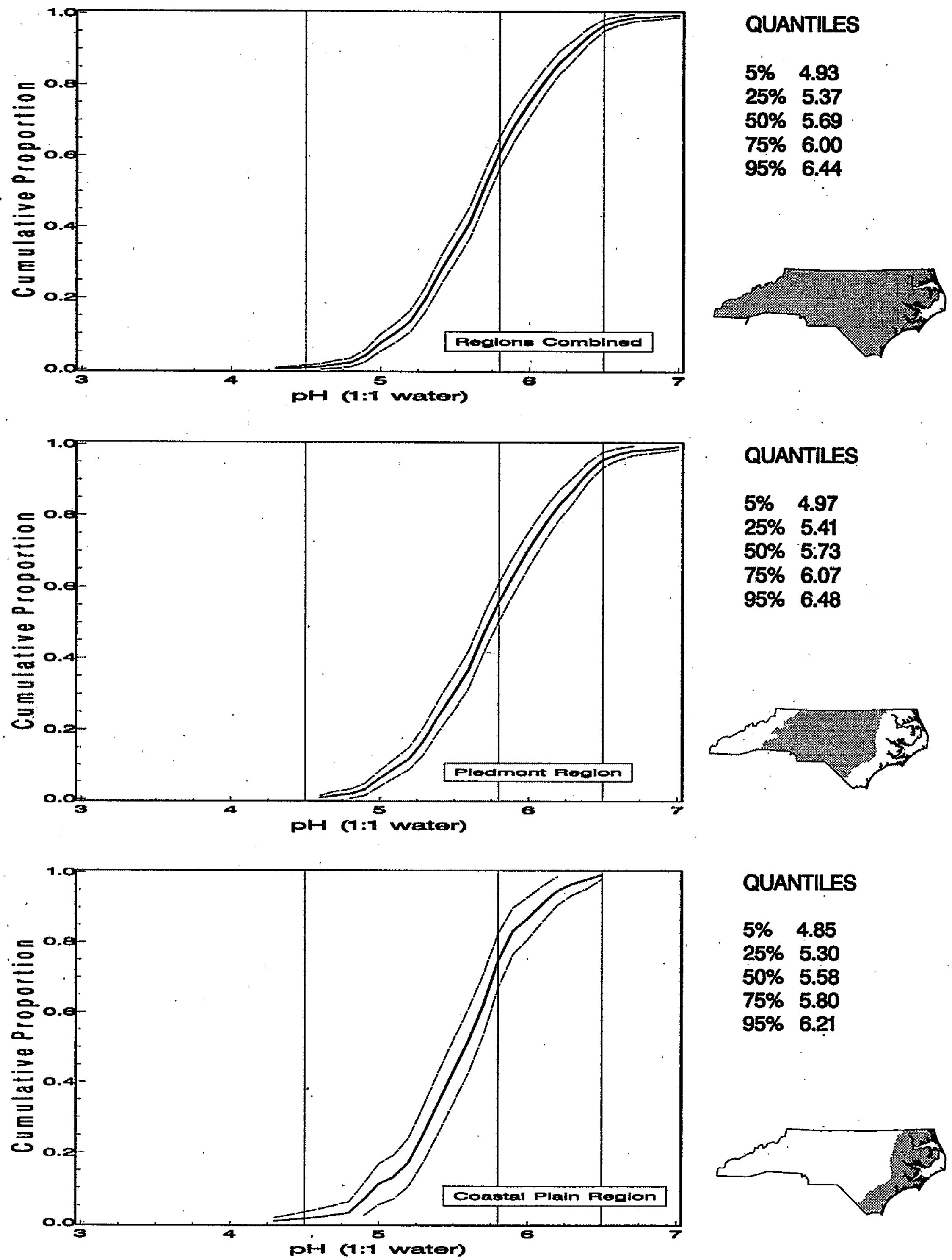


Figure 2.13 Cumulative distribution functions for pH (in water) for AP horizon. Vertical lines on the CDF represent boundaries for unacceptable (<4.5)and acceptable (>6.5) condition.

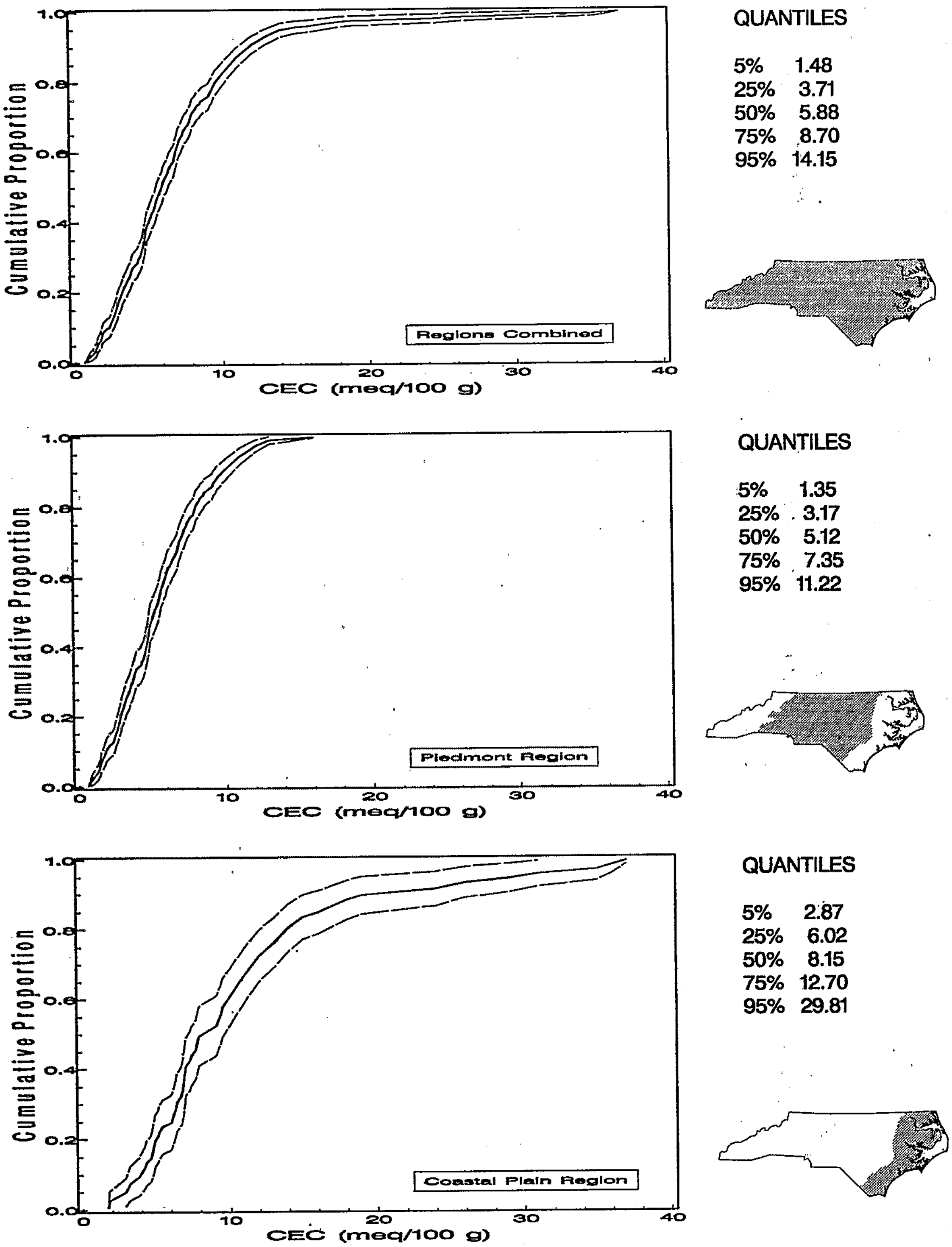


Figure 2.14 Cumulative distribution functions for cation exchange capacity (CEC) in AP horizon. Cumulative proportions refer to the land area cropped with annually harvested herbaceous crops.

2.3.2.3 Available phosphorus

Do the soils have adequate plant available phosphorus to sustain plant growth?

Phosphorus is a macronutrient that directly affects plant growth and health and is thus a common component in fertilization programs on any given field. In North Carolina, the soil phosphorus levels for Coastal Plain soils (using the Bray method with dilute hydrochloric acid and sulfuric acid extraction method) are classified as follows: <10 ppm, low; 11-31 ppm, medium; 31-56 ppm, high; > 56 ppm, very high (Olsen and Sommers, 1982). According to this scheme, more than 50% of the samples had very high phosphorus levels (Figure 2.15). Data are being re-evaluated to identify if the classification scheme cited by Olsen and Sommers is appropriate for the Bray analysis used on the Pilot samples.

2.3.2.4 Lead and cadmium

Do the soils have lead or cadmium levels that pose health risks to the ecosystem?

Lead and cadmium occur naturally in soils, but in excess quantities may threaten the health of the ecosystem. Soils that receive applications of municipal sludge are candidates for excess lead and cadmium levels, and soils near roads or some factories may contain high lead levels due to automobile traffic and emissions.

Average concentrations for lead in soils range from 15 to 25 ppm (Burau, 1982). A small proportion of the Piedmont samples contained relatively high levels of lead (> 40 ppm, Figure 2.16). Some of the Piedmont soils also contained relatively high levels of cadmium (Figure 2.17).

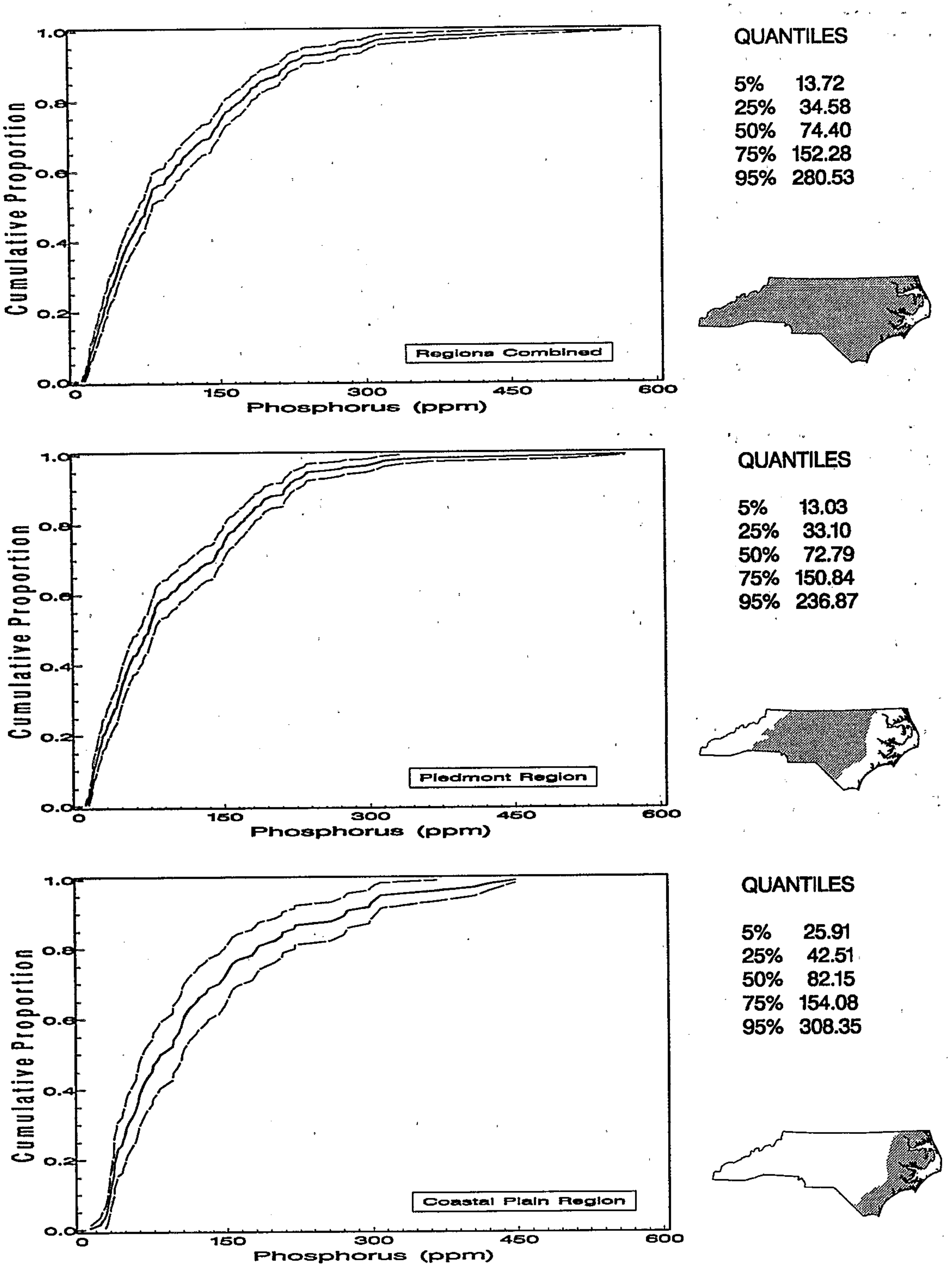


Figure 2.15 Cumulative distribution functions for available phosphorus in AP horizon. Cumulative proportions refer to the land area cropped with annually harvested herbaceous crops.

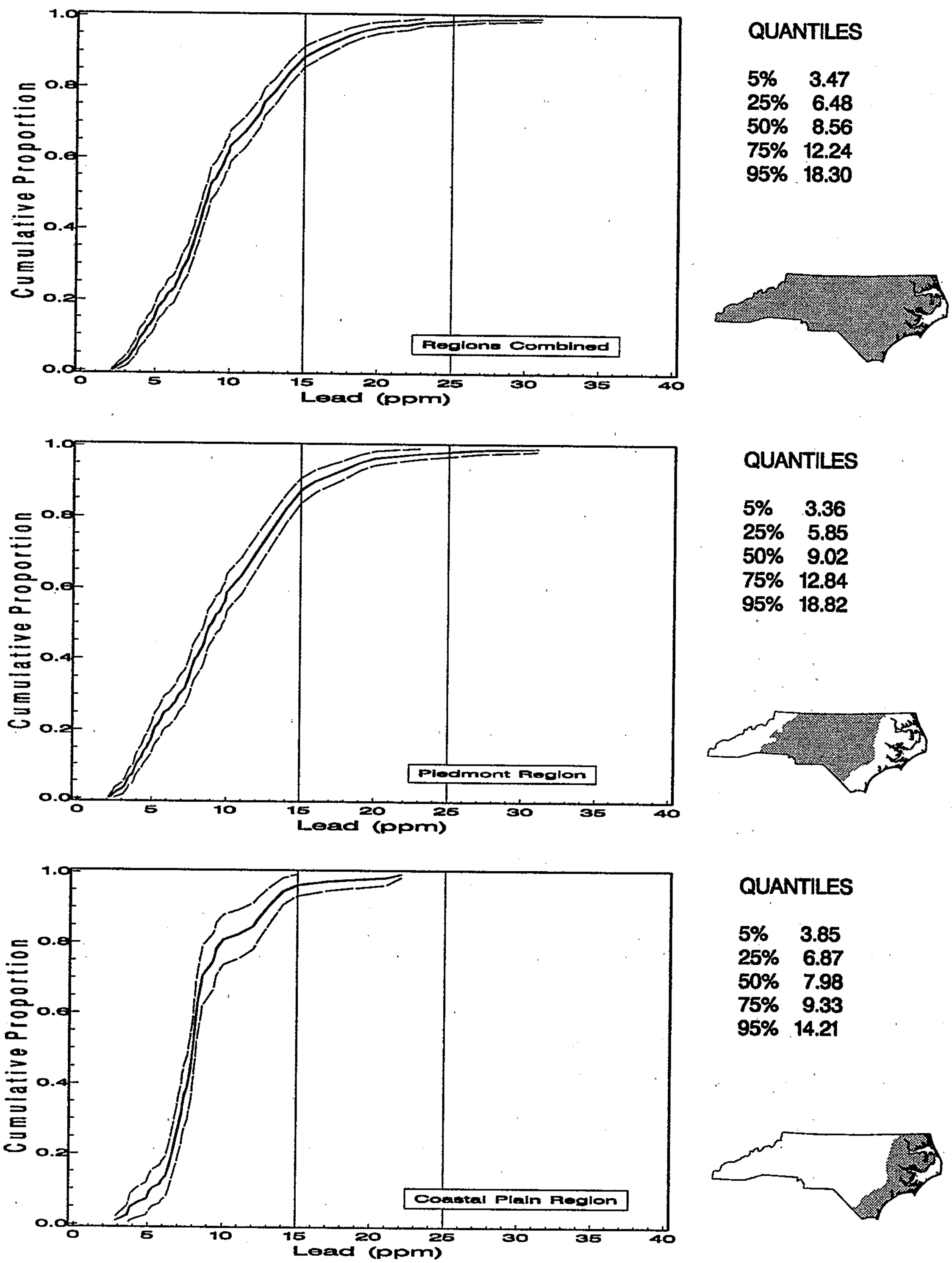


Figure 2.16 Cumulative distribution functions for total lead in AP horizon. Vertical lines on the CDFs represent a range of lead concentrations in soils (Burau, 1982).

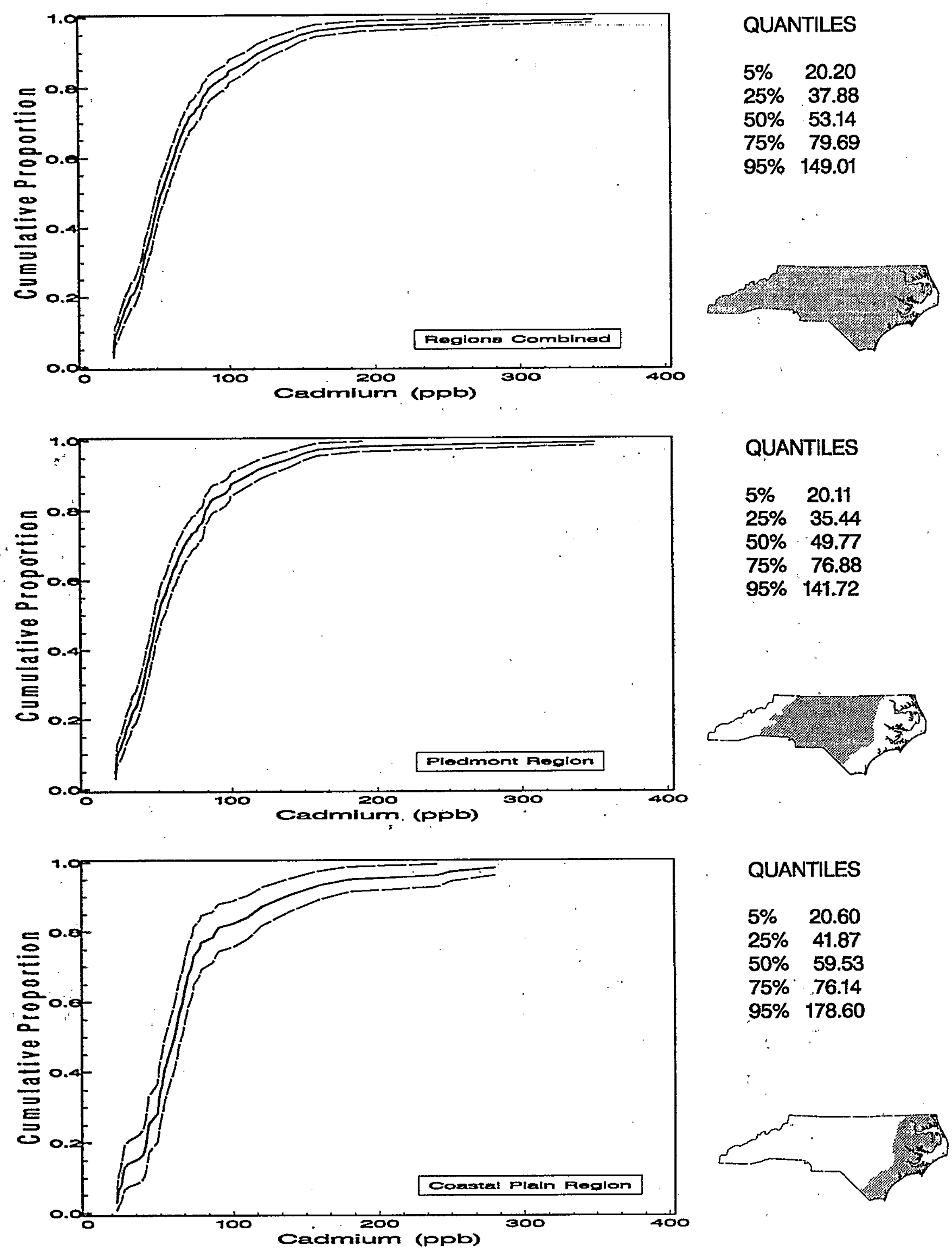


Figure 2.17 Cumulative distribution functions for total cadmium in AP horizon. Cumulative proportions refer to the land area cropped with annually harvested herbaceous crops.

2.4 SOIL BIOTIC DIVERSITY

Biotic communities within soil are responsible for the decomposition of organic matter, involved in many aspects of nutrient cycling, provide mechanisms for the development and maintenance of pore spaces through which gases and water flow, and interact in beneficial and detrimental ways with plant roots. Although soil biotic communities are influenced in many ways by natural elements of the soil physical structure and chemical composition, anthropogenic influences, such as tillage practices, cultural practices, and contaminants, also have profound effects on soil biotic communities. The status of the soil biota is, thus, vital to the overall characterization of soil condition or health.

Nematodes are found in most soils throughout the world. They are diverse in their feeding habits and occur as central members of the soil food web. Nematodes can be classified as bacterivores, fungivores, omnivores, plant-parasites, and predators. Predaceous nematodes occur higher in the trophic hierarchy than bacterivores and fungivores and are more sensitive to disturbance than the bacterivores and fungivores. Because of the range of feeding habits, trophic structure of nematode communites reflects the relative degree of stability or disturbance of the general biota in soils.

What is the degree of stability or maturity in nematode communities in soil?

A highly stable or mature community of nematodes would indicate that minimal disturbance or contamination has occured in the soil. Maturity indices for nematode community structure depend on whether particular nematode families are colonists (generalists or *r*-strategists) or persisters (specialists or *k*-strategists) (Bongers, 1990; Ricklefs, 1990), and therefore, reflect relative disturbance. Disturbances to the soil biota include cultivation, applications of agricultural chemicals, and soil compaction. A lower value for the maturity index for free-living or plant-parasitic nematodes indicates more disturbance, i.e., mostly colonizers and few persisters (Bongers, 1990; Neher and Campbell, 1994; Yeates, 1994).

For land cultivated in AHHC in the Piedmont and Coastal Plains regions of North Carolina (Figure 2.18 and 2.19), only a small proportion of communities of soil nematodes had a relatively high degree of stability or maturity. The proportion of land area cultivated with AHHC in North Carolina having values for the maturity index for free-living nematodes of less than 3.2 on a 1-5 scale (where 5 is the most mature or stable community) was estimated to be 0.95 (Figure 2.18). Although the median value for the maturity index for free-living nematodes was slightly greater for the land area cultivated with AHHC in the Coastal Plain (2.47) than for Piedmont (2.28), the reverse was true for the maturity index for plant-parasitic nematodes (2.48 vs. 2.74; Figure 2.19).

A higher proportion of larger values (>3.0) was found in the Piedmont than in the Coastal Plains region for both the free-living and plant-parasitic nematodes. Variability of maturity index values was much greater for land area cultivated with AHHC in the Coastal Plain than in the Piedmont region due to smaller sample size.

What is the degree of diversity of nematode communites in soil?

A diversity index such as the Shannon index of trophic diversity (Shannon and Weaver, 1949) describes the relative abundance and evenness of the distribution of nematodes across trophic or food-preference groups. In agricultural soils, higher diversity of trophic groups is correlated with an increase in the less abundant trophic groups, i.e. fungivores, omnivores, and predators, relative to the more generally abundant trophic groups, i.e., bacterivores and plant-parasitic nematodes.

Trophic diversity (Figure 2.20) values had a greater range for land area cultivated with AHHC in the Piedmont than in the Coastal Plain with more higher index values occurring in the Piedmont. Variability of the values for the trophic diversity index was also greater (i.e., wider 90% confidence band about the CDF) for the land area cultivated with AHHC in the Coastal Plain region due to smaller sample size than that of the Piedmont region.

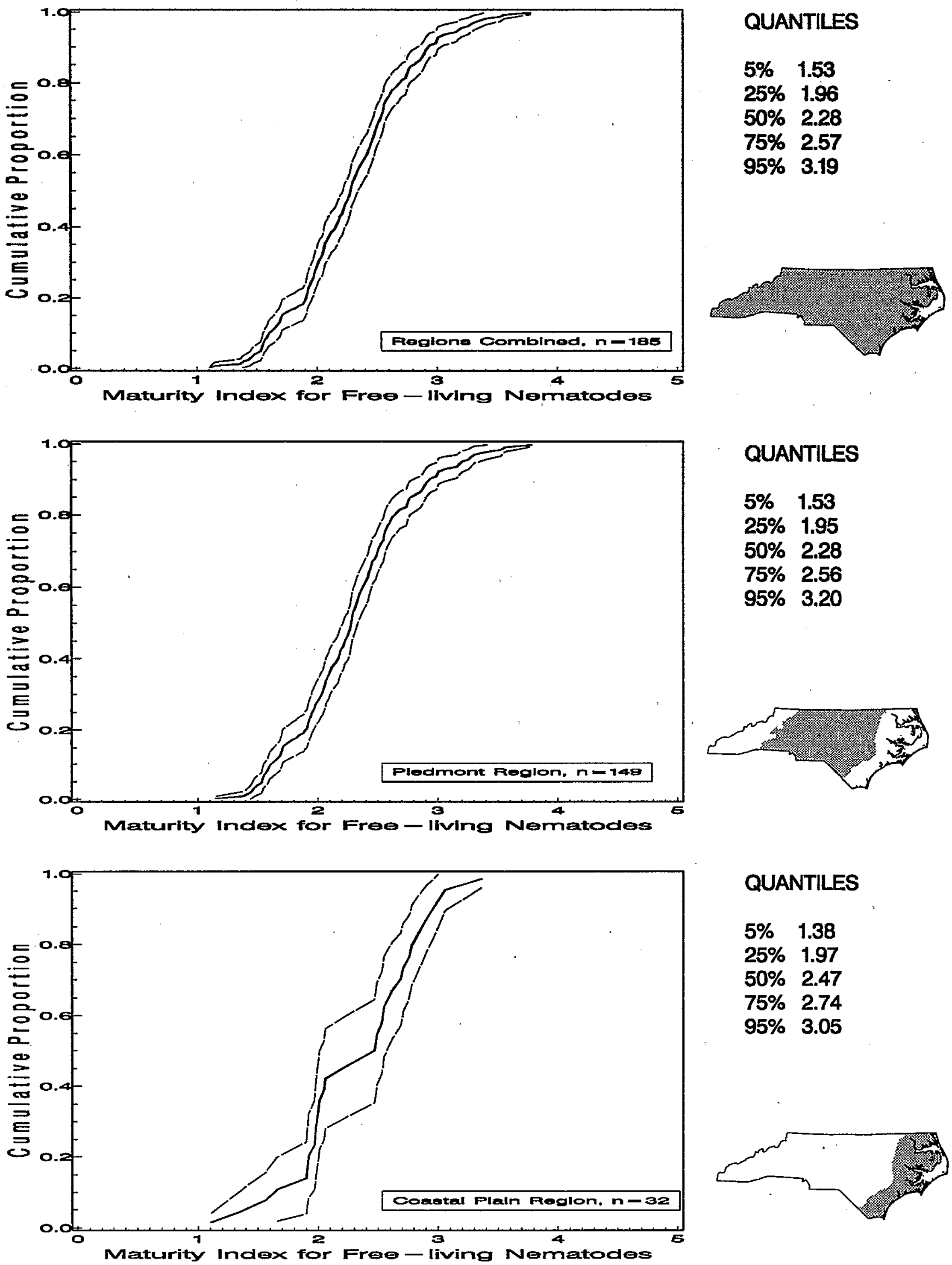


Figure 2.18 Cumulative distribution functions for maturity index for free-living nematodes in AP horizon of soil. Cumulative proportion refers to land area in annually harvested herbaceous crops.

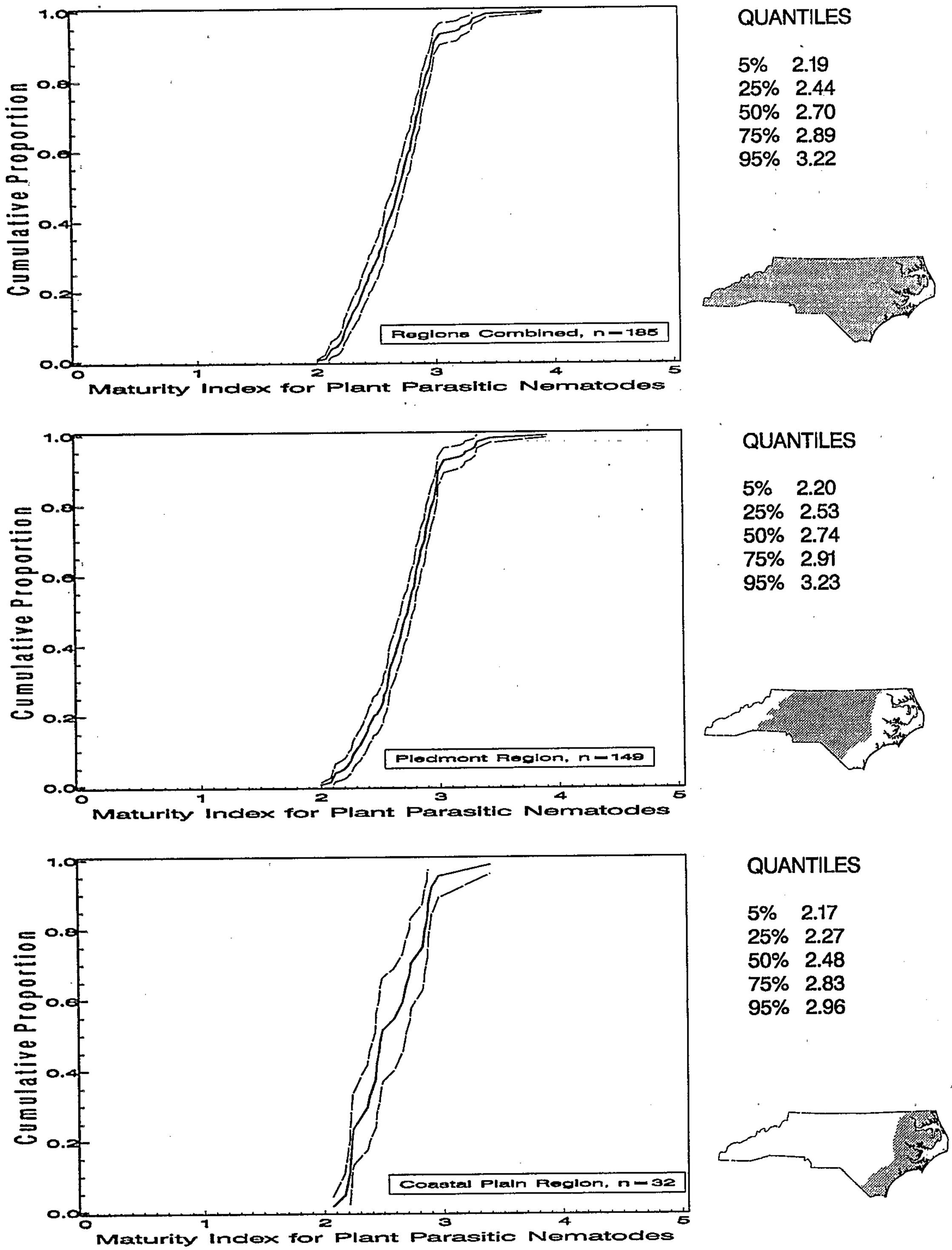


Figure 2.19 Cumulative distribution functions for maturity index for plant parasitic nematodes in AP horizon of soil. Cumulative proportion refers to land area cropped with annually harvested herbaceous crops.

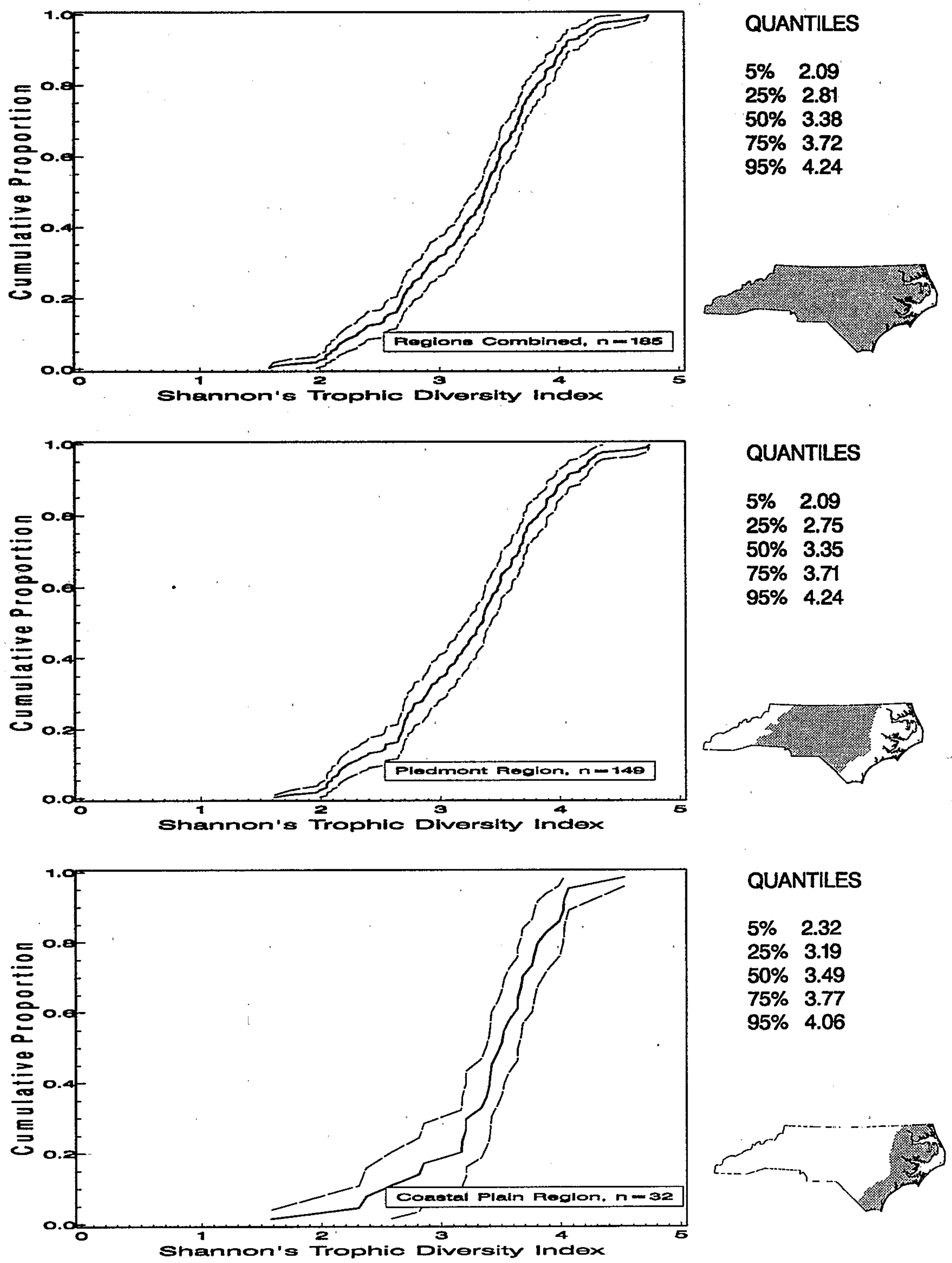


Figure 2.20 Cumulative distribution functions for Shannon's trophic diversity index for nematodes in AP horizon of soil. Cumulative proportion refers to land area cropped with annually harvested herbaceous crops.

2.5 WATER QUALITY

Water is essential to any agroecosystem, and it carries materials from the agroecosystem to the larger landscape. Two key places where water may be found are in ponds and underground. The 1992 Pilot Field Program sampled these water sources for two types of contaminants: nitrate and pesticides.

What is the distribution of nitrate concentrations in wells on farms in North Carolina? Specifically, what percentage of these wells have nitrate levels above the U.S. EPA maximum contaminant level (MCL) of 10 ppm nitrate-N?

Manure and fertilizer applications can result in leaching of nitrate into groundwater, where it is a potential health hazard. For example, nitrate can cause methemoglobinemia in infants, which is why the 10 ppm nitrate-N MCL was established (Bouchard et al., 1992; Fedkiw, 1991). Sixty-one wells were sampled across the state (Table 1.5). Although it was not a survey requirement that the wells be used for drinking water, about 90% of them were in that category. Samples were not taken of any treated well water. The detection limit was 0.20 parts per million of nitrate-N. Samples with detectable nitrate concentrations less than that value were assigned values of 0.20 ppm.

From the distribution of nitrate-N concentrations in the sampled wells (Figure 2.21) an estimated 3.5% of wells had greater than 10 ppm nitrate-N. Using 90% confidence limits, the upper bound of that estimate is 7.2% of wells and the lower bound is 0%. This result agrees with an unpublished study finding 3.2% of 9000 wells in North Carolina to have such nitrate concentrations (cited in Spalding and Exner, 1993.) The maximum concentration detected was 19 ppm nitrate-N. One of the three wells found to have greater than 10ppm nitrate-N was a hand-dug, bucket-type well.

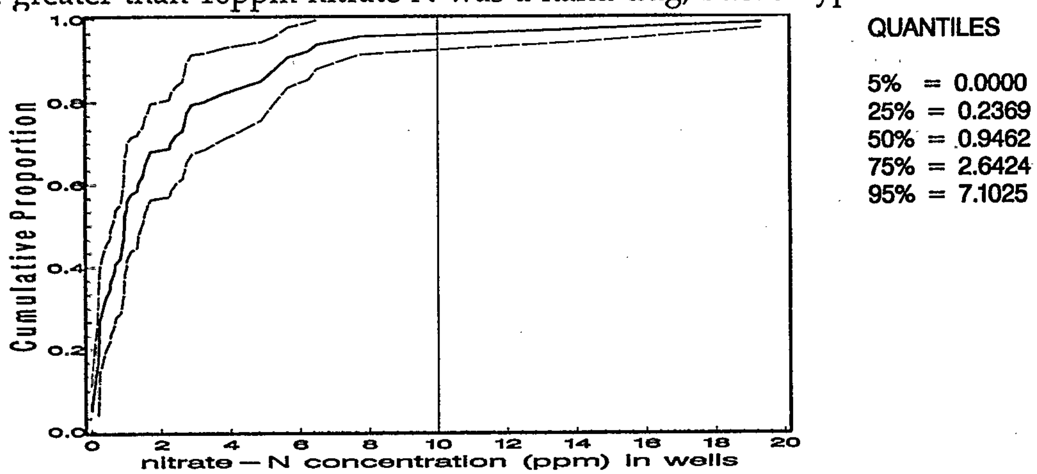


Figure 2.21 Cumulative distribution function for nitrate-N concentrations in wells on North Carolina farms, Fall 1992. Vertical line shows the EPA maximum contaminant level for drinking water.

What is the distribution of nitrate concentrations in farm ponds in North Carolina?

Nitrogen can be a nutrient that favors growth of green algae, but it is only one of the important elements (phosphorus is another). The nitrate concentration values by themselves have little ecological significance (Burkholder, personal communication), though levels of nitrate-N over 40-100 ppm may be harmful to livestock (Fedkiw, 1991).

Nitrate-N concentrations in pond samples (n=40) are illustrated by the CDF in Figure 2.22. Only the samples taken from the boat are included. Where replicate samples were taken in a pond, only the first was used. Concentrations were generally less than those found in wells (note the difference in the scales on the horizontal axes) with a median concentration in ponds of 0.2 ppm, about five times lower than the median value for wells.

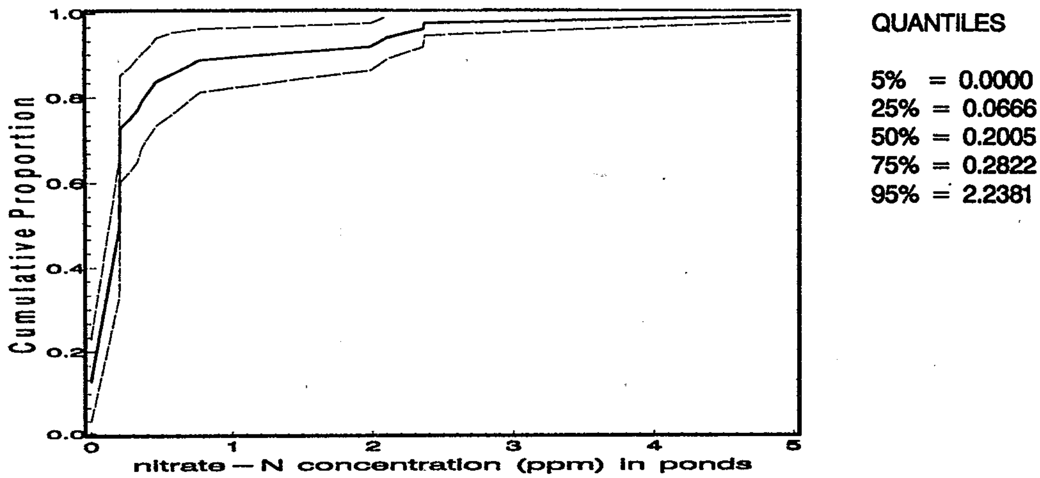


Figure 2.22 Cumulative distribution function for nitrate-N concentrations in samples from ponds on North Carolina farms, Fall 1992.

How extensive is pesticide contamination of farm ponds and wells in North Carolina?

There is concern about field-applied pesticides reaching ground and surface waters, mostly because of their possible impact on human health. Water samples were tested for the following pesticides, and none was detected at concentrations above one part per billion (1 ppb). Carbamates (aldicarb, carbofuran, methomyl, oxamyl) were analyzed by high performance liquid chromatography, the others by gas chromatography.

- Insecticides/nematicides:
  aldicarb, carbofuran, chlorpyrifos, endosulfan, methomyl, oxamyl

- Herbicides
  alachlor, atrazine, cyanazine, linuron, propachlor, simazine

Samples were taken in the fall, and it is possible that pesticides were present earlier in the year, closer to the time when such chemicals are widely applied to agricultural fields.

Chromatograms from a few of the samples showed the presence of unidentified compounds not on the above list, but this was not pursued further.

2.6 AGRICHEMICAL USE

How much stress are we putting on the ecosystems of North Carolina by applying chemical fertilizers and pesticides?

Next to human health, this question is at the heart of people's concerns about agricultural chemical use. Unfortunately, it is too broad to be addressed by simply finding out how much of which chemicals were used on North Carolina fields, though such data may be useful from four perspectives:

(1) As good or bad in themselves. According to one school of thought, sustainable agriculture must rely as little as possible on off-farm inputs.

(2) As stressors to agroecosystems. Data on fertilizer and pesticide use can be used in association with other indicators measured on agroecosystems. For example, nematicide use will eventually be an important variable for association with the nematode-based indices of soil biological condition.

(3) As variables against which to normalize. Fertilizer inputs will affect crop growth, and should be taken into consideration when designing crop productivity indices. An example of this is the nitrogen efficiency index shown in Section 2.2 of this report.

(4) As stressors to adjoining ecosystems and to ground water. There is a strong interest in identifying agriculture as a stressor to other systems. Such concerns may be justified, but that sort of assessment is beyond the scope of this report. Knowing what chemicals were applied is only a first step.

Recognizing these limitations, the use of four materials is summarized in Table 2.6.

- atrazine: an old and popular herbicide that has received recent publicity because of concerns about human health effects. It has become a water quality issue, and resistance to atrazine has developed in a number of weed species.

- Bt *Bacillus thuringiensis*, a bacterium used as a biological control agent for insect pests. Its use may be an indication that the farmer is trying to rely less on synthetic pesticides.

- carbofuran: a highly toxic insecticide/nematicide o nitrogen: an important plant nutrient, applied to fields in large quantities, that can leach to ground water. Only nitrogen from commercial fertilizers is shown in the table.

Table 2.6 Estimated use of fertilizer nitrogen and selected pesticides on annually harvested herbaceous cropland in North Carolina, 1992. Standard error is in parentheses. Units are kg/ha, except for Bt which is in billion international units of potency per hectare.

Material	Area (ha) of AHHCs treated at least once	Proportion of area of AHHCs treated at least once	Average number of treatments per year	Average rate[a]	Sample Number[b]
atrazine	299,000 (41,000)	0.178 (0.022)	1.06 (0.03)	1.65 (0.12)	54
Bacillus thuringiensis (Bt)	13,000 (7,000)	0.008 (0.004)	1.00 (--[c])	9.88 (--[c])	3
carbofuran	36,000 (14,000)	0.021 (0.008)	1.00 (--[c])	1.18 (0.08)	7
fertilizer nitrogen	1,269,000 (76,000)	0.755 (0.024)	1.72 (0.06)	84.3 (4.3)	242

[a]Average rates are active ingredient *per treatment*, except in the case of nitrogen which is *per crop year.*

[b]Number of sample units on which there were applications of the material.
[c]No standard error exists because all data values used to compute the mean were equal.

All estimates in the above table were derived solely from the Fall 1992 questionnaire, except that the areas treated were calculated by multiplying the proportion of area treated by the estimate of the total area of annually harvested herbaceous crops. This estimate was made using the June Enumerative Survey data.

Interestingly, the pesticides were each found on a specific crop in our sample. All applications of atrazine on sample fields were to corn (grain or silage) or popcorn. All of the fields with recorded carbofuran applications were either corn for grain or corn for silage. All of the Bt applications were to tobacco.

As expected, nitrogen applications dominate the picture, both in terms of area treated and the amount applied. This lends further support to the importance of nitrogen efficiency (Section 2.2).

Atrazine was used on a fairly substantial portion (almost 18%) of the area under annual crops. The area to which carbofuran and Bt were applied is much smaller, so small that the estimates associated with these materials have rather large standard errors.

3. EVALUATION OF DESIGNS, INDICATORS, AND ACTIVITIES

3.1 PRELIMINARY DESIGN COMPARISON

Design comparisons were performed primarily for fall questionnaire and soil sample data. Land-use indicators took advantage of the full June Enumerative Survey dataset in order to get the most precise estimates possible. Other indicators, including soil quality-physical/chemical, soil biotic diversity, and crop productivity indices, were only calculated from the fall questionnaire and soil sample data. The design comparison, then, is most relevant with respect to the these indicators.

Only statewide estimates were used and the comparison was done both accounting for, and ignoring, the stratification (Cotter and Nealon, 1987) of the Rotational Panel plan. Cost and precision have been considered in the evaluation of the relative efficiencies of the two sampling plans.

Typically, the efficiency of one design relative to another is defined as the ratio of the variances for the statistic of interest, adjusted for different sample sizes when necessary. For example, the relative efficiency of design one to design two is defined as the variance under design two divided by the variance under design one with respect to the parameter of interest. In EMAP the parameter of interest is the cumulative distribution function (CDF). A method which evaluates the overall precision of the CDF was used for comparing the two sampling plans for chemical measures of soil quality and crop productivity indicators. Nematode indices could not be used because nematode data were collected only with the Rotational Panel plan. Agrichemical indices are not included since they are not summarized with CDFs in this report. Crop specific indices of crop productivity are not presented because of relatively small sample sizes. For the relevant fall questionnaire and soil sample indicators, a CDF was computed under each plan. B_u was defined as:

$$B_u = \left(\sum_{t_1}^{t_{40}} Var_{rp}(t)\right) / \left(\sum_{t_1}^{t_{40}} Var_h(t)\right)$$

where $Var_h(t)$ is the estimated variance at t for the CDF under the Hexagon plan and $Var_{rp}(t)$ is the estimated variance under the Rotational Panel plan. The sums are over equally spaced increments ranging from

t_1 =max(min(Hexagon dataset), min(Rotational Panel dataset)),

to

t_{40}=min(max(Hexagon dataset), max(Rotational Panel dataset)).

The limits of the sum were defined as such to prevent the relative efficiencies from being dominated by variances in the tails of the CDFs. The method was also tried for a 20-point 't' vector and for the combined data 't' vector; results were very similar. To adjust for differences in sample sizes, B_u was multiplied by (n_{rp}/n_h), i.e., the ratio of the sample sizes. This adjusted value is a measure of the relative efficiency of the Hexagon sampling plan to the Rotational Panel sampling plan, ignoring costs (Table 3.1).

Table 3.1 Relative efficiencies[a] of the Hexagon design to the Rotational Panel design.

Indicator	Ignoring Stratification: Relative Efficiency	Ignoring Stratification: Relative Efficiency w/ Cost Factored in	Accounting for Stratification: Relative Efficiency	Accounting for Stratification: Relative Efficiency w/ Cost Factored in
Soil Quality-Physical/Chemical				
Phosphorus	0.81	0.58	0.80	0.57
Cadmium	0.80	0.57	0.78	0.56
CEC	0.72	0.51	0.69	0.49
Clay	1.00	0.71	0.97	0.69
Lead	1.13	0.81	1.11	0.79
Organic Matter	0.78	0.56	0.77	0.55
Sand	1.04	0.74	1.02	0.73
pH	0.90	0.64	0.89	0.63
Crop Productivity				
Nitrogen Efficiency (using nitrogen from commercial fertilizer and manure)				
All Seed Crops	0.80	0.57	0.77	0.55
Observed/Expected Yield				
All Crops (all crops for which expected yields were available)	0.95	0.68	0.92	0.66

[a]A relative efficiency less than one implies that the Rotational Panel design is more efficient, a relative efficiency greater than one implies that the Hexagon design is more efficient.

Relative efficiencies greater than one indicate that the Hexagon design is more efficient, whereas relative efficiencies less than one indicate that the Rotational Panel design is more efficient. For example, consider soil chemistry when stratification is ignored in the variance calculations under the Rotational Panel design. For estimating phosphorus content in soil, the Hexagon design was estimated to be 81% as efficient as the Rotational Panel design. When costs were factored in (to be described below), the Hexagon design was estimated to be only 58% as efficient as the Rotational Panel design. For estimating lead content in soil, on the other hand, the Hexagon design was estimated to be 13% more efficient than the Rotational Panel design when costs were ignored but only about 81% as efficient when costs were accounted for. Results are nearly the same when variance calculations accounted for stratification.

The adjusted relative efficiency can be thought of as the information per sample unit of the Hexagon plan relative to the Rotational Panel plan, where information is defined as the inverse of the average variance. To factor in costs, total costs, excluding salaries of the Agroecosystem Resource Group staff, were documented for each plan using records kept by NASS (Table 3.2). The Hexagon sampling plan required NASS enumerators to visit segments in June that they normally would not visit and, hence, costs were more per sample unit.

Table 3.2. Costs of the Hexagon and Rotational Panel sampling plans.[a, b]

Item	Hexagon	Rotational Panel
Constructing segments (Prorated over 20 years)	$ 586	0
June Enumerative Survey	$ 17,289	$ 8,025
SUBTOTAL	$ 17,875	$ 8,025
Number of segments	51	65
Cost per segment	$ 350	$ 123
Fall survey costs[c]		
NASS labor & administration	$ 13,005	$ 16,575
Equipment, shipping & lab analyses	$ 14,388	$ 19,323
TOTAL	$ 45,268	$ 43,923
Number fall questionnaires listed as complete	138	188
Cost per fall survey sample unit	$ 328	$ 234

[a]Since nematode data were only collected from the Rotational Panel plan, nematode costs are not included.
[b]Table entries are rounded.
[c]Cost differences for the Fall Survey reflect only differences in sample sizes.

The relative cost per sample unit of the Hexagon Plan ($328.00) to the Rotational Panel plan ($234) was calculated as 328/234=1.40. The relative information per unit was divided by the relative cost per unit to obtain the information per dollar of the Hexagon Plan relative to the Rotational Panel Plan. This index is a measure of the relative efficiency accounting for differences in costs (Table 3.1).

Accounting for stratification had little effect on the relative efficiencies. This is not surprising because NASS's stratification is designed to improve estimates of extent from segment level data and not specifically designed to improve estimates of field level data, e.g., fall questionnaire and soil sample indices. When costs were ignored for soil quality, the estimated relative efficiencies seemed to favor the Rotational Panel design for five of the eight variables and seemed to favor the Hexagon design for only lead content. The soil chemistry measures were, however, not all independent. The relative efficiencies, ignoring costs, for the crop productivity indicators nitrogen efficiency and observed/expected yield indices indicated that the Rotational Panel design was at least as efficient as the Hexagon design for both indices. When costs are factored in, the Rotational Panel design appears to be more efficient for all fall questionnaire and soil sample indices examined.

Entries in Table 3.1 are estimated relative efficiencies and, therefore are subject to sampling variability. Assessing the variability in the estimated relative efficiencies is difficult and we made no attempt to do so. Two conclusions can be reliably drawn from Table 3.1. First, the Rotational Panel design is no less efficient than the Hexagon design. Second, the Rotational Panel design is the more cost-effective design.

In addition to the above comparisons for fall data, the two sampling plans were compared with respect to the estimate of extent of annually harvested herbaceous crops from the June data (Table 3.3). Relative efficiencies in Table 3.3 were derived from the estimated variances of the estimates. NASS samples high-agriculture strata with greater intensity than they sample low-agriculture strata and, thus, segments in the Rotational Panel sample tend to be more agricultural than do segments in the Hexagon sample. The impact of NASS's stratification is evident in Table 3.3 where the Rotational Panel Plan appears to be considerably more efficient than the Hexagon Plan in the estimation of extent of annually harvested herbaceous crops.

Table 3.3 Relative efficiency[a] of the Hexagon design to Rotational Panel design with respect to estimation of extent of annually harvested herbaceous crops.

	Ignoring Stratification[b]		Accounting for Stratification[b]	
Indicator	Relative Efficiency	Relative Efficiency w/ Cost Factored in	Relative Efficiency	Relative Efficiency w/ Cost Factored in
Extent of AHHC	0.63	0.22	0.43	0.15

[a]A relative efficiency less than one implies that the Rotational Panel design is more efficient, a relative efficiency greater than one implies that the Hexagon design is more efficient.
[b]Refers to whether stratification was accounted for in variance computations.

Both designs provided good spatial coverage of North Carolina with the 51 Hexagon segments falling in 49 counties and the 65 Rotational Panel segments distributed over 55 counties. Prior to drawing the sample for the fall survey, the lists of acres were ordered first by crop and then by segment number. The ordering by crops diminished the spatial coverage of the two sampling plans for the fall questionnaire and soil sample. Because segments were numbered according to location, and the lists are sampled systematically, ordering first by segment number should help maintain the spatial coverage. For the 1993 Nebraska Pilot the annually harvested herbaceous cropland acres will be ordered only by segment number and the spatial coverage will be preserved better.

In summary, when differences in costs were considered, the Rotational Panel design appeared to be more efficient than the Hexagon design. If costs are ignored, the Rotational Panel design tends to perform as well or better than the Hexagon design for most fall questionnaire and soil sample indices, with the only exception being lead content of soil. For estimation of extent from the June Survey, the Rotational Panel design was more efficient than the Hexagon design.

3.2 SUCCESSES AND CHALLENGES IN INDICATOR DEVELOPMENT AND EVALUATION

3.2.1 Crop productivity

At least three important questions surround the development of a crop productivity indicator. These relate to combining data across crops, indicator performance, and conceptual and operational challenges. Something about each of these was learned in the 1992 Pilot.

How can data from different crops be brought together in a single productivity index?

In order to make broad regional statements, it would be useful to be able to "add apples and oranges" into one overall index of productivity. A simple method of standardizing yield was attempted with the data from the 1992 Pilot.

Ideally, standardization of any index should rely on a long-term baseline of data. As a test, county average yields were used. These were obtained from NASS for the period 1980-1990. The index summarized in Figures 3.1 and 3.2 is calculated by dividing the reported yield for each sample unit by this 11 year county average. The composite graphs in Figure 3.1 include most but not all AHHCs: barley, field corn (maize), cotton, hay, oat, peanut, grain sorghum, soybean, sweet potato, tobacco, and wheat. Data were available to calculate this index for 77% of the valid sample units, much more than the 49% for the combined nitrogen index for seed crops (Section 2.2).

The median values of all the illustrated observed/expected indices are greater than 1.0, indicating that technology or perhaps weather has caused the yields as a whole to be greater than they were during the 11-year reference period. This index presented for a single year has little meaning in itself, except to allow many species to be plotted on the same graph. Over the long term, however, this type of index should reflect changes and trends in overall crop productivity.

It is interesting to look at the pieces that go into the overall index. The CDFs for the observed/expected index for three individual crops (Figure 3.2) are quite different from each other. The median index value for land in soybeans is 1.04, indicating a near-average situation. Land in corn showed greater productivity than during the baseline years, and almost all of the distribution for hay lies to the right of the 1.0 line.

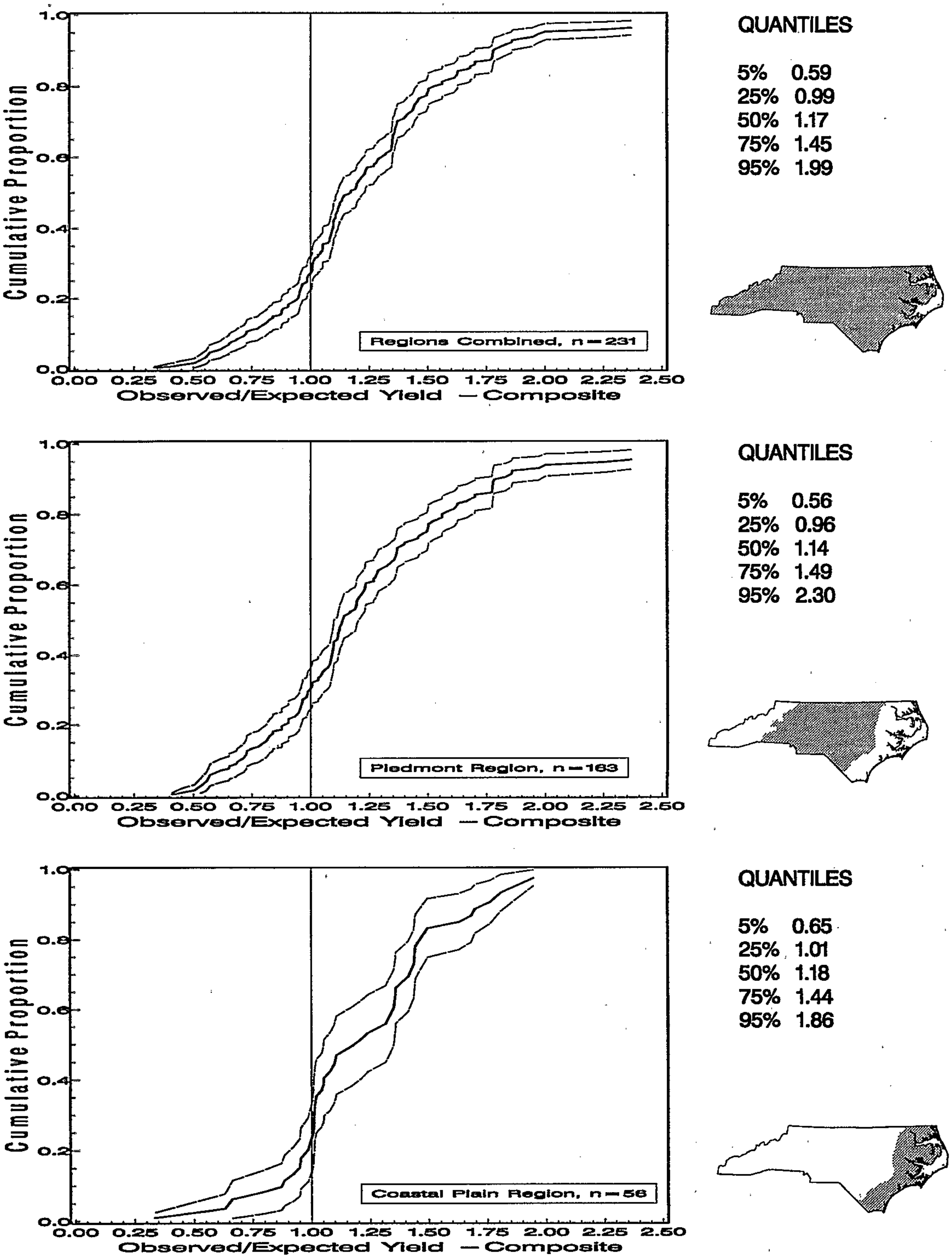
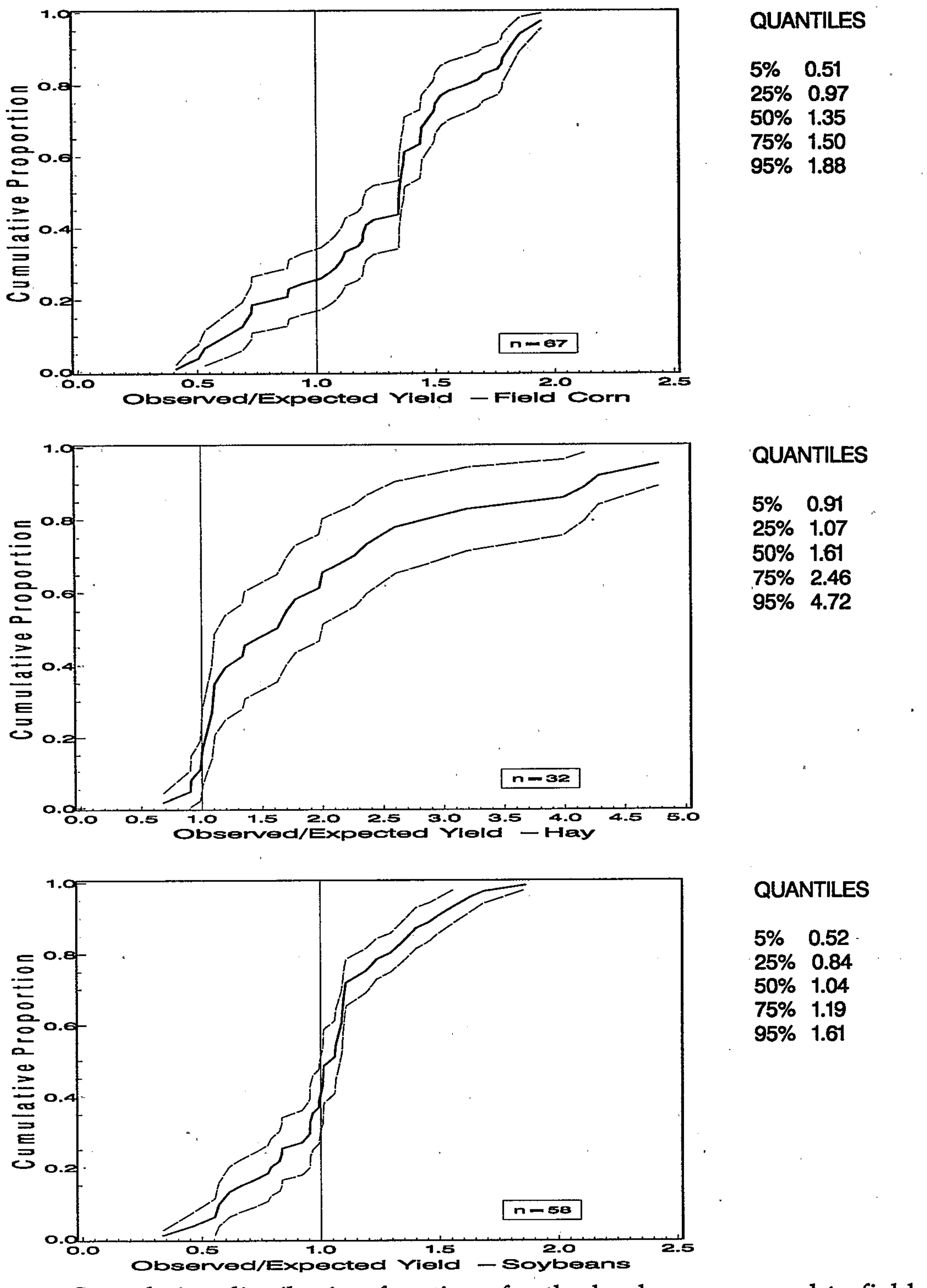


Figure 3.1 Cumulative distribution functions for the observed/expected yield index, composite for land area cropped to barley, field corn (maize), cotton, hay, oat, peanut, grain sorghum, soybean, sweet potato, tobacco, and wheat.



Figures 3.2 Cumulative distribution functions for the land area cropped to field corn (maize), hay, and soybean, respectively. A vertical reference line shows where the index value is 1.0.

How did the two crop productivity indices perform?

Variability is the only performance criterion that has been examined so far. In order to detect trends it is desirable that within-year variance and random year-to-year fluctuation be small compared to the trend "signal". Low noise would also be helpful for making comparisons among regions. Tables 3.4 and 3.5 show the coefficients of variation (CV) for the nitrogen efficiency index and the observed/expected ratios, as compared to the CVs for the corresponding unadjusted yields. Number of sample units differ between the two tables; see "Data completeness and quality" below.

Table 3.4 Comparisons of coefficients of variation (=100*(σ/μ)) for the nitrogen efficiency index and unadjusted yields (weight per unit area). Data for the entire state of North Carolina. Nitrogen calculations based on manure plus commercial sources.

Category	CV for unadjusted yield	CV for nitrogen efficiency index	No. of sample units
eight seed crops, combined data	58	100	146
corn for grain	37	80	60
wheat	21	62	31
soybean	34	139	57
cotton	22	89	15

Table 3.5 Comparisons of coefficients of variation (=100*(σ/μ)) for unadjusted yields and for the observed/expected yield index. Data presented are for the entire state of North Carolina.

Category	CV for unadjusted yield	CV for observed/ expected index	No. of sample units
all crops with available data	83	46	231
corn for grain	36	33	67
soybean	34	31	58
hay	72	63	32

In every case, the nitrogen index shows a much greater CV than the corresponding unadjusted yield. This is disappointing, and further work is needed to find the best indicators of efficiency. The observed/expected ratio gives slight to moderate reductions in the coefficient of variability. The expected yields used in the denominator were county average yields for each crop, so the resulting index essentially corrects for crop species (a substantial effect) and location within the state.

What challenges were met, and what obstacles remain?

Data completeness and quality

The index values, especially the combined index for nitrogen efficiency, include data from only a portion of the sample. Sometimes sample units were unusable because questionnaire data were simply missing, for example when crops were not harvested by the time of the fall visit. In other cases there was a lack of conversion factors or reference yields, leading to the exclusion of otherwise usable data. Further research is needed in this area. The inclusion of nitrogen from manures is a special challenge (see below).

Many issues surround the data that go into the indices. The factors used to convert yields to dry weights and to estimate the nitrogen content of manures are of unknown accuracy. The survey data themselves are subject to a number of nonsampling errors.

Where yield=0, data could not be used in the nitrogen efficiency index (no division by zero). This may skew the distribution downward slightly. For the nitrogen use efficiency of seed crops, 24 such sample units were excluded. It is difficult to determine why those yields were 0, but in 14 of the 24 samples, the crop seemed not to have been intended for harvest (usually a cover crop). Crop failure is the apparent explanation for the other 10 sample units. To include those units in the calculation would be to ignore the fact that biological production was not actually zero. This index is not strictly biological, of course, since only the harvested portion of biomass is considered. When yield data were available for only one of two crops in a single year, that sample unit was included in the nitrogen efficiency index for the individual crop but excluded from the composite. Many other such judgements had to be made during index development.

Estimating the amount of nitrogen contributed by application of manure

The estimated nitrogen content of manures was assigned to the crop to which the material was applied, though realistically only about half will be available for plant uptake during the first year. No conversion factors have been obtained for cases where the animal waste was in the form of a slurry (n=4 sample units). Sample units that received sewage sludge applications in the last five years (n=3) were excluded from the calculation, since no details were known on the rate or time of the sludge application. If a cover crop received manure or commercial fertilizer, the nitrogen will be credited to the next harvested crop, though this situation did not occur for the seed crops illustrated in this report.

An alternative nitrogen efficiency index can be calculated that includes only commercial sources of nitrogen, which depend heavily on nonrenewable resources. Fewer conversion factors are necessary to calculate this index. The resulting CDF is not illustrated, but it is nearly identical to Figure 2.9 for the state, because so few sample units received manure applications.

Classifying index values as good or bad

A major limitation at this stage is that no acceptable/unacceptable interpretation can be given for the crop productivity indices. Such cutoff points must either be defined or the index will have to be used only for the difficult job of trend detection.

Future directions for indices of crop productivity

Nitrogen use efficiency and standardized yield are only two of many possibilities for evaluating productivity in agroecosystems. It is suggested by Lal (1991) that the most limiting or nonrenewable input be used as a part of a larger index to quantify sustainability. In this spirit, water use efficiency may be included when the ARG begins work in the western states. Regression approaches also may be tried.

The use of a simple long-term average in the observed/expected index is only for testing purposes. Variants of it may be used in associative studies. It is desirable to have expectations that are somehow tied to soil map unit as well as location, but this has not yet been accomplished because the expected yield data available from the Soil Conservation Service are not all from the same time period. An adjustment for time trends will be needed before the data are useful.

In addition to soil and management, the key variable driving productivity is weather. Neither of the indices shown here accounts for variability due to temperature, precipitation, etc., and efforts are underway to use crop simulation models to do so.

3.2.2 Successes and Challenges in Assessing Soil Quality

We have met with a number of challenges in attempting to define and assess soil quality. Issues emerged related to societal values, integrating soil assessments with other indicator assessments, relating the measured values to agroecosystem condition, and identifying whether or not our values represented sustainable conditions or not. But our first task was to evaluate the soils data. In many ways, we haven't graduated beyond this step, although after accomplishing this, we have a much clearer direction of where we need to go.

Our first step was to make the task of data evaluation manageable, and we accomplished this by examining each of the distributions of the various parameters and developing CDFs for each. Then we assigned ranges of acceptable and unacceptable to each of the parameters in relation to crop production. These ranges were related to generic plant growth; for example, we identified the "ideal" ranges of clay content, pH, etc. We then grouped the parameters according to whether they related to soil physical, chemical, or biological condition. Soil physical condition was estimated by evaluating organic matter and clay content; soil chemical condition was determined by evaluating pH, cation exchange capacity, available phosphorus, and some trace elements; and soil biological condition was evaluated by evaluating the nematode communities. Framing our data in this way enabled us to pose fairly specific assessment questions, which in turn facilitated our data summaries.

We found this approach to be very useful for summarizing data about individual soil properties, but after all our summaries were complete, we noticed that we still could not draw very meaningful conclusions about soil quality *as related to agroecosystem condition.* For the soil physical properties, then, we combined the clay and organic matter data and redefined the proportion of agroecosystems in North Carolina that had "nominal" soil physical condition. While this approach was useful, it does not go far enough. We recognize that additional data integration needs to proceed. Data need to be integrated within indicators (such as clay and organic matter) but also need to be integrated across categories. We identified three deficiencies in our approach to the 1992 Pilot data. These are described below, along with potential approaches to addressing these issues in forthcoming studies.

3.2.2.1 The approach did not account for different plant requirements.

Data evaluation did not incorporate information about which plants were being grown in the soil or the specific requirements of the plant. Implicit in this approach is that there is an "ideal" soil for all types of annually harvested herbaceous plants. This approach required that the ranges for soil properties be sufficiently large to include the span of requirements for a wide range of crops, which included crops grown extensively such as wheat, corn, cotton and tobacco, and crops grown in relatively small areas such as watermelons and cucumbers. The data ranges could be narrowed if we took into account the specific soil requirements for each crop.

We possess a wide range of crop and management information for each of the fields from which our soil samples were drawn. We should identify the specific crop that is grown in each soil and assign a suitability value for each parameter that relates the crop requirements to the soil conditions. Using this approach would then enable us to evaluate whether the soil was being used for the type of crop for which it is best suited. Such information may be useful for linking soil parameters to sustainability or agroecosystem condition.

3.2.2.2 Agronomic potential and management information are not considered.

Our approach to data analysis does not consider soil potential or land management practices. Consequently, there is no way to determine whether a soil is in good or bad condition relative to its potential. Management practices are an integral factor of sustainability in agroecosystems.

Some soils have naturally high potential for crop and non-crop productivity, while others do not. It is not enough to merely measure soil properties and proceed with an assessment of sustainabilty. The link between soil properties and sustainability requires information about management factors. Examination of the management factors alone does not give us information about whether the management practices are producing desired results. Examination of the soil properties alone does not provide information about sustainability. Soil property evaluation may show a "good" quality soil in the midwest and a "marginal" soil in North Carolina. The reality may be that the deep, midwestern loessial soil may be in the process of degrading due to rill erosion or continuous cropping, whereas the North Carolina soil may be holding steady or improving due to responsible management. Future data analyses will integrate information about management practices and soil type to arrive at an indicator of sustainability.

3.2.2.3 Societal values were not integrated into the assessment.

The integrated assessment of clay and organic matter contents revealed some low clay content soils with unexpectedly high amounts of organic matter. A closer evaluation of these outliers revealed that these supposedly "nominal" or "optimal" soils were actually organic (muck) soils, probably drained wetlands. This finding challenged our soil rating system by touching the nerve of wetlands protection. In turn, this challenged our agricultural approach of evaluating soils and confirmed that we were really evaluating "crop-ability" of the soils, and not tying this evaluation to other societal values and issues. In order to resolve the issue of how to evaluate cropped wetlands, we may need to frame our questions in the context of the much-debated societal values. Since the societal values, as well as the indicators, are still evolving, we do not have a recommended procedure for addressing this elusive issue. At this point, we simply note that it is an issue and will continue to be aware of potential conflict of evaluating the "crop-ability" of soils without considering the broader issues of societal values.

3.2.2.4 Some successes were realized in simultaneously analyzing two soil properties.

For soil physical properties, we have met with some successes in jointly analyzing the clay and organic matter data. The integrated evalulation of clay and organic matter contents revealed that the two properties were inversely related. This leads us to conclude that it may be necessary to establish organic matter ranges for each range of clay class (0-18%, 18-35%, and over 35% clay). The deterministic relationship between these properties requires that we develop indices of soil quality that integrate a myriad of properties. The integration of clay and organic matter contents is an attempt to integrate factors to make statements about physical condition of the soil; now we need to attempt to integrate information about physical, chemical, and biological condition, as well as soil potential and management factors to make statements about the overall soil condition and sustainability. Our first step towards such a holistic assessment is framing the data in the context of the soil type. This will free us from the concept of an ideal soil, and will preclude us from concluding our investigations with a map of agroecosystem condition that looks remarkably like a Major Land Resource Areas map.

3.2.2.5 Suggestions for future efforts.

While we met with some successes in analyzing organic matter and clay contents together, we recognize that we still fall short in incorporating all of the factors that we need for a truly integrated assessment. Several soils with nominal organic matter, for example, had undesireable pH levels. It is clear that several factors need to be considered simulataneously, and it may not be appropriate to arrive at an assessment of soil physical condition in addition to soil chemical condition and soil biological condition. We may need to integrate all three of these into one finding. We will approach this integrated effort by utilizing the Soil Conservation Service's Soil Rating for Plant Growth (SRPG) model, developed by Joyce Scheyer at the National Cooperative Soil Survey. This model incorporates information about the surface properties of soil (which we could measure) and some profile information pertinent to the soil series (which we could obtain from the SCS databases). Results from this model will get us closer to a truly integrated assessment for plant productivity, yet we will still need to incorporate information about management practices and biological indices to approach an assessment of sustainabilty.

In addition to using the SRPG model, we will design our next sample collection around the use of some intact soil samples. In the 1992 Pilot, we utilized composited samples collected by NASS enumerators. These samples were compromised of 10 subsamples that were homogenized in the field then ground up in the laboratory prior to analysis. This approach to sample analysis precludes any evaluation of the soil tilth. While tilth is largely a qualitative and intuitive measure, it can be semi-quantified by measuring bulk density, pore size distribution, aggregate size and stability, mechanical impedance, organic matter content, and some measure of water movement (e.g., infiltration, permeability, or hydraulic conductivity). Most of these measurements require intact core samples rather than ground samples. We plan to use John Doran's (USDA-ARS) kit for collecting and analyzing field samples, which he developed specifically for assessing soil quality.

Other ideas that are being considered include additional measurements and analyses and different ways of analyzing the data, e.g., with the use of indices. These are briefly outlined below.

Measurements. We are currently looking at comparing bulk density of soil cores with reconstituted bulk density, which is determined in the laboratory by grinding, packing, and wetting the soil to create a controlled sample for comparison purposes. A ratio of reconstituted to field bulk density could be compared across all fields and soil types.

Ratios of cation exchange capacity to clay may also be evaluated in the different soil types to identify the theoretical maximum clay fraction of the soil material, and compare it with the measured clay fraction. Since CEC is also related to organic matter, we would expect to see relationships between the theoretical versus measured clay content and the organic matter content of the soil.

Further studies could evaluate the relationship between soil pH and macro and micronutrient levels in soils. Consideration should be given to native soil conditions (e.g., jarasite and sodium-affected soils) and to management practices (e.g., lime application, irrigation). A ratio of lime application to pH could be attempted to examine the efficiency of lime application, which would be related to sustainability of management practices.

Analyses. Subsequent analyses are planned to examine variance components resulting from grouping the soils by textural class and climatic regime, and possibly by soil order (e.g., forest soil, grassland soil, wetland). Such an approach will attempt to facilitate comparisons among and between soil types. Future efforts will also focus on the type of crops that are supported and the resulting yield from those crops. This approach will facilitate the comparison of the expected to the observed yield based on soil type.

Indices. Several possible approaches to providing an index of soil quality are possible. To answer the question of whether the conditions are suitable for plant growth, we will evaluate the SPRG model developed the SCS National Cooperative Soil Survey. This model provides incorporative, implementive measures of soil physical and chemical quality and a rating for potential biomass production. The model does not currently incorporate any measures relating to the soil biota; however, attempts will be made to add this component to the model. The model has also been tested only in South Dakota and Indiana, and thus we intend to validate the model for other regions. Calibration may be based on soil type or climatic data.

3.2.2.6 Components of variance and reliability ratios for pilot data

Analysis of the 1992 Pilot data revealed that the current sampling design of one transect per field and one composite sample per transect is sufficient to detect differences among segments and fields. Except for cadmium, the ranking of variance components from high to low was among segments, among fields, among transects, and within samples (Table 3.6). For cadmium, the ranking from high to low variance was among transects, among fields, among segments, and within samples. Similar variance component analyses will be applied to the SPRG model results to identify if differences can still be detected when all the data are combined.

Table 3.6 Variance components for related soil chemical and physical measures.

Parameter	Among Segments(σ^2_s)	Among Fields(σ^2_f)	Among Transects(σ^2_t)	Within Samples(σ^2_d)
Sand (%)	354.72	64.41	20.98	4.00
Clay (%)	94.12	16.68	5.93	3.56
pH (1:1 water)	0.10	0.07	0.04	0.02
Organic Matter (%)	1.75	0.65	0.60	0.04
Phosphorus (µg/ml)	3920.	3164.	1139.	419.
Cation exchange capacity (meq/100g)	13.43	4.58	3.13	0.83
Lead (µg/ml)	15.73	10.03	2.58	1.08
Cadmium (µg/ml)	775.9	823.4	940.3	124.1

Reliability ratios were calculated to determine the amount of signal versus noise in the data. Reliability ratio is defined as variance (signal)/(variance(signal) + variance(noise)) where the "signal" refers to the sampling or among segment and among field variability and "noise" refers to the measurement or within field and among samples variability (Table 3.7). As a reference point, a reliability ratio of 0.5 implies that the variance due to noise is as large as the variance of the signal. Conversely, a reliability ratio near 1.0 indicates that the variability from noise is a very small part of the overall variability. Ratios greater than 0.9 were calculated for sand content (%) and clay content (%) for all regions combined and separate regions with a sampling design that includes one transect per field and one laboratory determination per transect or composite soil sample (Table 3.7). Reliability ratios were greater than 0.70 for soil pH, organic matter (%), extractable phosphorus, and total lead for separate regions with a sampling designed with one transect per field and one laboratory determination per transect (Table 3.7). For cation exchange capacity, reliability of estimates in the Coastal Plain region of NC exceeds 0.70 with a design of one transect and one laboratory determination per transect, but a design of two transects and one laboratory determination per transect is necessary to exceed a reliability ratio of 0.70 in the Piedmont region (Table 3.7). For total zinc, estimates of reliability equal 0.9 in the Piedmont region of NC with the current sampling design, but require one transect with two laboratory determinations per transect to exceed a reliability ratio of 0.7 in the Coastal Plain region of NC (Table 3.7). For cadmium, estimates of reliability equal 0.81 in the Coastal Plain of NC with the current sampling design but require two transects with one laboratory determination each per field to exceed a reliability ratio of 0.70 in the Piedmont region of NC (Table 3.7).

Gabriel's biplot is an informative plot derived from principal component analysis that visually shows (1) the relationship among the independent variables, (2) the relative similarities of the individual data points, and (3) the relative values of the observations for each independent variable (Rawlings, 1988). It is called a "biplot" because both row (observation) and column (variable) information are displayed in one plot. The vectors in the plot represent the variables. If a vector length is close to unity, then it is well represented by the plot. If any vectors are not close to unity, then they are poorly represented by the plot and determining relationships among these variables should be avoided. The angle between two well represented variable vectors reflects their pairwise correlation. The correlation is the cosine of the angle because the variables were centered before the eigenanalysis was done (Rawlings, 1988). Therefore, angles of 0, 90, or 180 degrees indicates correlations of 1.0, 0.0, or -1.0, respectively. The numbers on the bottom and left axes are the lengths of the projected vectors in that dimension. The numbered points (+) on the graph represent the observations. Points close together have similar values with respect to the set of independent variables, and vice versa. The numbers on the upper and right axes are scale values for the numbered observations for the corresponding principal components.

The biplots illustrate associations among soil chemistry variables for all regions combined (Figure 3.3), Piedmont region (Figure 3.4), and Coastal Plains region (Figure 3.5). As expected, Bray 1 and Bray 2 methods for measuring extractable phosphorus were highly correlated, and sand and clay content were inversely correlated (Figures 3.3-3.5). Amount of calcium and magnesium found in soil samples was moderately correlated, and percent organic matter and clay content were positively correlated in the Piedmont region but not in the Coastal Plain region (Figures 3.4 and 3.5).

Table 3.7 Reliability ratios for soil chemical and physical measures[a] for the Piedmont and Central Plain regions and for regions of NC combined.

Parameter	t[b]	d[c]	Piedmont Region	Coastal Plain Region	Regions Combined
% Sand	1	1	0.96	0.92	0.94
	1	2	0.96	0.93	0.95
	2	1	0.98	0.96	0.97
% Clay	1	1	0.91	0.94	0.92
	1	2	0.92	0.96	0.93
	2	1	0.95	0.97	0.96
pH (1:1)	1	1	0.73	0.79	0.76
	1	2	0.75	0.86	0.79
	2	1	0.84	0.88	0.86
% O.M.	1	1	0.80	0.81	0.79
	1	2	0.81	0.82	0.79
	2	1	0.89	0.90	0.88
P (Bray II)	1	1	0.92	0.72	0.82
	1	2	0.93	0.75	0.84
	2	1	0.96	0.84	0.90
CEC	1	1	0.62	0.88	0.82
	1	2	0.63	0.90	0.84
	2	1	0.77	0.94	0.90
Zn	1	1	0.90	0.62	0.89
	1	2	0.92	0.71	0.91
	2	1	0.95	0.77	0.94
Pb	1	1	0.90	0.75	0.88
	1	2	0.91	0.79	0.89
	2	1	0.95	0.86	0.93
Cd	1	1	0.55	0.81	0.60
	1	2	0.56	0.83	0.61
	2	1	0.71	0.89	0.75

[a] calculated as signal/(signal + noise) which corresponds to:
$\sigma^2_s + \sigma^2_f$ / $[\sigma^2_s + \sigma^2_f + (\sigma^2_t/t) + (\sigma^2_d/td)]$; the higher the value the greater the relative amount of signal and the lower the relative amount of noise in the data collected.
[b] number of transects within a field
[c] number of laboratory determinations per transect or composite soil sample.

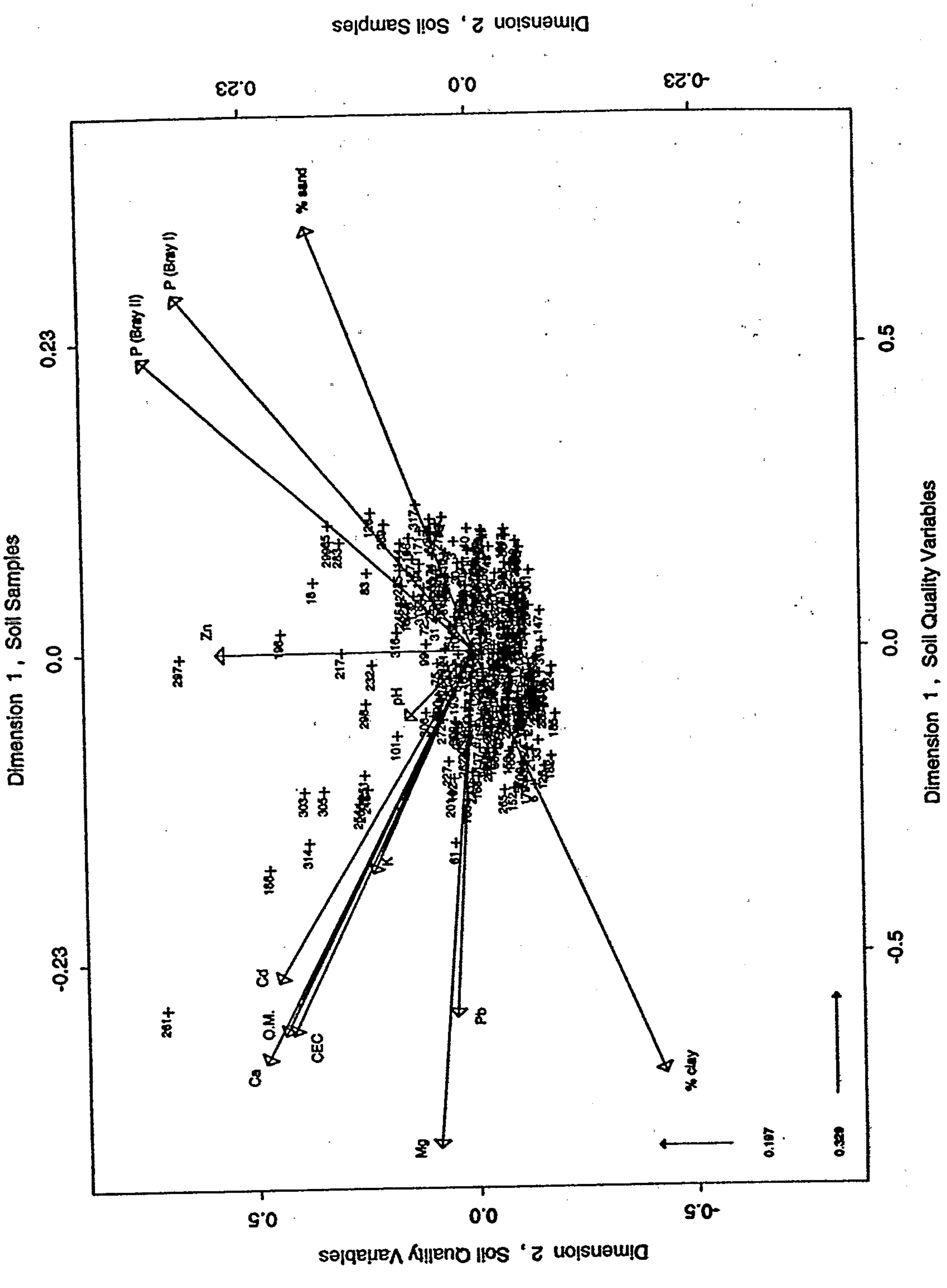


Figure 3.3 Biplot for all regions combined.

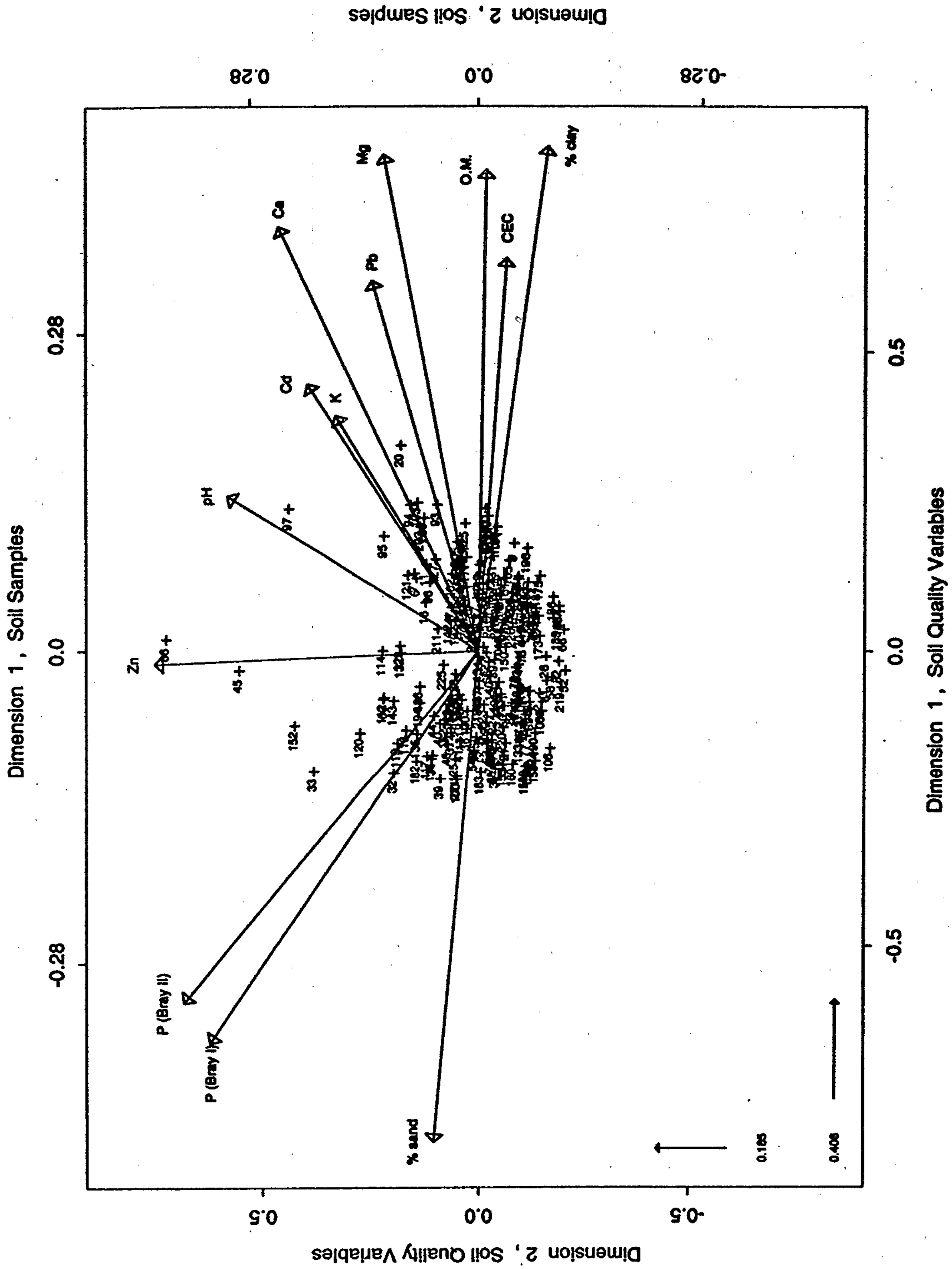


Figure 3.4 Biplot for Piedmont region.

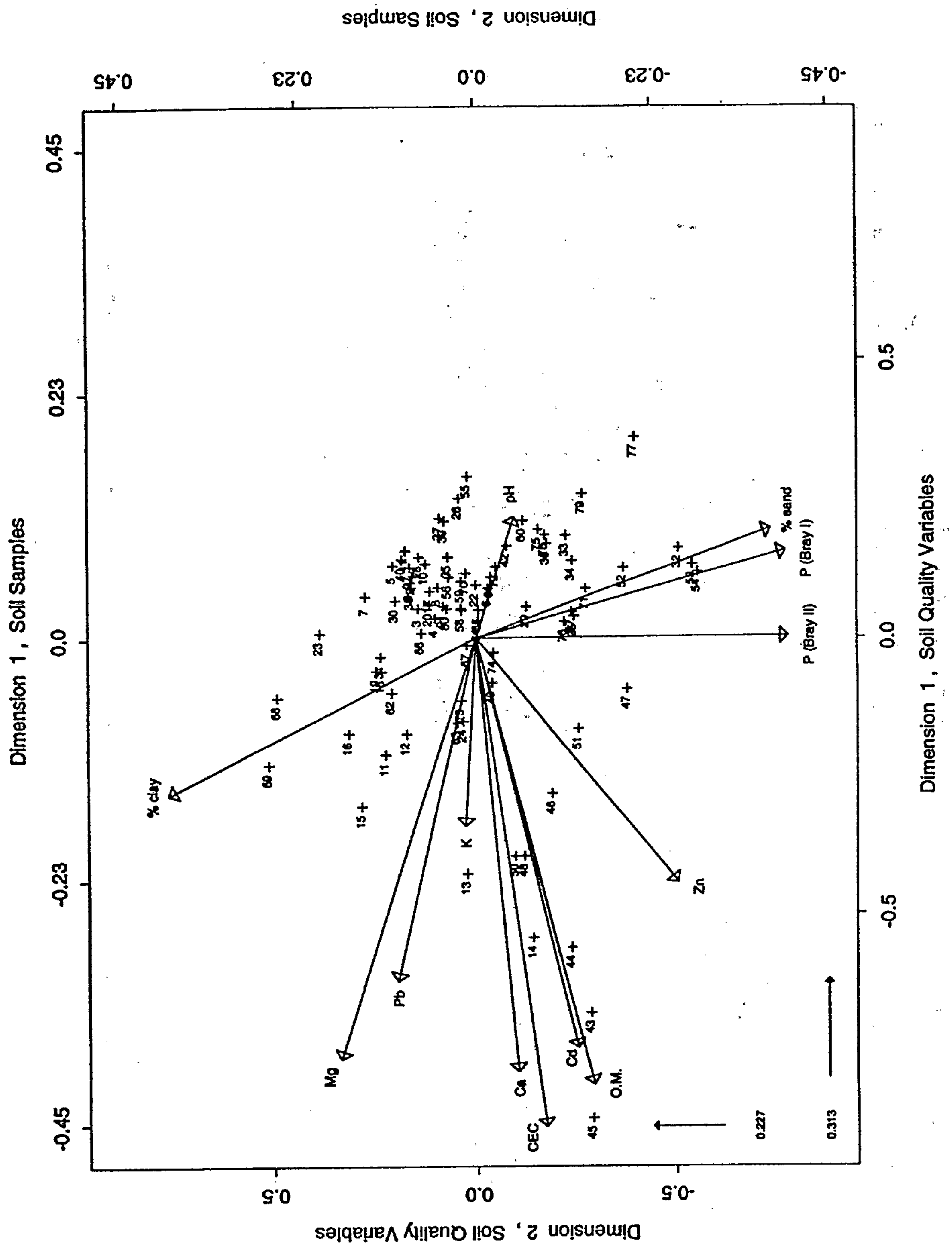


Figure 3.5 Biplot for Coastal Plain region

3.2.3 Soil Biotic Diversity

Three indices were computed for the nematode community in each soil sample:

- Maturity index for free-living nematodes (MI) (Bongers, 1990)
- Maturity index for plant parasitic-nematodes (PPI) (Bongers, 1990)
- Shannon index of trophic diversity (Shannon and Weaver, 1949)

Diversity of nematodes by feeding preference were estimated using the Shannon diversity index, $N1 = \exp[-\Sigma P_i(\ln P_i)]$, where P_i is the proportion of trophic group i in the total nematode community (Ludwig and Reynolds, 1988).

The MI was calculated as the weighted mean of the values assigned constituent nematode families (and the genera and species they contain) (Bongers, 1990): MI or PPI= $(\Sigma\ v_i * f_i)/n$ where v_i=the colonizer-persister (*c-p*) value assigned to family i, f_i=the frequency of family i in a sample, and n=total number of individuals in a sample. C-p values range from 1-5; however, plant-parasitic taxa are assigned c-p values from 2-5 because there are no plant-parasitic colonizers designated as 1 (Bongers, 1990).

3.2.3.1. Associations among soil properties and with nematode communities

Nematode maturity indices were compared to various soil chemical and physical properties thought to influence populations of nematodes i.e., clay content, sand content, pH, organic matter, extractable phosphorus, exchangeable potassium, exchangeable calcium, exchangeable magnesium, cation exchange capacity, total lead, and total cadmium (Table 3.8).

Table 3.8 Coefficients from a principal components analysis of variables of soil properties and nematode community indices. The first five principal components for MI, PPI, Shannon's index of trophic diversity (SHAN), and related physical and chemical parameters for soils are given.

	MI	PPI	SHAN	Sand	Clay	pH	OM	P
1	-0.01	-0.10	-0.03	-0.37	0.37	0.09	0.35	-0.26
2	0.07	-0.30	0.00	0.06	-0.13	0.65	-0.18	0.15
3	0.68	0.15	0.66	0.02	0.03	-0.02	0.07	-0.12
4	0.11	0.20	-0.00	0.16	-0.28	0.01	0.09	0.49
5	-0.09	0.66	-0.29	0.25	-0.08	0.13	0.22	-0.24

	K	Ca	Mg	CEC	Pb	Cd
1	0.21	0.32	0.37	0.30	0.28	0.27
2	-0.01	0.44	0.30	-0.25	-0.20	-0.11
3	-0.15	0.03	0.05	0.07	-0.13	-0.01
4	0.60	0.12	-0.04	0.41	-0.15	0.19
5	-0.33	0.20	0.20	0.13	-0.28	-0.01

Principal component analysis involves the linear transformation of the original set of variables to a set of orthogonal variables such that the first principal component accounts for the largest possible amount of the total dispersion, the second principal component accounts for the largest possible amount of the remaining dispersion, and so on. The first five principal components accounted for about 74% of the dispersion in the data. The first principal component (i.e., variables with the largest coefficients) was primarily soil chemical and physical properties. Conversely, the third principal component consists primarily of MI and trophic diversity (i.e., SHAN), and the fifth principal component consists primarily of PPI. Nematode indices tend to load on different principal components than do the soil chemical and physical properties, thus implying that the nematode indices provide information different from that provided by soil chemical properties.

3.2.3.2 Relationship among nematode community indices

Among the nematode indices only the MI and Shannon's diversity index highly correlated with each other (Table 3.9) Many nematode families that feed on bacteria are colonists (Bongers, 1990; Neher et al., 1994) and the presence of such families would decrease the Maturity Index for free-living nematodes (Table 3.9). Bacterivores are generally abundant in agricultural soils. As a result the trophic group, bacterivores, is inversely related to the maturity index of free-living nematodes and the Shannon trophic diversity index.

Table 3.9 Spearman correlations between indices and trophic groups (n=185) of soilborne nematodes.

	PP	SHAN	BACT[a]	FUNG[b]	OMNI[c]	PRED[d]	PLPAR[e]
MI	0.05	0.43***	-0.64***	-0.15*	0.25***	0.02	-0.12
PPI		0.01	-0.04	0.07	0.02	0.01	0.18*
SHAN			-0.17*	0.26**	0.39***	0.40***	-0.36***
BACT				0.37***	0.28***	0.25***	0.24**
FUNG					0.19*	0.16*	0.24***
OMNI						0.27***	0.06
PRED							0.18*

* $p<0.05$, ** $p<0.01$, *** $p<0.001$
[a]number of bacterivorous nematodes
[b]number of fungivorous nematodes
[c]number of omnivorous nematodes
[d]number of predaceous nematodes
[e]number of plant-parasitic nematodes

Table 3.10 Variance components for three indices associated with nematode community structures in soils. Actual variance values are presented.

Indices	Among Segments	Among Fields	Within Fields	Among Samples
Maturity Index (MI)	0.031	0.066	0.017	0.127
Plant Parasitic Index (PPI)	0.026	0.014	0.032	0.031
Shannon Trophic Diversity	0.002	0.080	0.173	0.134

Various components among samples and within fields were generally numerically greater than among fields and among segments (Table 3.10). These differences in magnitude of variance components are probably due to biological variation in numbers of nematodes within each composite sample and within fields. These differences in magnitude and relative ranking of the variance components is a source of concern about the acceptability of this indicator of condition.

The ability to differentiate ecological condition of soil among fields with statistical confidence improves if the variance among segments and among fields exceeds the variance within fields and within samples (Table 3.11). A measure of this is the reliability ratio and values $\geq$ 0.70 are desirable (see Table 3.7 for method of calculations for the reliability ratio). The results suggest that sampling plans which have multiple measurements per field may greatly improve the ability of these indices to differentiate ecological condition of soil among fields with statistical confidence.

One of the biggest challenges with this indicator is to locate additional personnel who have the expertise and time to identify nematodes to taxonomic family, especially nematodes that are not parasites of higher plants.

Table 3.11 Reliability ratios for several indices of nematode community structures for various sampling plans[a]

Indices	*t*[b]	*d*[c]	Reliability Ratio
Maturity Index (MI)	1	1	0.402
	1	2	0.546
	1	3	0.620
	2	1	0.574
	2	2	0.707
	2	3	0.766
	3	1	0.669
	3	2	0.783
	3	3	0.831
Plant Parasitic Index (PPI)	1	1	0.388
	1	2	0.457
	1	3	0.486
	2	1	0.559
	2	2	0.627
	2	3	0.654
	3	1	0.656
	3	2	0.716
	3	3	0.739
Shannon Trophic Diversity	1	1	0.211
	1	2	0.255
	1	3	0.274
	2	1	0.348
	2	2	0.406
	2	3	0.430
	3	1	0.445
	3	2	0.506
	3	3	0.531

[a]Calculated using the variance component estimates in Table 3.10.
[b]Number of samples taken within a field.
[c]Number of subsamples of a composite soil sample analyzed by the enumeration laboratory.

3.2.4 Water Quality: Method Development

How do the two methods of sampling pond water compare?

The standard method of sampling from a body of water requires a boat and a device to retrieve water samples from several depths. Water was collected from three points per pond and mixed. At each point, water was taken at three depths with a Kemmerer sampler. An alternate method was tested in which a stoppered teflon bottle at the end of a 5-m pole was used to draw a shallow (0.3 m) water sample from four points along the bank of each pond. Water from the four bank locations was mixed to form a composite sample. There was no significant difference ($P>0.6$) in pond nitrate concentration due to the sampling method, and the results of the methods were highly correlated ($r>0.99$). However, precision was slightly greater with the boat sampling method, based on variances between replicates within ponds (data not shown).

3.2.5 Extent and Geographic Distribution of Annually Harvested Herbaceous Crops

3.2.5.1 Successes and lessons learned

Full June Enumerative Survey used for extent estimates

Extent estimates are based on data from the complete National Agricultural Statistics Service (NASS) June Enumerative Survey, the largest sample available. During data analysis, extent was also estimated using the rotational panel and hexagon-selected samples. These samples, which are much smaller than the June Enumerative Survey sample, provided less precise estimates of extent. We decided to use data from the full June Enumerative Survey because they gave the most precise and accurate estimate of extent.

Strata map

The digitized NASS strata map is a very useful presentation tool and served as the focal point for describing the geographic distribution of the annually harvested herbaceous cropland resource. It provides a clear picture of the distribution of the agricultural resource throughout the state.

Strata maps for many of the other states will be unavailable for several years. Creating these maps is presently a labor-intensive and inexact process, both for us and NASS. Fortunately, NASS has developed a new system of area frame construction that will both simplify the construction of strata maps and improve their accuracy. As new frames are created for each state, these maps will become available. In the meantime, it's possible that an Advanced Very High Resolution Radiometer (AVHRR)-derived land cover map may provide similar information.

Diversity index selection

Measures of diversity are used to answer two questions:

(1) How many crops is one likely to find in an area?
(2) How even is their relative abundance in that area?

The first question is answered with a species richness measure. For each segment, the observation is the number of different annually harvested herbaceous crops in the segment. For this measure of diversity, segments with no crops must be included.

In the ecological literature, the Shannon index is the most commonly used measure of relative abundance. However, explaining the Shannon index is always difficult, largely because it does not have a direct biological interpretation. While it is relatively easy to explain a trend in the Shannon index, it is difficult to describe what a single value means. For these reasons, *the Shannon index will not be used here.*

Another index - the *Berger-Parker* index - is both simpler to calculate and easier to interpret. For each segment *i*, the Berger-Parker index is:

$$d = \frac{MAX(n_{i,j})}{N_i}$$

where N_i is the acres of annually harvested herbaceous crop in the segment and $MAX(n_{i,j})$ is the number of acres of the most common crop in the segment. *The Berger-Parker index is a measure of* dominance (*the inverse of evenness*): *the proportion of cropland occupied by the crop with the largest acreage.* The index varies $0 < d \leq 1$, where a value near 1 means that a single crop *dominates* the area, and smaller values mean the crops are more *even*ly represented. For this measure of diversity, segments with no crops must be excluded because the index is undefined. The Berger-Parker index was found to be closely, negatively correlated to the Shannon index for our data (Spearman r = -0.91).

Sanity Test

The estimated total land area of North Carolina was calculated from NASS data by summing the expanded area of all segments. This gave an area estimate of 77,525,091 ± 495,202 (at 95% confidence) hectares. According to the 1987 Census of Agriculture, the total land area of North Carolina is 77,184,583 hectares. This is within the 1% Measurement Quality Objective.

3.2.5.2 Challenges and issues

Most of the challenges result from differences between what we would like to get from June Enumerative Survey data, and what NASS expects from those data. These differences are, for the most part, minor. The order in which the issues are arranged should not be taken to imply relative importance.

Measurement error

Each datum used to calculate resource extent is the area of a field. Although NASS takes great pains to control the quality of their data these measurements are not 100% accurate. *We do not know the frequency and magnitude of measurement error*, and it is not accounted for in the estimates of variance in the *Land Use and Cover* report. The variance calculations used here assume that each measurement is completely accurate. We don't know how to deal with this problem yet, but our statisticians are thinking about it.

"Other crop" identification

During the June Survey, NASS records the acreage of infrequently occurring crops in an "Other crops" category. Although the exact crop is recorded on the survey form, some crops are aggregated during data entry and the dataset we receive does not allow us to determine what these "other crops" are. We would like to have a way of identifying the crops for which acreage is recorded in "Other crops."

The diversity of annually harvested herbaceous crops is one of our indicators. Calculation of diversity requires as input the number and area of each different crop found in each segment. During the analysis of data from the 1992 North Carolina June Enumerative Survey we found three "Other crop" variables: CROPOTHP, COPTL1PL, and COPTL2PL. CROPOTHP and COPTL2PL were both used to record "undetermined crops"; COPTL1PL was used to record all vegetables, including melons. The estimated extent of "Other crops" is larger than the estimated extent for some crops that are specifically identified (*e.g.* COPTL1PL = 8535 hectares; sunflower = 383 hectares). This makes it difficult to justify writing "other crops" off as insignificant. Also, for segments in which multiple "other crops" are recorded, we have no way of determining how many of these are different; and therefore no way of calculating the number of different crops in the segment.

We are also aware that "other crops" include crops that are *not* annually harvested herbaceous crops. These appeared in the CROPOTHP variable, which is one of the variables used to select fields for EMAP sampling.

This issue has been brought to NASS's attention in the form of a Survey specification for the 1994 North Carolina and Nebraska surveys. One solution which we would find acceptable is to have the identity of the "other crop" coded into another variable. For example, if acreage is recorded in CROPOTHP, code the crop name in another variable called CROTHPID (or something like that). Other solutions might be to assign unique variable names to *all* crops, or to assign individual "other crops" to the COPTLnPL variables (instead of groups of crops).

Idle land

We found the recording of "idle land" to be very confusing. Although a separate field was expected for idle land, it is sometimes combined with crop area. The reason for this has to do with farming practices, particularly for cotton and tobacco. A large amount of idle land can occur in a tobacco field as farmers plant their acreage allotment leaving spaces throughout the field. Further, the number of idle acres could be reported in a different column in the June Enumerative Survey table. For example, the tobacco acreage could be reported in field 2, while its idle acres might be reported in field 15. BUT these were not drawn off as separate fields on the photograph; rather, the two numbers (2 and 15 in this example) were written in the same field on the photo.

No action has been taken to rectify this issue because we have yet to determine how idle land fits into our resource classes. Two possibilities are:

(1) Include it with cropland or pasture, depending on its previous (and intended future) use.
(2) Establish an idle land resource class.

Once we have determined how we wish to view idle land, we will work with NASS to sort out the confusion.

For 1992, idle land was included for extent, but was not eligible for fall sampling.

Cropland pasture

NASS defines and records two kinds of pasture: permanent pasture and cropland pasture. Permanent pasture is grass and rangeland that is not in a regular crop-pasture rotation. Cropland pasture is land in a crop-pasture rotation, being used for pasture in the current year. The area of permanent pasture and cropland pasture sometimes appear in the same record as a crop. This was a surprise, as we expected pastures would be drawn off separately. The acreage involved was usually small.

A decision must be made regarding whether cropland pasture is part of the cropland or pasture resource. This decision will affect both extent estimates and indicator development. If cropland pasture is included in cropland, it is unclear how the productivity indicator would be measured on a field that is currently cropland pasture.

Cropland pasture was not eligible for fall sampling in 1992, nor was it included for extent.

Hay

"Grains harvested for hay" duplicates acreage that appears as either winter wheat planted, barley planted, oats planted, or rye planted. For extent, this problem was accounted for by use of the subtractive method.

Geographic coordinates for June Enumerative Survey (JES) data

We would like to have the latitude and longitude of the centroid of all segments that are part of the JES. During the 1992 pilot, coordinates were included on each record of the EMAP samples (both the rotational panel and hexagon-selected samples), but not for the complete JES. Although the variables MLAT and MLONG appeared in the JES dataset, the values were all set to missing. The solution we prefer would be an ASCII file containing segment number, latitude, and longitude. This could be directly imported into our geographic information system (GIS). Another solution would be to include the coordinates on each JES record. A 1994 JES survey specification requesting this change has been submitted and will be discussed with NASS.

Reporting on consistent regions

The regions used to summarize extent and geographic distribution were different than those used to summarize other indicators. Most indicators were regionalized by Major Land Resource Areas (MLRA); extent and distribution were reported by NASS strata. We could have reported extent within each MLRA by post-aggregating either rotational panel or hexagon selected data by MLRA. However, the sample sizes were small and extent estimates very imprecise; instead we chose to use the full June Enumerative Survey.

As part of our future analyses, and for the development of reports, we would like to post-aggregate extent data to the same ecological and political regions used for other indicators (e.g. hydrologic units, Major Land Resource Areas, ecoregions). Because we have access to these regions in a GIS format, availability of segment centroids would allow us to post-aggregate data to any and all of these regions. Analyses and reports could then be made within consistent boundaries. An alternative to obtaining segment centroid coordinates would be to have NASS provide a table of segments cross-referenced to specific regional identifiers. In the long run, this alternative is likely to be more costly and time-consuming than using segment centroids.

Communication of data needs

There was much confusion surrounding specific attributes of the survey data when we received it. Not all of the data were included and the meaning of variables in the datasets was unclear. The confusion was a result of our failure to communicate *exactly* what we needed to NASS, which was, in part, a result of our lack of understanding of NASS procedures.

Data presentation

Although cumulative distribution graphs provide a lot of information, they may not be the best presentation tool for all audiences. For example, histograms may be more interpretable for discrete data like crop diversity (Figure 2.3). The appropriate presentation techniques are very audience-dependent. Statisticians and scientists may be comfortable with cumulative distribution functions; histograms, bar charts, and pie charts may work better for general audiences.

Link to societal values

The few statements in the land use and cover section that attempt to relate the measures to societal values are weak and point out the need for further work. Further analysis of our 1992 data for correlation among indicators may help, but field research may be required to answer other questions. For now, use of these data is best limited to statements of the extent and geographic distribution of the annually harvested herbaceous crop resource.

At what scales should crop diversity over space be measured?

In this report it was measured over a 260 hectare area – the size of a NASS segment. This may not be the (only) appropriate scale. If we're interested in making statements about spatial diversity and vulnerability to disease, pest infestation, and erosion we must first measure them at various scales and find out which (if any) are correlated. Likewise, if we want to make statements about crop diversity and wildlife. Keep in mind that these findings are all likely to be region-specific. Even if we don't want to make correlations, if diversity is measured at all it should be done at several scales. (Perhaps the crop diversity of each ecoregion should be calculated.)

We need a measure of crop diversity over time – rotation

This could be an important indicator of vulnerability to disease and pest infestation, and of sustainability in general. We are working on a crop rotation index that may be calculated using crop history data obtained during the Fall Survey.

3.2.6 Landscape Structure

The work on landscape structure described in the 1991 Pilot Plan (Heck et al., 1992) has been slowed by several obstacles, including the learning curve for ARC/INFO, data management difficulties, and changing priorities. Although no results are presented, much has been learned.

3.2.6.1 Successes and lessons learned

Appropriate data format: Raster, not vector

For broad geographic areas covered in fine spatial detail, the vector format is not appropriate; raster data are strongly preferred.

Preliminary work was conducted using Thematic Mapper-based land cover data for the Albemarle-Pamlico Drainage Basin, a 27,500 square-mile area that includes parts of North Carolina and Virginia. Classification was at a spatial scale of (approximately) 28 meter x 28 meter pixels. These are the only extensive, classified data presently available for North Carolina. The land cover data were delivered in ARC/INFO vector format, separated into files representing the USGS 1:100,000 quad sheets. The classification system included 19 categories of land cover. We wanted to aggregate land cover into eight classes, in a seamless coverage of the region.

Working with this much data in vector format proved unacceptable. Software and hardware limitations were encountered; even when things worked, analyses were very slow. After the study began, a raster-based module (GRID) was released for ARC/INFO. Portions of the data were converted from vector to raster format, which proved much easier to work with.

Calculating of measures of landscape pattern

Once the data were available in a raster format, developing algorithms to calculate most measures landscape pattern detailed in the Pilot Plan was straightforward. In most cases, the procedure involved exporting information from ARC/INFO and performing the analysis using SAS.

3.2.6.2 Challenges and issues

Thematic classification: How much detail?

The level of thematic classification detail required is unknown. To date, EMAP-Landscape Characterization has attempted to produce a single classification system for land use and cover. This approach results largely from the belief that one can define a one-to-one association between land cover and EMAP Ecosystem Resource Groups.

Combined land use and cover classifications confound the issues of use and cover, with a resulting loss of information. *Separate coverages of land use and cover are preferred,* using GIS to overlay them as needed. Ecosystem Resource Group membership could then be decided using information from any number of thematic data layers (*e.g.*, land use, land cover, ecoregion, soil type, elevation).

Classification of broad geographic areas

Completing a consistent, and accurate, land cover map for the nation is a formidable task, yet it is critical to the development of landscape indices. Vegetation over specific areas can be highly variable, yet disparate vegetation can have similar reflectance; the same vegetation can even have different reflectance at different times of the day or year. Even within the relatively small area of the Albemarle-Pamlico basin, data for adjacent scenes were acquired months apart, different parameters were used for different geographic regions, and some cover classes were dropped for particular regions. Calculation of measures of pattern across scene lines using these data is thus difficult, if not meaningless.

Appropriate scales of space and time

The appropriate spatial and temporal scale(s) for monitoring and research depends on the question(s) being asked and the scale(s) at which pertinent processes occur. We need to gain a better understanding of the scales at which agroecosystem processes occur and collect data at those scales. We look to cooperative work among EMAP Resource Groups and the EMAP Landscape Ecology and Landscape Characterization groups to help us address these issues.

Ecosystem boundaries

The question of where one "ecosystem" ends and another begins is unanswerable without reference to particular processes or organisms. Further, one cannot define the extent of an ecosystem without reference to particular processes or organisms. While it is possible to delineate and calculate the extent of land covers at various scales without reference to a particular process or organism, a patch of a single cover does not necessarily bound an ecosystem. Again, the belief that one can define a one-to-one association between land cover and EMAP Ecosystem Resource Groups leads us astray. This is another argument against combining classification themes into a single data layer; doing so hampers our ability to redefine system boundaries for different processes and organisms.

Measurement error and confidence in measures of landscape pattern

Landscape ecologists have been using GIS representations of remotely-sensed data to calculate measures of landscape pattern but have devoted little effort to quantifying the uncertainty in these measures. Without statistical confidence in the measures used, scientists cannot evaluate correlations between landscape pattern and ecological processes. Without statistical confidence one cannot use measures of pattern to detect differences in landscapes over space or changes in a landscape over time. This problem should be of great interest to all EMAP Resource Groups using measures of landscape pattern, and should be on the research agendas of both the EMAP Landscape Characterization and Landscape Ecology groups.

Classified land cover data in a GIS, usually derived from remotely-sensed data, are frequently used to calculate measures of landscape pattern. Classification error is usually reported in the form of an error matrix based on an accuracy assessment of the classified data. An error matrix clearly describes the accuracy of each classification category, as well as the nature of the confusion among categories. Further uncertainty is introduced when multiple GIS maps are overlaid – a common procedure among GIS users. Although awareness of these and other error-related issues is common in the remote sensing and classification communities, the implications of these issues have yet to be addressed by landscape ecologists. *There are no procedures for using an error matrix to generate confidence intervals for measures of landscape pattern. We are working to overcome this problem.*

Interpretation of measures of landscape pattern

Theoretical work indicates that landscape pattern measures may reflect the ability of organisms to inhabit and traverse a landscape, the potential for materials or disturbances to move from one part of the landscape to another, or the types of processes that are shaping the landscape. But these relations are hypothetical and difficult to test. We look to cooperative work among landscape ecologists, the EMAP-Landscape Ecology Group, and the EMAP Resource Groups to begin addressing this issue. This implies that work at the landscape scale must be designed carefully to serve not only as a method to monitor patterns, but also as a method to test hypotheses at multiple spatial scales.

Interactions with EMAP-Landscape Characterization and EMAP-Landscape Ecology

The operational details of the interaction between Resource Groups and the EMAP-Landscape Characterization Group need to be clearly defined. There have been major changes within EMAP-Landscape Characterization over the last 4 years. During these changes, the roles and responsibilities of the Landscape Characterization Group have not been clearly communicated to the Resource Groups. Further discussions are essential to determine how the Resource Groups and EMAP-Landscape Characterization can interact successfully.

One of the changes in the EMAP-Landscape Characterization group was the spinoff of an new group, EMAP-Landscape Ecology. The manner in which the EMAP-Landscape Ecology group and the Resource Groups will interact is also undefined.

3.2.7 Agrichemical Use

How does EMAP-Agroecosystems pesticide use information compare to what other groups do?

More extensive pesticide surveys are currently done by others for specific crops, for example the Agricultural Chemical Usage reports published by NASS for selected crops (e.g. Agricultural Chemical Usage: 1992 Field Crops Summary, USDA-NASS, Agricultural Statistics Board, Washington, D.C., March 1993, Ag Ch 1 (93)).

One strength of the EMAP-Agroecosystems survey is that it included most field crops and, in that respect, can provide more complete information about our resource than the current NASS Agricultural Chemical Usage report. The NASS report is based on a sample of only major crops and is only conducted in the states that account for most of the area of each crop. In a very few cases, a comparison can be made between the EMAP survey data and the Agricultural Chemical Usage findings. According to the latter, the average rate of atrazine applied to corn in North Carolina was only 1.43 kg/ha/treatment in 1993, considerably less than our value. Yet that same report would indicate that the total area to which atrazine was applied was 363,000 ha for corn alone, much more than our estimate of the total. These differences are approximately 1.5 to 2 standard errors in magnitude. Attempts to reconcile the numbers were unsuccessful.

3.3 EVALUATION OF PILOT ACTIVITIES

3.3.1 Interactions with USDA-NASS

The 1992 Pilot was a successful test of cooperation between EMAP-Agroecosystems and NASS. NASS personnel were involved in four areas of the Pilot: (1) collecting data during their regular JES, (2) drawing the sample of fields to be visited in the fall, (3) designing and conducting the fall questionnaire and, simultaneously, (4) collecting fall soil and water samples. Two primary contacts were provided by NASS: one in Washington, D.C. and one in North Carolina. One member of the ARG served as liaison officer to NASS; however, the day-to-day cooperation was very decentralized. The technical director, indicator leads, statisticians, and information manager all worked with various NASS personnel as needed.

One area of particular sensitivity has been the confidentiality of data from individual NASS surveys. Each member of the ARG who participates in collection or use of the data must sign a "certification and restrictions on use of unpublished data (off-site)" (NASS Form ADM-043). This form states that the signee will abide by United States Code Title 18, Sections 1902 and 1905, and Title 7, Section 2276. These codes deal with crop market information divulgence and personal identity of the data source. Any violations of these codes is punishable by fines and/or imprisonment. A memorandum of understanding is being drafted between NASS and the USDA-Agricultural Research Service (ARS) that will guide future handling of confidential data (see Section 3.4.4).

Both the interview and sampling for the 1992 Pilot were done by part-time "enumerators" who work for NASS through the National Association of State Departments of Agriculture (NASDA). They are accustomed to doing various NASS surveys and adapted well to the new job of taking soil and water samples (see Section 3.3.3). The Agroecosystem program is not restricted to using only enumerators. Some future monitoring may require field crews with more technical training and expertise.

The ARG and NASS together developed the questionnaires and enumerator's manual. Training was also a joint responsibility. All printed materials, including labels, were produced by NASS. For the 1992 Pilot, the ARG purchased all equipment, assembled the kits, and distributed them at the training school. In the future, more of these logistical responsibilities will be handled by NASS. Surveys and sampling as well as all other NASS activities were conducted on a reimbursable basis and were funded through an interagency agreement between U.S. EPA and NASS.

Although everything eventually fell into place, there were some elements, such as water sampling procedures, that were being revised up until the last minute. This is something that is of concern to NASS, so the ARG is trying to adjust our timetables for future programs.

Other aspects of cooperation with NASS are covered in Sections 3.4.4 (Information Management) and 3.1 (Preliminary Evaluation of Designs).

3.3.2 Logistics

The 1992 Pilot in North Carolina was generally a logistical success. The logistics activities can be broken down into questionnaire performance and survey administration, soil sampling, water sampling, shipping and sample tracking, laboratory relations, and reporting back to the respondents. The survey and sampling period ran from the end of October until mid-December.

Questionnaire performance and survey administration. Most aspects of the survey went well. The questionnaire was completed for 326 (94%) of 348 sample units (Table 1.3). There were 11 questionnaires not completed because of operator refusal and 11 for which the operator was inaccessible. Of the 326 "completed" questionnaires, some had partial data: two were missing fertilizer information, and 11 more were missing both fertilizer and pesticide information. One drawback of the sampling scheme was that often fields were sampled more than once (Table 1.4), so that failure to contact or receive cooperation from one operator meant the loss of several samples. This problem will be corrected in the 1993 pilot by drawing the sample from a larger set of fields.

Areas for improving the questionnaire were identified. For example, the occurrence of double crops and cover crops had to be deduced on a case-by-case basis, because no specific code was provided to identify them. That code was added in 1993. The questions designed to arrive at fuel use (not reported here) were also difficult to enumerate. This section was changed for 1993 so that field operations, but not fuel use, were tallied. The question about crop rotations as a pest control practice was too vague to give meaningful data for 1992.

Soil sampling. Soil was not collected for 27 of the sample units (Table 1.5), usually because the farmer refused or was inaccessible. On six of those 27, however, the field was a woodland or wetland and should not have been selected for sampling. In two other cases the enumerator made a mistake and failed to get the soil sample. Also, several farmers allowed soil sampling but declined the interview. Only one farmer (representing seven sampling units) gave the interview but refused to allow soil sampling.

The soil sampling procedure itself was done satisfactorily (see Section 3.3.3). Some of the enumerators reported difficulty with laying out the transect, use of the soil sampler, and mixing the soil.

Water sampling. Water was not collected for 11 of 51 ponds and 20 of 61 wells (Table 1.5) due to farmer refusals or inaccessibility. Ponds and wells were sampled only from the Rotational Panel design.

Water sampling procedures and equipment purchases were not finalized until the last minute. Considerable effort was made to borrow equipment for this part of the project, but in the end most of it had to be purchased, including depth finders and several small jon boats. To prevent enumerators from getting steel splinters, nylon clothesline cords were substituted for the cables that came with the Kemmerer samplers. However, one of the cords broke during use. The sampler was recovered from the bottom of the pond, and all enumerators were given the original steel cables to use.

Shipping and sample tracking. The use of Federal Express as a courier for shipping samples was successful. At first the enumerators were calling to have the samples picked up at their homes, but then the NASS office determined where the nearby Federal Express offices were, so that samples could be dropped off later each day. The procedure went smoothly and samples were delivered in a timely manner. Note that the water samples were sent via "priority" overnight service, whereas the soil went under the regular government contract service.

There were a few accounting difficulties with Federal Express. Without a weight written on each "airbill" label, a large default weight was sometimes charged. It also seemed that several of the charges on the final billing report may have been incorrect.

Sample tracking was done by the ARG logistics officer, who received database files from NASS and faxed sheets from the laboratories. Identification numbers of the samples sent were compared to those of the samples received. Although a rather tedious procedure, it was useful in identifying, for example, transcription errors at the laboratories. Federal Express was very efficient in providing information about packages that were thought to be lost (using their airbill identification number).

Laboratory relations. Identification time for nematodes exceeded initial estimates, and a few errors were found in the soil chemistry data, but in general the laboratories performed well. Some delays were also experienced with receipt of data from the EPA Athens-ERL, where the water samples were analyzed. Nitrate testing was done by a federal scientist and pesticide testing by a contractor: Technology Applications, Inc.

Reporting back to the respondents. By mail, NASS reported some soil and water data back to the individual farmers who participated in the 1992 Pilot. One unfortunate incident was that there was a mix-up in the units in which the nitrate data were reported. Thus, farmers were told that concentrations were about 4 times their actual values. NASS, fearing that it would create confusion and negative public relations, did not send out the corrected results. This is not the ideal situation, but at least the concentrations were not understated, and some people had had their wells retested.

3.3.3 Quality Assurance

Two separate quality assurance (QA) programs were used during data collection for the 1992 Pilot. NASS has a well-established QA program, which was used during the collection of the questionnaire data for its annual JES and for the EMAP-Agroecosystems fall survey. All other organizations participating in sampling and analytical activities were required to follow QA procedures developed by the ARG, in addition to their own quality assurance activities. These organizations included:

- the ARG;
- NASS for the collection of soil and water samples;
- U.S. EPA Environmental Research Laboratory (ERL) in Athens, Georgia, for water sample analysis;
- U.S. EPA Environmental Monitoring Systems Laboratory in Las Vegas, Nevada;
- Agrico Agronomic Services Laboratory in Washington Court House, Ohio, for soil sample analysis; and
- N&A Nematode Identification Service in Davis, California, for nematode identification and enumeration.

Overall, the quality assurance program, although immature, was successful in obtaining data of high and known quality during the Pilot. Quality assurance activities occurred at six specific steps detailed below.

3.3.3.1 Area frame development

During the development of the area frames used to determine sampling locations, quality assurance activities were performed by NASS in accordance with the procedures specified in Area Frame Design for Agricultural Surveys (Cotter and Nealon, 1987). The NASS area frame was successfully converted to a geographic information system (GIS) coverage in ARC (a commercial package). The GIS lead for the ARG verified that the primary sampling units developed around EMAP hexagon centroids actually contained the appropriate centroid.

3.3.3.2 Enumerator training

Training of enumerators for the collection of questionnaire data (June and fall) was the responsibility of NASS. NASS conducted a national training school in April 1992 to train the NASS supervisors. A state school was held in May 1992 to train enumerators for the JES. Each enumerator was provided with a manual that included the standard operating procedures for the collection of the verbal survey data. The school for the Fall Survey was conducted in October 1992. Enumerators were instructed on collection of soil and water samples as well as questionnaire data. This later training school was conducted cooperatively by NASS and the ARG, including the EMAP-Agroecosystems Quality Assurance Officer (QAO). An enumerator's manual was again provided.

For administration of questionnaires, enumerator training consisted of a thorough discussion of all questions and possible appropriate answers. Enumerators were also instructed on how to deal with unusual or unexpected situations. For soil and water sampling, all procedures were presented and discussed in a classroom setting, followed by a field demonstration and practice.

3.3.3.3 Collection and handling of survey data

Survey data collection, processing, and output were the responsibility of NASS. Complete NASS QA procedures were not used, but the following steps were taken. As usual, enumerators each worked under the guidance of a supervisory enumerator. Questionnaires were sent back to the NASS state office, where they were checked for completeness and reasonableness in the first part of a two-stage edit process. If this manual edit turned up incomplete or questionable data, the enumerator was contacted again, or if necessary, the farmer was phoned directly from the state office. The step of having a supervisory enumerator review two surveys from each enumerator's workload was skipped because of the tight sampling schedule.

After the manual edit, a detailed computer edit was done to test that data were within expected ranges and were internally consistent. Any problems at this stage were brought to the attention of the statistician coordinating the survey in North Carolina, who resolved them based on the original questionnaires and notes made by enumerators. A statistician in Washington, D.C. also reviewed pesticide usage for outliers and reasonableness. Any questions were relayed to the statistician in Raleigh who resolved them. It was not necessary to re-contact any enumerators or farmers for this step. After edits, a final summary was written.

3.3.3.4 System and performance audits

Field technical system audits of soil and water sampling were performed. Two technical systems audits were performed on two different NASS survey crews. These audits were conducted by the EMAP-Agroecosystems QAO in cooperation with a member of the North Carolina NASS office.

The NASS enumerators performed their soil and water sampling efficiently. Only a few problems or concerns were identified, and they were minor and were not expected to have any significant influence on the integrity of the samples.

A technical systems audit was performed on the Agrico Agronomic Services Laboratory in Washington Court House, Ohio, by the EMAP-Agroecosystems QAO. The purposes of the audit were: to observe the operations of the analytical laboratory, to review existing preliminary data as a pre-QA check (preliminary performance audit), and to answer questions regarding the analyses to be performed.

In general, the Agrico laboratory was found to be generating acceptable data and was following ARG protocols. Most areas of concern were relatively minor and could easily be incorporated into laboratory operations. However, two areas of major concern were noted during the audit. First, samples were processed through a 14-mesh sieve whereas most soils are processed through a 10-mesh sieve to meet the standard USDA size definition for soil (i.e., 2-mm or finer). The other area of concern was that sample aliquots were obtained for analysis on a volumetric and not on a gravimetric (i.e. sample with known weight) basis. Whereas this type of sample aliquoting may be appropriate for soil fertility testing, the potential variability in the sample results can affect the long-term monitoring efforts of the EMAP Agroecosystems program.

It was noted during the review that laboratory quality control (QC) samples were prepared and analyzed at a higher frequency at Agrico than required for most U.S. EPA programs indicating the laboratory's concern and effort to obtain high quality results. Additional laboratory control samples were incorporated into each analytical batch. These samples were collected and processed by Agrico in bulk and have established acceptance windows to allow the technicians to rapidly assess, during ongoing sample analysis, the functioning of the equipment and the accuracy of the measurement system.

During the performance audit phase of the laboratory audit, results of the other QA/QC samples (to be discussed) were examined for both accuracy and precision. Where comparable procedures were used, accuracy results were generally within the determined acceptance windows for those given parameters. Precision among the analytical replicate and field split samples was generally within a relative percent difference of 10% or less for all parameters undergoing analysis indicating that the sample preparation techniques in the field and laboratory were functioning well. Upon examination of the data from the field replicate samples, larger relative percent differences were identified, indicating a high variability within the field, in particular for phosphorus.

3.3.3.5 Use of QA/QC samples

Soil samples

Precision was assessed through the analysis of analytical duplicates created in the laboratory, field split samples (i.e., two samples obtained from a single composite soil sample), and field replicates (i.e., two different samples collected from two different transects within the same field). Field split and replicate samples were sent blind to the laboratory. Although no measurement quality objectives had been set for precision in the EMAP-Agroecosystems program at the time of the audit, with few exceptions, the data generated for analytical duplicates were within a relative percent difference of 10% or less for all parameters examined.

In addition to the precision samples, a performance evaluation sample was obtained by the EMAP-Agroecosystems QAO from the EMAP-Forests program and incorporated into the sample stream. Several replicates of this sample were submitted throughout the time period soil sampling was being performed by the NASS enumerators. The laboratory did not know which samples were the performance evaluation samples. Where comparable determinative procedures were used, accuracy results were generally within the pre-determined acceptance windows for those given parameters.

Water samples

Water samples were analyzed for nitrate during November and December 1992, following the procedures outlined in Payne (1992). No technical systems audit was performed on the Athens laboratory by EMAP. Aqueous samples were directly injected into the ion chromatograph, so no fortified samples were needed beyond the nitrogen standards used for calibration. The correlation coefficients indicated no significant day-to-day variance in the standard curve. Forty-eight samples (29% of the total) were split for replication. The average ratio between duplicate and original came out to be 99.98%, with a standard deviation of 2.9. Six samples (3%) were considered outliers and analyzed in triplicate.

Of the 12 pesticides for which the water was tested, none were detected. All available evidence indicates that these zero values are legitimate. A second extraction and chromatography were done on 10% of the samples, all of which showed no detectable levels of the compounds. Fortified spike samples for each compound were analyzed at 1, 2, and 3 ppb (only 1 ppb for alachlor, aldicarb, and cyanazine). Average percent recoveries for the replicated 1 ppb spikes ranged from 60% (aldicarb) to 285% (cyanazine). This shows that the compounds would have been detectable at or above 1 ppb, though the target accuracy of 70% and target precision of ± 10% would not have been met. Laboratory blanks were included for each batch (10% of total). No field blanks were sent to the laboratory, but field contamination was clearly not a problem.

Note that sample bottles were chilled on ice and then shipped overnight in styrofoam cases within cardboard boxes. They arrived the next morning in good condition. The temperature of the first few samples was tested and found to be 10-15 C. This shows either that the samples stayed cool during transit or that they re-chilled en route in the aircraft. Stability studies showed that the samples were extracted (24 hr) and the extracts analyzed (1-2 mo), well within the periods during which the compounds were stable in storage (7 d and 6 mo, respectively).

3.3.3.6 Data validation

Upon receipt of the completed data sets from NASS, Agrico Agronomic Services Laboratory, N&A Nematode Identification Service, and ERL-Athens, ARG statisticians and indicator leads, in cooperation with the ARG information manager, validated the assembled database. For example, an ARG statistician evaluated the minimum and maximum values of the soil parameters to ensure that the values were within the range of possible values. An ARG statistician and indicator lead also checked all soil parameter values which were greater than five standard deviations from the sample mean to be sure that values were correct. The ARG ensured that data were missing from the dataset only when the soil samples had not been taken. Questions about specific values for soil or water parameters that appeared to be anomalous were referred to the appropriate personnel at the respective analytical laboratories.

Although NASS checks for reasonableness and consistency, the EMAP-Agroecosystems logistics coordinator followed up on apparent problems with the questionnaire data when they arose. Some of these problems indicate that data will need to be collected differently in the future, while other problems show that extra edit checks are needed. In most cases, the questionnaire data were not changed once they reached the ARG.

No attempt has been made to quantify these nonsampling errors in the questionnaire data, but several deserve mention. During the fall visit it sometimes happened that a different crop was in the sample field than what was expected; sometimes the field had no annually harvested herbaceous crops at all. Survey information and soil samples were taken anyway, except when the site was found to be woods or wetland. Some, but not all, of the differences can be explained as changes that occurred between June and fall. The sample was drawn using data from the June Enumerative Survey; however, even field sizes were sometimes reported differently between June and fall, indicating that a different field was visited on the two trips. The yield and management information are surely subject to similar uncertainties. The fact that the "unit weight" was not consistent within crop species gives another indication of possible survey problems, although they were used as given.

Unexpectedly, there were a few cases where an enumerator "estimated" yield and other data when a farmer was unavailable (out-of-state). This will be expressly forbidden for the 1993 survey in Nebraska. When the respondent was available but harvest was not complete, yield data were usually recorded by NASS as missing values; however, in some of those sample units, an estimated yield and harvest date were recorded. Estimated data were not excluded from the 1992 Pilot analyses.

QA checks were not performed on the determination of soil map units, but the results did vary when some were done twice by different people. The procedure involves comparing the NASS aerial photograph (showing the outline of the sample field) to SCS soil survey maps. In the recent 1993 Pilot, some map units will be determined on the ground by SCS scientists as a check on the accuracy of the map-and-photo method.

The primary reporting tool for this report is the cumulative distribution function (CDF). ARG statisticians submitted data sets that had been analyzed by hand to the CDF computer programs to verify that the estimated distribution and its confidence bands were being computed correctly.

3.3.4. Information Management

3.3.4.1. Introduction

Information management activities were focused in three areas: 1) the development of a working relationship with NASS, particularly in the area of data confidentiality and data movement; 2) the development of an initial infrastructure for the Agroecosystems Information Center (AIC) - items such as hardware, software, and networks were acquired, installed, and configured to satisfy the requirements of the ARG for pilot activities; and 3) data management. Because this was the first field season for the ARG, much of the work was new and involved documenting activities and working closely with indicator leads and statisticians.

3.3.4.2 NASS interactions

The development of a cooperative relationship with NASS was a primary goal of the 1992 Pilot. The process of familiarization with each other's activities, reasoning, terminology, methods, personnel, etc., was very successful. This success has already shown benefits in the planning for the ARG 1993 Pilot Field Program.

Working within the confines of the NASS data confidentiality regulations was a challenging activity for information management. Sample level identifiers were masked to remove any systematic identification and replaced with unique, "random" identifiers. When sample level data were required with identifiers, ARG personnel traveled to the NASS State Office in North Carolina to acquire data and perform aggregation or masking until that data would pass screening for removal from the premises. This proved to be very cumbersome because the types of analyses required relied heavily on the sample level identifiers.

To address the issues of data confidentiality and transfer for the future pilot activities, a Memorandum of Understanding (MOU) between NASS and ARS is in the draft stage, which will specify that the ARG can have the data with identifiers at the ARG headquarters. The MOU will require that ARG personnel sign NASS Form ADM-043 (Certification and Restrictions on Use of Unpublished Data) and that the Information Manager develop a security plan and procedures which are subject to periodic audit by NASS. Restrictions on the data would still remain for any "public" use of the data that does not adhere to the requirements of Form ADM-043 and pass a confidentiality screening by NASS. The provisions of the MOU will allow the ARG to integrate and analyze data in a much less restricted manner while protecting data confidentiality. Once the EMAP Information Management System is on-line and the ARG is ready to become a functioning node of that system, the question of "public" release of the data will be addressed.

The movement of pilot data from NASS to the ARG proved to be more difficult than expected. Several data sets were required from NASS: JES for all of North Carolina; the JES that was performed on the EMAP Hexagon segments; the Fall Survey data; and several reference files. Most of the difficulties encountered were either related to miscommunication or data confidentiality. A post-pilot meeting was held with NASS to address these problems. As a result of that meeting, an e-mail link between NASS and the ARG has been established, protocols and standards have been developed for data interchange, and the MOU with NASS is under development.

3.3.4.3. The Agroecosystem Information Management System Infrastructure

Emphasis was given to the acquisition, installation, and learning of a technology infrastructure for the ARG for the 1992 Pilot. The analysts used Sun workstations configured with at least 24 megabytes of memory, with local and shared disks, and color displays. The remaining staff each have an MS-DOS personal computer with at least 8 megabytes of memory.

A suite of software was available to the analysts. SAS and S-Plus for statistical analysis; WordPerfect for word processing; ARC/INFO and ArcView for GIS; and Corel Draw for graphics. Indicator leads, management, and support staff had WordPerfect, Freelance, and various spreadsheets available for use, as well as access to the workstations.

All of the computers are attached to a local area network (LAN). The network has a print, file, compute, and mail server attached to it that is accessible to all users. By structuring the network in this way, expensive devices can be shared by all of the staff and they can be configured in the most efficient manner. The network's most important feature, file sharing, made the analysis of the 1992 Pilot data logistically simpler than if files had not been shared. By providing access to current versions of common data files, version control was easier to maintain.

The LAN is connected to the NCSU campus network, which is in turn connected to the Internet. This allows members of the ARG housed in locations other than the ARG headquarters to have access to all of the facilities described above through the NCSU network and modem bank. Other important aspects of the Internet connection are electronic mail to NASS and other EMAP personnel and the ability to become a node on the EMAP Information System that is currently under development.

3.3.4.4. Data management

The movement and management of data (Fig. 3.6) was critical to the success of the pilot. Planning for data management consisted of NASS/ARG interactions to determine process, schedule, and flow of data from the field to the ARG. The processes were monitored carefully and adjustments made as necessary. The flow of data, although somewhat awkward, worked as planned. Once results were received from the analytical laboratories, they were merged with the survey and sample design data.

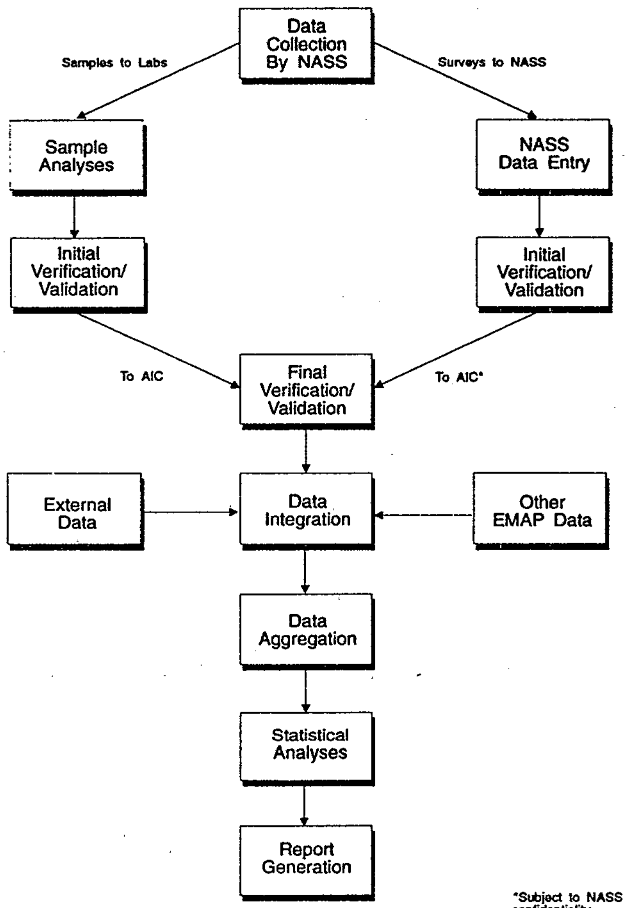


Figure 3.6 Data flow for the 1992 Agroecosystem Pilot.

From the perspective of the ARG, NASS performed their IM related duties well. When questions or problems arose, they were solved promptly. The ability to go back to the source of the data (the NASS enumerator) was tested and was successful. NASS showed great flexibility in adapting to changing needs during the pilot, sometimes transferring similar data sets to the ARG several times. A person from NASS to serve as the specific liaison with the ARG would greatly improve understanding and coordination, decrease confusion, and generally benefit the ARG effort. This arrangement is being pursued.

As verification and validation routines were performed on the data, errors were encountered. For errors in the laboratory sample data, direct communication with the lab cleared up all mistakes. For errors in the fall survey data, there was not a clearcut procedure for tracing back in the data stream to find the source of the error and to correct it. Most of the time the ARG went to the NASS State Office and consulted the original paper form. This worked reasonably well for a one state pilot, but it would have to be refined for larger field projects.

When errors were discovered in the data, corrections were made only by the Information Manager. Corrected data sets were numbered consecutively and notes of the corrections were attached to the data by means of documentation files. These files are accessible to the ARG by means of a Gopher server on the ARG workstation server. The Gopher entries were linked dynamically to the files so that changes would be available immediately to the staff. This method is also being used to document the ARG GIS Library.

A challenge that confronted the analysts was to integrate the sample and survey data into one data set and analyze duplicate and split-sample data without specific identifiers attached to the data sets. Both of these tasks were difficult to accomplish because the identifiers were withheld or changed to ensure data confidentiality. To solve most of the problem, a cross-reference file was provided by the NASS that used a sequential number (value from 1 to N, where N is the number of sample units) as the identifier to link the samples and surveys together. This did not solve the problem of analyzing the duplicate and split samples because the sequential number applied only to the primary sample for that sample site. Eventually, enough information was included in the cross-reference file to distinguish the specific type of sample. This problem was encountered again when comparisons of the two sample designs required strata and sub-strata identifiers for the NASS segments. NASS is understandably reluctant to release segment numbers because when combined with a map that identifies them, they define exact location of a 1 square mile area.

The ARG made some use of external data during the 1992 Pilot. The North Carolina State Soil Survey Database was used for the soil quality and crop productivity indicators, the Soil Conservation Service's Land Resource Region boundaries were used as ecologically meaningful regions for data summarization, NASS strata were used for the land use indicator, NASS aerial photography was used for soil map unit identification, and climate data were obtained for use in indicator research.

3.3.4.5. Future Efforts

The focus for information management during the 1992 Pilot was not on systems development, but ongoing through the procedure for the first time with close attention to the individual processes. Most of the data management and analysis tools were created *de novo* and many times performed manually, which allowed for a thorough understanding of the details. This understanding will be used to review the requirements for the Agroecosystem information system.

For future monitoring efforts, the ARG and NASS need to cooperate more closely in the information management arena. Much remains to be learned about each other's operations, equipment, software, and expertise that could be used to benefit the EMAP program. NASS has recently finished an exhaustive search for a relational database management system to use for their future enterprise data management system. Since the EMAP Information Management System is being developed with a relational database management system, it is to the benefit of our working relationship to be aware of the parallel development efforts. The Soil Conservation Service is starting a similar process to re-engineer their soils databases.

Areas such as version control, data documentation, database management, security, data modelling, metadata, verification/validation, and sample tracking need to be developed. The management of the data in 1992 was accomplished using SAS. Using SAS for management of data was found to be unacceptable. It is imperative that the ARG procure and began using a relational database management system as soon as possible.

4. LITERATURE CITED

Angle, J.S., Gross, C.M., Hill, R.L., and McIntosh, M.S. 1993. Soil nitrate concentrations under corn as affected by tillage, manure, and fertilizer application. *J. Environ. Qual.* 22:141-147.

Barker, J.C., Zublena, J.P., and Campbell, C.R. 1991. Livestock manure production rates and nutrient content. pp. 322-324 in *The 1991 North Carolina Agricultural Chemicals Manual.* The College of Agriculture and Life Sciences, North Carolina State University at Raleigh. January 1991.

Bongers, T. 1990. The maturity index: an ecological measure of environmental disturbance based on nematode species composition. *Oecologia* 83: 14-19.

Bouchard, D.C., Williams, M.K., and Surampalli, R.Y. 1992. Nitrate contamination of groundwater: sources and potential health effects. *Journal AWMA.* September 1992. pp. 85-90.

Burkholder, JoAnn (NCSU Dept. of Botany), personal communication re. significance of nitrate in ponds. 10 September 1993

Cotter, J., and Nealon, J. 1987. Area frame design for agricultural surveys. U.S. Department of Agriculture, National Agricultural Statistics Service, Research and Applications Division, Area Frame Section. Washington, DC.

EMAP. 1993. *Environmental Monitoring and Assessment Program Guide.* EPA/620/R-93/012. Research Triangle Park, NC: U.S. Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory.

Fedkiw, J. 1991. *Nitrate Occurrence in U.S. Waters (and Related Questions): A Reference Summary of Published Sources from an Agricultural Perspective.* United States Department of Agriculture. Washington, D.C. September 1991.

Heck, W.W., Campbell C.L., Finkner A.L., Hayes C.M., Hess G.R., Meyer J.R., Munster M.J., Neher D., Peck S.L., Rawlings, Smith C.N., and Tooley M.B. 1992. Environmental Monitoring and Assessment Program - Agroecosystem 1992 Pilot Project Plan. EPA/620/R-93/010. U.S. Environmental Protection Agency, Washington, D.C.

Lal, R. 1991. Soil structure and sustainability. *Journal of Sustainable Agriculture* 1(4):67-92.

Ludwig, J.A., and Reynolds, J.F. 1988. Statistical Ecology: A Primer on Methods and Computing. John Wiley, New York.

Neher, D.A., and Campbell, C. L. 1994 Nematode communities and microbial biomass in soils with annual and perennial crops. *Appl. Soil Ecol.* 1: (in press).

Neher, D.A., Peck, S.L., Rawlings, J.O., and Campbell C.L. 1994. Measures of nematode community structure for an agroecosystem monitoring program and sources of variability among and within agricultural fields. *Plant and Soil* (in press).

Nelson, D. W., and Sommers L. E. (1982). "Total carbon, organic carbon, and organic matter." In *Methods of Soil Analysis, Part 2*. American Society of Agronomy and Soil Science Society of America, Madison, WI.

Olsen, S. R., and Sommers L. E. (1982). "Phosphorus." In *Methods of Soil Analysis, Part 2*. American Society of Agronomy and Soil Science Society of America, Madison, WI.

Overton, W.S. 1993. Probability Sampling and Population Inference in Monitoring Programs. Department of Statistics, Oregon State University, Corvallis, OR.

Overton, W.S., White D., and Stevens D.L. 1991. Design Report for EMAP - Environmental Monitoring and Assessment Program. EPA/600/3-91/053. U.S. Environmental Protection Agency, Washington, D.C.

Payne, William R. (Project Officer). 1992. QAP for Agroecosystem EMAP North Carolina Pilot including Sampson County Project: Nitrate Analysis. Task#: 416. Internal Document of ERL-Athens. October 26, 1992.

Pimentel, D., Hurd, L.E., Bellotti, A.C., Forster, M.J., Oka, I.N., Sholes, O.D., and Whitman, R.J. 1973. Food production and the energy crisis. *Science* 182: 443-449.

Rawlings, J.O. 1988. Applied Regression Analysis. Wadsworth & Brooks/Cole Advanced Book & Software. Pacific Grove, California.

Ricklefs, R. E. 1990. Ecology. Third Edition. Chiron Press, New York.

Shannon, C.E., and Weaver, W. 1949. The Mathematical Theory of Communication. Univ. Illinois, Urbana.

Southwell, P.H., and Rothwell, T.M. 1977. Report on Analysis of Output/Input Energy Ratios of Food Production in Ontario. March 31, 1977. School of Engineering, University of Guelph, Ontario, Canada. Contract serial number OSW76-00048.

Spalding, R.F., and Exner, M.E. 1993. Occurrence of nitrate in groundwater -- a review. *J. Environ. Qual.* 22:392-402

Stevens, D.L., Jr. 1993. Implementation of a National Environmental Monitoring Program. submitted to the Journal of Environmental Management.

Yeates, G.W. 1994. Modification and qualification of the nematode maturity index. *Pedobiologia* (in press).

USDA-National Agricultural Statistics Service. 1992. Agricultural Chemical Usage: 1992 Field Crops Summary, USDA-NASS, Agricultural Statistics Board, Washington, D.C., March 1993, Ag Ch 1 (93)

☆U.S. GOVERNMENT PRINTING OFFICE: 1994 - 550-001/80394