Statistical Methodology for the NRI-CEAP Cropland Survey

J. Jeffery Goebel
Natural Resources Conservation Service
U.S. Department of Agriculture
Beltsville, MD

[draft – May 2009]

I. Introduction

The National Resources Inventory (NRI) is a longitudinal survey conducted by the USDA Natural Resources Conservation Service (NRCS), in cooperation with the Iowa State University Center for Survey Statistics and Methodology. The purpose of the NRI is to provide support for agricultural and environmental policy development and program implementation. The NRI is a panel survey of land use and associated natural resource attributes, conducted at 5-year intervals from 1982 through 1997, and annually from 2000 through the present.

Scientists, economists, resource managers, and policy makers have found the NRI to be a source of scientifically credible and nationally consistent data that help them formulate policy proposals and analyze economic and environmental impacts. The NRI was developed in the 1970's by the United States Department of Agriculture (USDA) as a tool to assess status, condition, and trend of soil, water, and related resources on the Nation's non-Federal lands, as mandated by the Rural Development Act of 1972 and the Soil and Water Resources Conservation Act of 1977 [see Goebel (1998)]. The NRI survey system was built upon the survey system used for the Conservation Needs Inventories of 1956 and 1967 [see Nusser and Goebel (1997)].

A variety of natural resource issues have been analyzed using the NRI. These issues include: land use change with emphasis on loss of agricultural lands to urban development; conservation provisions of the 1985 Farm Bill; trends in soil erosion; gains and losses of wetlands due to agricultural activities; transport of agricultural chemicals into water supplies; the role of agriculture in sequestering carbon.

USDA has made significant investments in developing a methodology that assesses the environmental effects of various conservation practices and systems. This methodology utilizes physical process models and the framework provided by USDA's NRI survey program. The models simulate resource condition changes that occur over time on a cropland field, taking into account both natural factors and farm operator decisions. The methodology developed by USDA has been used to: estimate environmental effects of existing conservation programs; assess current conservation needs; analyze on-going Technical Assistance; and examine the effects of alternate conservation systems and proposed conservation policies and programs. The NRI provides basic inputs into the models and a scientifically-credible method for assimilating and interpreting model results. [see Goebel and Kellogg (2002)]

Previous NRI modeling efforts used generalized information about farming practices that were not specific to the soils and climate data associated with the specific sample sites [see Potter et al, 2006]. Modeling capabilities have been greatly enhanced for the NRI-CEAP Cropland Assessment, to provide simulations that better reflect specific cropland field conditions and to provide results for a broader suite of resource issues. These enhancements required that additional data be collected for NRI sample sites and that the scope of existing process models be broadened. These additional data were acquired through the "NRI-CEAP Cropland Survey", which was designed to obtain detailed data describing farming activities and conservation practices for the field associated with selected NRI sample points. Trained enumerators working for the National Agricultural Statistics Service conducted interviews with the individuals (farm owner and/or operator) responsible for making management decisions for these fields.

II. Overview of Survey Design

The objective of the NRI-CEAP Cropland Survey was to obtain additional site specific data needed to utilize the field-level process model APEX to estimate field-level effects of conservation practices. The process model was run for a sub-sample of NRI sample points; inputs for a sample point included historical NRI site specific data, data obtained from the NRI-CEAP Cropland Survey for the agricultural field where the sample point is located, additional information on conservation practices from Field Office records, soil properties and characteristics associated with the particular soil at the sample point location, and climate data associated with the sample point location. The input data associated with a particular point describe a "representative field;" outputs from the process model runs include losses of materials (such as sediment and chemicals) from this field and changes in condition (such as accumulation of carbon). These outputs are used to estimate both on-site and off-site effects.

The APEX model outputs can be treated like other NRI variables; the site specific results for each sample point can be aggregated or averaged for some meaningful portion of the landscape using statistical weights. The statistical (survey) weight for an NRI sample point is the acreage value assigned to that sampling unit based upon the sampling design and certain control figures [derivation of weights for the NRI-CEAP Cropland Survey is discussed in Section VI, Estimation Procedure]. The APEX model outputs also serve as inputs into hydrologic models that simulate transport of water, sediment, and chemicals from the land into and through stream networks and eventually into estuaries and oceans. The NRI-CEAP data and the models can then be used to estimate changes in in-stream concentration of sediment and chemicals that result from changes in land management.

The sampling strategy utilized for the NRI-CEAP Cropland Survey was to select a sub-sample of NRI sampling units from the NRI Foundation Sample; in particular, a subset of sample points was selected from those sampling units used for the 2002 and 2003 Annual NRI surveys. Sampling strategies for the NRI Foundation Sample, Annual NRI surveys, and the NRI-CEAP survey are discussed below. The NRI sampling structure provided a natural framework for the data collection and modeling activities needed to support the CEAP national cropland assessment; it also provided efficiency to the process because sample locations were already identified and significant data already existed for these sites. The full collection of NRI sample sites provides a statistically credible representation of the diversity of soils, climate, cropping systems, and natural resource issues for the Nation's agricultural lands. Data collection activities were spread over a four-year period because of financial constraints and operational considerations. A different set of sample points was selected for each year. The goal was to develop a data base that supported statistical analysis of the benefits of conservation practices at the national and regional levels.

III. NRI Foundation Sample

The universe of interest for the National Resources Inventory (NRI) survey consists of all surface area [land and water] of the U.S., including all 50 states, Puerto Rico, the U.S. Virgin Islands, and certain Pacific Basin islands. The NRI Foundation Sample only covers the 48 contiguous states, Hawaii, Puerto Rico, and the U.S. Virgin Islands; it does not cover Alaska and the various Pacific Basin islands. The sample does cover all land ownerships categories including Federal, although NRI data collection activities have historically concentrated on non-Federal lands. Federal land area is covered by the NRI Foundation Sample for several reasons, including: ownership maps and data bases have typically been out-of-date and incomplete; ownership patterns change over time; and data collection activities have occasionally included all ownership categories.

The NRI Foundation Sample was selected on a county-by-county basis; units analogous to counties were used where county designations do not exist. Each county sample was selected using a stratified, two-stage, area sampling scheme; specific procedures and sampling rates varied from county to county. The foundation sample is basically that which was used for the 1997 NRI. The sample was initially established for the 1982 NRI; modifications occurred for a number of counties during the 1990's [see Goebel and Baker (1987), and Nusser and Goebel (1997)]

Area Sampling

The NRI survey system uses pre-defined areas of land and specific point locations as sampling units. Fixed geographical locations are used rather than features that can change due to the effects of human activities and natural occurrences. Some agricultural surveys use farms, farm enterprises, fields, tracts, farmers, and/or operators as sampling units. The boundaries or definitions for those units are subject to change over time; this would not work for the NRI, where the same sampling units are visited periodically.

Stratification

Stratification serves to make sampling more efficient by subdividing the entire population of interest (data universe) into non-overlapping portions (layers or strata) that are more homogeneous than the population as a whole. The NRI design uses both geographical stratification and stratification based upon specific resource conditions and general ownership patterns.

The Public Land Survey System (PLSS) provides the basis for geographical stratification in counties covered by this system, except for Arkansas and Louisiana. The PLSS subdivides counties into townships and sections. Townships are nominally square areas of land that are six miles on a side; a township is typically divided into 36 one-mile square areas of land called "sections". Sections are numbered from 1 to 36 in a serpentine manner, starting in the northeast corner of the township. For sampling purposes, three strata of 12 sections each were formed within each township; one stratum contained the top two rows of sections [sections 1 thru 12], a second stratum contained sections 13 thru 24, and the third contained sections 25 thru 36. Each of these geographical strata, therefore, was a two -mile by six -mile rectangular areas of land. A number of townships and sections have dimensions that differ from the "standard" system, due to operational difficulties encountered while conducting a land survey.

The PLSS does not cover the 13 northeastern states, Texas, Hawaii, parts of Ohio, and the southeastern states of Georgia, South Carolina, North Carolina, Kentucky, and Tennessee. In

these 6 southern states and the non-PLSS portions of Ohio, lines analogous to township lines and section lines were superimposed on county highway maps, and geographical strata were developed in the same manner as counties covered by the PLSS. For counties in the 13 northeastern states, a sampling system based upon latitude and longitude was developed; strata are rectangular areas of land two minutes of latitude by four minutes of longitude.

Stratification based upon specific resource conditions occurred mostly in large counties located in the Western portion of the U.S. Three or four types of strata were usually constructed, in addition to the geographic stratification described above.

- First, many areas that were under irrigation (or potentially irrigated) were identified and placed into specific strata, and then sampled in a manner analogous to that used in irregularly shaped counties.
- The second type of strata typically included land that supported non-irrigated farming operations.
- Additional resource-based strata were formed for areas that would not support farming operations. Sometimes large tracts of Federal land were identified and placed into specific strata.

Stratification was modified and sampling was augmented in about 200 counties between 1982 and 1997. Analysis of historical NRI data showed these modifications were needed to better estimate conversion of prime farmland and other rural lands into urban lands. Most augmentation occurred in exurbia areas outside of already established cities and suburban communities.

Selection of Segments

Each two-mile by six-mile stratum contains 12 sections. For most standard counties the sections were subdivided into four half-mile square "quarter sections", with nominal size 160 acres. This meant there were 48 quarter sections within a standard stratum, and each quarter section was considered a potential sample unit, or segment. Sampling rates differed from county to county, and sometimes within county, depending upon factors such as county size, complexity of soils and agricultural practices, and number of counties within the state; workload and budget issues were balanced with statistical reliability. The most common sampling rates were one per stratum (1 out of 48, or approximately 2%) and two per stratum (approximately 4%). Some strata had lower sampling rates; for example, a one-half percent sampling rate was accomplished by grouping four adjacent strata and randomly selecting one of the 192 quarter sections. For portions of counties where some sections (and partial sections) did not fit within the regular two-mile by six-mile strata, groups of 12 sections were formed and segments were randomly selected in a manner similar to the procedure used for regular strata.

Many counties in Western states contained additional stratification and different sized segments. Strata formed in irrigated areas often contained 40-acre segments because of the greater heterogeneity in these areas. Strata formed in relatively homogeneous areas of range, forest, and barren land often contained 640-acre segments, both for statistical reasons and because of the difficulty in locating sample sites on the ground in areas where landmarks are limited.

For counties in the 13 Northeastern states, sampling units are 20 seconds of latitude by 30 seconds of longitude; because of the curvature of the earth, they range in size from 96 acres in northern Maine to 113 acres in Virginia. Each stratum contains 48 segments, and selection procedures were handled in a manner similar to that described for 2-mile by 6-mile strata.

The Universal Transverse Mercator (UTM) grid system was used to define sampling units for counties in Arkansas and parishes in Louisiana. For Arkansas, the sampling units were square kilometers of land [approximately 247 acres]. Segments were numbered sequentially in a geographic order starting in the northwest corner. The initial sample in 1982 was selected systematically at a rate of 1 out of 10; less than a half of these sample segments [about 6,100] are part of the NRI Foundation Sample. For Louisiana, the sampling units were half-kilometer squares of land and the strata were 4-kilometer squares. Randomization with control was used to select segments within strata.

Selection of Points within Selected Segments

Three specific sample point locations were selected within most selected segments. Fewer points were originally selected for segments in Louisiana, Arkansas, and 40-acre segments within portions of the Western U.S. The procedure used for a standard 160-acre segment is outlined below; it is a restricted random procedure that assured the points were spread throughout the sample segment.

- Step 1. The segment was conceptually sub-divided into 36 square blocks, each approximately 440 feet on a side. These 36 blocks were assigned numbers from 1 to 12, with each integer assigned to three blocks [see Figure 1]. As indicated by the double lines in Figure 1, the segment could also be pictured as having three rows and three columns of blocks; numbers were assigned to blocks so that each integer occurs only once in each row and once in each column. In addition, no two blocks with the same number are contiguous. These restrictions assured as wide a dispersal as possible within the segment.
- Step 2. Sample point #1 was determined by selecting two random integers between 1 and 2,640. These numbers designated the point location in terms of feet north and east from the southwestern corner of the segment.
- Step 3. Sample points #2 and #3 were located in the other two blocks having the same number [from 1 through 12] as the block in which Point #1 fell; the designation of which was Points #2 was made randomly. The relative location of Points #2 and #3 within these blocks was the same as for Point #1.

The selection of two sample points within a standard 40-acre segment can be expressed as follows:

- Step 1. Sample point #1 was determined by selecting two random integers between 1 and 1,320, say N_1 and E_1, where N_1 designates feet north and E_1 designates feet east of the southwest corner of the segment.
- Step 2. The coordinates for point #2 were then:

$$N_2 = \text{mod}_{1320}(N_1 + 660), \text{ and}$$
$$E_2 = \text{mod}_{1320}(E_1 + 660).$$

In 640-acre sample segments and in the 13 northeastern states, three sample points were selected in a manner analogous to that used for standard 160-acre segments. In Arkansas and Louisiana, initially only one point was randomly selected within each sample segment; this has been modified so that those segments also contain three sample points.

Remarks

The stratification described above was mostly established for the 1956 Conservation Needs Inventory (CNI), except for counties and parishes in Arkansas, Louisiana, and the 13

Northeastern states. Where possible, subsets of segments established for the 1956 CNI were used for the 1982 NRI; soils information had already been detailed for these segments, and this provided a more efficient start-up. Stratification and segment selection for Arkansas, Louisiana, and counties in the 13 northeastern states were new for the 1982 NRI. Development of strata and selection of sample segments was done manually, by staff at the Statistical Unit, Iowa State University; precise rules and random number tables were supplied.. The procedure for computerized selection of sample points within sample segments was developed for the 1977 and 1982 NRI surveys; sample point coordinates were printed on self-adhesive labels that were affixed to the data collection worksheets.

Figure 1. Sub-Division of Segment for Sample Point Selection

1	2	3	4	5	6
7	8	9	10	11	12
4	6	12	7	1	3
11	10	2	5	9	8
5	3	1	6	4	2
9	12	8	11	7	10

IV. Annual NRI Sample Design

Introduction

The National Resources Inventory (NRI) was conducted every five years during the period 1977 through 1997. During the second half of the 1990's, NRCS worked with the Statistical Unit at Iowa State University on development of a statistical design for transitioning to an annual survey approach. Several types of factors needed consideration:

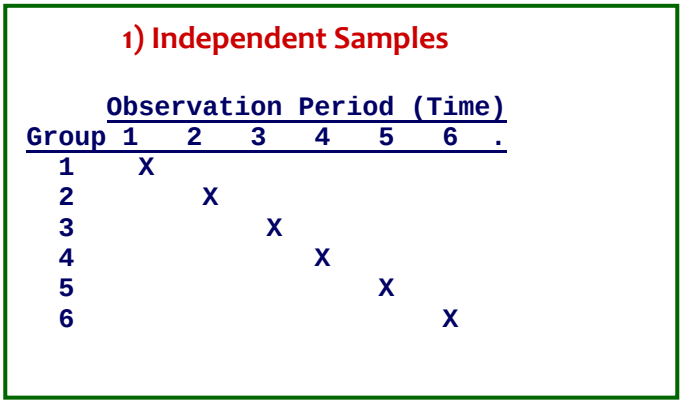
- programmatic factors – budgets and staffing requirements needed to be constant from year to year
- statistical factors – reasonable statistical efficiency for estimators of short-term change, long-term change, and condition/level for any given year
- analysis capabilities – the new approach needed to continue support of development of a data base that supports complex analysis
- inter-agency cooperation – an annual survey approach makes it easier to work with other Federal agencies on development of collaborative inventory approaches
- adaptability to changing information needs.

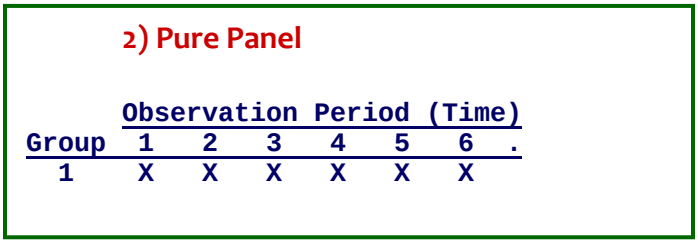
The desire was to continue to use the NRI Foundation Sample sampling units but to use some subset each year rather than to sample all units every five years. Four alternative longitudinal designs were considered [see Figure 2]:

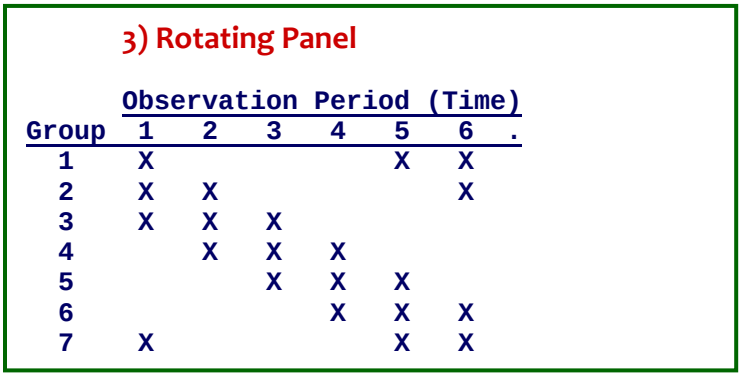
- Independent Samples – the entire collection of sample units would be split into equal non-overlapping groups (panels), and each year one of the panels would be sampled until all panels were completed, at which time the first panel would be re-sampled – this is the basic design used for the USFS Forest Inventory and Analysis (FIA) program, which operates on a 5 – 10 year cycle depending upon the state; there could be variations, such as sampling with replacement and having varying rates of re-sampling for certain categories of sampling units
- Pure Panel – each year's sample would be the same as that used for the previous year, which means 100% of the sample would be revisited every year – this was the basic NRI design between 1982 and 1997, except the panel was observed every five years rather than annually
- Rotating Panel – the samples are split into groups (panels) of samples, and in any one year two or three panels are being sampled, which means that a given panel is sampled for two or three consecutive years and then "rotated" out of the sample and replaced by the next panel – this design has been used by the National Agricultural Statistics Service (NASS) for some agricultural surveys
- Supplemented Panel – this is a type of split plot design where one specified (core) panel is sampled each year and each year a rotating panel is also sampled

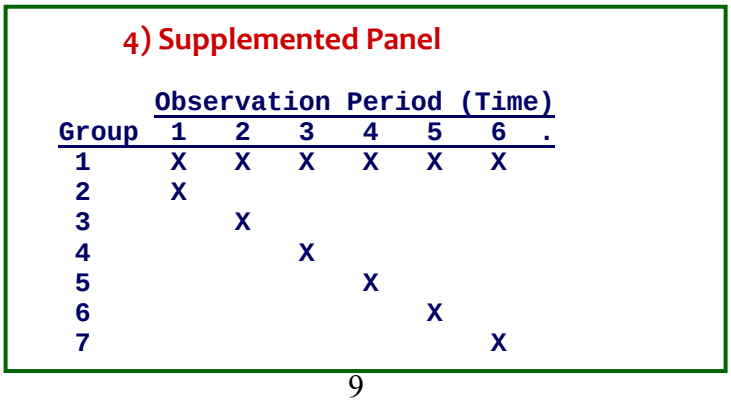
The supplemented panel design was selected for the Annual NRI survey system because it serves as a compromise design. The Core Panel that is sampled every year provides an efficient method for estimating net change over time; the Rotation or Supplemental Panel provides efficiency for estimation of status at a given point in time. [see Fuller (1999) and McDonald (2003) for further discussion]

Figure 2: Longitudinal Survey Designs Considered for Annual NRI









NRI Special Study data collected during 1995 – 1997 for 5,972 sample segments in 538 counties were used to estimate correlations and measurement error variances, so that the four designs could be compared [see Breidt and Fuller, 1999] Observations from 1982, 1987, and 1992 were also available for these sample units.. A first-order autoregressive process was assumed. Year-to-year correlations were quite high, with the estimated Autoregressive Coefficient, ϕ, being > 0.95 for all tested variables, and > 0.99 for many. Measurement error was significant for soil erosion estimates, non-cultivated cropland, large water bodies, and small streams; the estimated portion of variability due to measurement error, ν, is 23% for soil erosion and 13% for non-cultivated cropland. Measurement error estimates were needed to study the make-up of the autoregressive process because it has a large effect upon the estimated First-Order Autocorrelation, $\rho(1)$. This parameter is above 0.96 for items like forest land and area in roads that change very little from year to year and are measured with very little measurement error; $\rho(1)$ was 0.74 for soil erosion estimates and 0.84 for non-cultivated cropland. Area in CRP was the only estimate studied that did not show a good fit to the first-order autoregressive process; this is understandable given the nature of that type of land.

The estimated parameters for the measurement error and the first-order autoregressive process were used to study Θ, which is the portion of the sample to place in the pure panel (or Core Panel). Recall that $\Theta = 1$ provides the best efficiency for measuring change and $\Theta = 0$ provides the best estimates for a single point in time. Moving from using $\Theta = 1.0$ to $\Theta = 0.5$ showed that estimates of change degraded slower than the gains made in estimating a period mean (i.e., status/condition at one point in time, or level). The conclusion of this analysis was that $\Theta = 0.5$ was a good statistical compromise and also provided adaptability if there was a need to change survey items [see Breidt and Fuller (1999)]. Several of the effects of implementing a Supplemented Panel longitudinal survey design for the Annual NRI survey, using $\Theta = 0.5$, are shown in Tables 1 and 2.

The design that was implemented in 2000 for the Annual NRI survey process was a Supplemented Rotation Panel design with Θ slightly greater than 0.6. This design was selected because there is only a moderate loss of precision relative to other designs, and it provides flexibility for status and gross change estimates. Several strategies have been implemented to minimize loss relative to the 5-year cycle (full 300,000 segment panel) estimates, including: selection of Core Panel and Rotation Panels using stratification and non-proportional allocation; utilization of historical information from the full NRI Foundation Sample of 300,000 segments in estimation process; and implementation of techniques such as constrained generalized least squares.

Selection of Segments for Core Panel P00

The Core Panel P00 for the Annual NRI survey program was selected from the 300,000 segments in the NRI Foundation Sample. Samples were selected on a state-by-state basis. The same basic procedure was used for the contiguous 48 states, Hawaii, Puerto Rico, and the U.S. Virgin Islands.

Segments were placed in categories based upon historical data available for all samples, and then varying sampling rates were assigned to each category.

Table 1: Projected Increase in Width of Confidence Intervals

[for Annual NRI Sample Design, Compared with Previous 5-Year Cycle]

Type of Estimator	Estimator	Cultivated Cropland	Developed Land
Status [for 2002]	Single year	+ 17 %	+ 18 %
	Average of 3 years	+ 8 %	+ 7 %
Change [1997 – 2002]	Single year	+ 61 %	+ 61 %
	Average of 3 years	+ 27 %	+ 27 %

Table 2: Projected Size of 95 % Confidence Intervals, in Millions of Acres

	National	Missouri	Washington
Area of Cultivated Cropland	325.4	10.72	5.60
• 5-year cycle	(± 1.7)	(± 0.25)	(± 0.32)
• Annual – GLS	(± 2.0)	(± 0.29)	(± 0.37)
• Annual – Mod	(± 1.8)	(± 0.27)	(± 0.34)
5-Year Change in Cultivated Cropland	- 25.4	- 1.61	- 0.64
• 5-year cycle	(± 1.02)	(± 0.15)	(± 0.16)
• Annual – GLS	(± 1.64)	(± 0.24)	(± 0.26)
• Annual – Mod	(± 1.30)	(± 0.19)	(± 0.20)

Classification of Segments:

1. Wetland: If the segment contained at least one point with a Cowardin wetland classification for 1992 and/or 1997;
2. CRP: If the segment was not in Category 1, and at least one of the points in the segment had a land cover/use of "land in CRP" for 1987 and/or 1992 and/or 1997.
3. Urban Change: If the segment is not in Category 1 or 2, and less than 90% of the segment is classified as "Urban and Built-up" land, and either "Urban and Built-up" land within the segment changed from 1987 to 1997 or acres of roads with the segment changed by more than four acres.
4. Urban: If the segment is not in Category 1, 2, or 3, and the segment contains some land classified as "Urban and Built-up" but less than 90% of the segment is classified as "Urban and Built-up" land.
5. High Erosion: If the segment is not in Category 1, 2, 3, or 4, and at least one point within the segment had a land cover/use of "cropland" for 1997 with (usle + weq) > 2T for 1997 [where 'usle' represents the estimated sheet and rill erosion rate, 'weq' is the estimated wind erosion rate, and 'T' is the T-factor]
6. Cropland: If the segment is not in Category 1, 2, 3, 4, or 5, and at least one of the points had a land cover/use of either "cultivated cropland" or "non-cultivated cropland" for 1982 and/or 1987 and/or 1992 and/or 1997.
7. Pastureland: If the segment is not in Category 1, 2, 3, 4, 5, or 6, and at least one of the points had a land cover/use of "pastureland" for 1992 and/or 1997.
8. Rangeland: If the segment is not in Category 1, 2, 3, 4, 5, 6, or 7, and at least one of the points had a land cover/use of "rangeland" for 1997.
9. Forest Land: If the segment is not in Category 1, 2, 3, 4, 5, 6, 7, or 8, and at least one of the points had a land cover/use of "forest land" for 1997.
10. One Hundred Percent Urban or Water: If the segment is not in Category 1, 2, 3, 4, 5, 6, 7, 8, or 9, and 100% of the area of the segment is classified as "Urban and Built-up" plus "Water".
11. One Hundred Percent Federal or Water: If the segment is not in Category 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, and 100% of the area of the segment is classified as "Federal Land" plus "Water".
12. Remainder: All remaining segments

The conditional probabilities of selection were proportional to the "Weights" given in Table 3. The state sample sizes were based upon a variety of factors, including the importance of certain segment types for specified estimates, regional and state-level considerations, and budgets and workforce availability. Small states were generally sampled at higher rates than large states.

There were two major complicating factors in selection of sampling units for the Core Panel:

1. the 5,972 Special Study sample segments were specified as part of the Core Panel
2. the Core Panel was to include at least one sample segment in each HUCCO that contained any 1997 NRI sampling units, where a HUCCO is a geographical unit defined as the intersection of county and four-digit hydrologic units.

Table 3. Segment Weights for Annual NRI Sampling Strategy

Segment Category		Weight
1.	Wetland	3.4
2.	CRP	3.2
3.	Urban Change	3.2
4.	Urban	2.0
5.	High Erosion	2.0
6.	Cropland	1.5
7.	Pastureland	1.0
8.	Rangeland	1.0
9.	Forest Land	1.0
10.	100% Urban or Water	0.7
11.	100% Federal or Water	0.6
12.	Remainder	1.0

The first step was to select sample segments for "small HUCCOs". For HUCCOs containing only one 1997 NRI segment, that single sample segment was included in the Core Panel with certainty. For HUCCOs containing two 1997 NRI segments, both segments were included in the Core Panel with certainty if at least one of the segments was a Special Study segment or was in Segment Categories 1 – 5; otherwise, one segment was selected with each being given an equal chance of selection. For HUCCOs with three to nine 1997 NRI sample segments, selection rules were based upon the number of segments that fell within each of three groups – the first and second groups were comprised of Segment Categories 1 – 3 and 4 – 6 respectively, and the remainder were placed into the third group. Selection within each group for a small HUCCO was with equal probability, except for Special Study sample segments being included with certainty. In general, about one-half of the segments in the first group were selected for the Core Panel, a third for the second group, and a fourth for the third group.

For "large HUCCOs" with more than nine 1997 NRI sample segments, the selection procedure took into account the specified state sample size for the Core Panel, the number of 1997 NRI sample segments in each Segment Category, the weights in Table 1 for each Segment Category, the Special Study segments automatically placed into the state's Core Panel, and the sample segments selected for small HUCCOs. The process can be thought of as a systematic sampling procedure with the following characteristics:

- ordering or arrangement of eligible segments – the segments were ordered by Segment Category within HUCCO within county, with the Segment Category order reversed for adjacent HUCCOs
- the eligible segments were those that were not part of the Special Study sample and were within HUCCOs that contained at least ten 1997 NRI sample segments
- $(n - n_a - n_b)$ segments were selected, where n = total Core Panel sample size specified for the state, n_a = number of Special Study sample segments in the state, and n_b = number of sample segments in the state already selected for the Core Panel in small HUCCOs

- an unequal sampling interval was used that took into account the Segment Category weights and the distribution of n_a and n_b by Segment Category.

Selection of Segments for Supplemental Panel P01

The Supplemental Panel for 2001, called P01, only contained sample segments in large HUCCOs. The state selection process was a systematic sampling procedure that was quite similar to that used for the Core Panel. The conditional probability that a particular segment was selected for panel P01, given that the segment was not part of the Core Panel, was the same as for any other segment that was in the given Segment Category and was not in the Core Panel; the relative rates are the same as the Table 1 values used for selection of the Core Panel. This procedure caused selection probabilities to become slightly more equal. The probability that segment k in county q was in the 2001 Annual NRI sample was:

$$p_{1qik} = p_{0qik} + (1 - p_{0qik})\, T^{-1}\, S_i\, n_{2001}, \qquad \text{if segment is in a large HUCCO}$$

$$= p_{0qik}, \qquad \text{if segment is in a small HUCCO}$$

where: p_{0qik} = probability that the segment was selected for the Core Panel

S_i = the Segment Category weight given in Table 1 for category i

$T = \sum_i T_i$

$T_i = S_i * \check{Z}_i$

$\check{Z}_i$ = number of segments in category i in the state that are in large HUCCOs and not selected as part of P00

n_{2001} = number of segments specified for P01 for the state

The selection probabilities depend upon the county q because the Core Panel probabilities differ depending upon whether the county was included in the Special Study sample. The closed form for p_{0qik} is not provided here. Table 4 provides panel sample sizes for the 48 states by Segment Category, and Table 5 provides panel sample sizes by state. Panel P02 was selected using the procedure used for P01; the selection procedure for subsequent panels has taken into account numbers of samples already selected and remaining by Segment Category within state. Note that the sample for Louisiana was modified following the 2001 Annual NRI survey; two additional sample points were selected for each sample segment and sample sizes were reduced in order to gain efficiency. This modification to the Louisiana sampling strategy is the reason that Table 4 does not contain segment counts for P01.

Table 4: Annual NRI Sample Design – Panel Sizes for 48 Contiguous States, by Segment Category

	NRI Foundation Sample				
Segment Category	Large HUCCO's	Small HUCCO's	P00	P02	P03
1	42,162	429	9,810	7,416	7,477
2	9,836	144	2,960	1,604	1,653
3	22,705	165	5,202	4,010	4,184
4	17,050	139	2,614	2,100	2,069
5	15,747	183	2,603	2,003	1,995
6	60,777	803	7,427	5,753	5,909
7	15,401	204	1,198	1,038	1,028
8	30,511	431	2,901	2,313	2,312
9	34,739	501	2,558	2,253	2,234
10	16,957	167	923	805	810
11	22,813	370	1,300	1,101	1,119
12	2,046	22	148	150	139
Total	290,744	3,558	39,644	30,546	30,929

Table 2: Annual NRI Sample Design – Panel Sizes for 48 Contiguous States, by State

State	Available(*)	P00	P01	P02	P03	P04	P05
Alabama	6,033	778	622	622	600	608	608
Arizona	2,879	338	242	242	242	236	324
Arkansas	6,134	799	601	601	606	612	606
California	8,658	1,127	898	898	912	891	998
Colorado	7,453	959	691	691	718	704	721
Connecticut	1,213	261	199	199	194	202	201
Delaware	493	190	100	100	113	101	109
Florida	6,991	1,102	848	848	910	906	876
Georgia	7,928	836	764	764	751	743	808
Idaho	6,970	909	691	691	723	689	690
Illinois	8,387	1,264	936	936	995	988	955
Indiana	5,827	828	672	672	692	690	675
Iowa	7,060	996	750	750	741	736	738
Kansas	9,174	1,142	833	833	819	813	815
Kentucky	6,590	798	652	652	679	683	646
Louisiana (#)	14,435	1,117	2,458	932	835	834	837
Maine	2,661	377	323	323	312	317	332
Maryland	3,166	571	458	458	466	435	455
Massachusetts	1,900	368	257	257	310	259	259
Michigan	7,897	1,172	903	903	989	984	889
Minnesota	8,161	1,377	973	973	1,049	1,045	998
Mississippi	6,713	971	729	729	674	678	701
Missouri	8,706	1,214	911	911	963	966	924
Montana	6,348	823	652	652	662	650	674
Nebraska	7,382	972	728	728	756	763	744
Nevada	3,952	393	307	307	296	300	347
New Hampshire	1,686	261	239	239	260	246	243
New Jersey	2,038	322	278	278	242	248	284
New Mexico	5,166	782	593	593	560	594	635
New York	6,933	964	736	736	675	680	680
North Carolina	6,264	884	689	689	680	683	691
North Dakota	7,362	1,150	850	850	882	877	872
Ohio	6,874	952	723	723	683	689	678
Oklahoma	7,414	881	669	669	671	677	653
Oregon	5,905	674	601	601	670	604	609
Pennsylvania	7,243	1,035	765	765	752	749	756
Rhode Island	614	180	110	110	108	106	114
South Carolina	4,642	729	521	521	520	514	515
South Dakota	7,070	1,017	808	808	787	789	837
Tennessee	7,017	870	680	680	646	642	641
Texas	21,912	2,820	1,980	1,980	2,025	2,028	1,971
Utah	3,667	435	362	362	389	367	367
Vermont	2,023	294	256	256	274	245	258
Virginia	7,748	949	801	801	859	860	799
Washington	5,298	724	651	651	631	649	651
West Virginia	3,942	448	378	378	390	384	391
Wisconsin	6,587	1,067	783	783	814	808	777
Wyoming	3,788	524	401	401	404	413	402
TOTAL	294,304	39,644	32,072	30,546	30,929	30,685	30,754

***** Indicates number of segments in NRI Foundation Sample

\# Design for Louisiana modified in 2002; only 7,800 segments now available for selection

V. Sub-Sampling Procedure for NRI-CEAP

The target population for the NRI-CEAP Cropland Survey was all land in the 48 contiguous states that is classified by NRI as having a land cover/use of "cultivated cropland" or "land in CRP." Cultivated cropland is defined by NRI as "land in row or close-grown crops, including hayland and pastureland in rotation with row or close-grown crops;" land in CRP is "land that was under a Conservation Reserve Program (CRP) contract."

The sampling approach utilized for the NRI-CEAP Cropland Survey was to select a sub-sample of Annual NRI sample points. In particular, the sample comes from sampling units selected initially for the 2002 and 2003 Annual NRI surveys. The sampling strategy developed for the farmer surveys included:

- Collect data for 20,000 sample sites over a four year period, in order to obtain a full representation of the diversity of cropping systems, resource concerns, farming activities, conservation practices, soils, climate, and other natural resource conditions on cultivated cropland; and to obtain insight into implementation of conservation systems associated with the 2002 Farm Bill. [sample sites are cropland fields associated with NRI sample points; the Foundation NRI sample contains about 200,000 cropland points].
- Sampling and data collection for 2003 and 2004 were to focus on developing a good base-line for the most predominant cropping and conservation systems, to make sure that credible statistical analyses could be made on a national basis for all U. S. cultivated cropland.
- Sampling and data collection for 2005 and 2006 were to have a complementary focus: (a) to obtain data for areas and systems that are less extensive but usually more environmentally sensitive (vulnerable); and (b) to obtain data on actual changes in conservation systems and practices that occurred due to implementation of 2002 Farm Bill provisions – data collection in 2005 and 2006 provided a fuller and broader perspective, since some practices were not installed until after 2003.

An NRI sample point is used to identify a field in order to determine land cover/use and management systems; similar protocols are used to determine the natural or inherent features, such as soil type or erosion equation factors. The NRI utilizes points as the sampling units rather than farms or fields; land use and land unit boundaries change frequently in some parts of the country, and factors such as soil type do not follow human-induced boundaries such as land unit boundaries. Sample point coordinates are known based upon Digital Ortho-Photo Quadrangle (DOQ) base maps and standards. The temporal nature of desired results was handled in several ways: (i) the NRI-CEAP farmer survey collected site specific data for several years, and historical NRI data are available for each sample point; (ii) conservation practices, other agricultural management systems, and acts of nature have long-term effects upon the environment – the process models used to quantify effects produce results by year and season; (iii) the Annual NRI utilizes a supplemented panel survey design, wherein each year's sample includes a Core Panel (sampling units observed each year) and a Supplemental (or rotating) Panel – this provides the flexibility to revisit sample units over the course of time.

Sample for 2003 Survey

The sample for the 2003 NRI-CEAP Farmer Survey was selected from the 2002 Annual NRI sample points classified as having a land cover/use of either cultivated cropland or land in CRP for the 2002 growing season. In particular, the samples were selected from the supplemental panel P02, as follows:

(a) Any sample point in P02 classified as "land in CRP" for 2002 was included.
(b) Sample points classified as "cultivated cropland" were selected as follows:
    - it was determined which segments in P02 contained at least one point classified as "cultivated cropland" for 2002
    - within each of those segments, one point classified as "cultivated cropland" in 2002 was selected randomly.
(c) For South Dakota and North Dakota, one-half of these points were not sampled; systematic sampling was used to select half of the points. The sampling rate was reduced due to lack of available interviewers within these two states.
(d) An additional 333 points were removed from the sample because they represented farm operators that had also been selected for the ARMS-II survey. These samples were removed from the survey so that respondent burden for ARMS-II would not be affected. An initial examination of these overlap samples indicated that no bias should be expected; the samples were distributed across the country in proportion to cropland occurrence. This will be verified as part of a post-survey statistical evaluation of non-response, which will utilize historical NRI information and operator information collected from NRCS field offices.

Sample sizes by state are presented in Table 6. The sample included 2,236 CRP sample points and 9,580 cultivated cropland points.

Sample for 2004 Survey

The sample for the 2004 NRI-CEAP Cropland Survey was selected from the 2003 Annual NRI sample points classified as having a land cover/use of either cultivated cropland or land in CRP for the 2003 growing season. In particular, the samples were selected from the supplemental panel P03, as follows:

(a) Any sample point in P03 classified as "land in CRP" for 2003 was included.
(b) Sample points classified as "cultivated cropland" were selected as follows:
    - it was determined which segments in P03 contained at least one point classified as "cultivated cropland" for 2003
    - within each of those segments, one point classified as "cultivated cropland" in 2003 was selected randomly.

Sample sizes by state are presented in Table 6. The sample included 2,268 CRP sample points and 10,148 cultivated cropland points.

Sample for 2005 Survey

The sample for the 2005 NRI-CEAP Cropland Survey was selected from the 2003 Annual NRI sample points classified as having a land cover/use of either cultivated cropland or land in CRP for the 2003 growing season. In particular, the samples were selected from the Core Panel P00, as follows:

(a) Any sample point in P00 classified as "land in CRP" for 2003 was included.
(b) Sample points classified as "cultivated cropland" were selected as follows:

- it was determined which segments in P00 contained at least one point classified as "cultivated cropland" for 2003
- within each of those segments, one point classified as "cultivated cropland" in 2003 was selected randomly.

(c) The following randomization process was used to eliminate all cropland sample points in 10 states:
- Minnesota and Wisconsin were paired [placed in Stratum A]; each was given an equal chance of selection. Minnesota was kept in the sample and Wisconsin was selected for elimination.
- North Dakota and South Dakota were paired [placed in Stratum B]; each was given an equal chance of selection. South Dakota was kept in the sample and North Dakota was selected for elimination.
- The states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, and Connecticut were combined into a New England Grouping. New York and the New England Grouping were paired [placed in Stratum C]; each was given equal chance of selection. New York was kept in the sample and the New England Grouping was selected for elimination.
- The states of Montana, Colorado, Wyoming, Utah, and New Mexico were grouped [placed in Stratum D]; each was given an equal chance of selection. Colorado, Montana, and Utah were kept in the sample; Wyoming and New Mexico were selected for elimination.

(d) Sample sizes for cultivated cropland were reduced in 11 states, as follows:
- randomization techniques were utilized that reduced the sample by one-third in four states: Kansas; Minnesota; North Carolina; Ohio
- randomization techniques were utilized that reduced the sample by one-half in two states: South Dakota; Texas
- randomization techniques were utilized that reduced the sample by two-thirds in five states: Illinois; Indiana; Iowa; Missouri; Nebraska

(e) No cropland points in Florida, Nevada, and West Virginia were included for the 2005 survey; problems had been encountered in the 2003 and 2004 surveys. These three states were included for the 2006 survey.

Sample sizes by state are presented in Table 6. The sample included 3,893 CRP sample points and 7,489 cultivated cropland points. The sample size for cultivated cropland was about 25% less than for each of the earlier years; less funding was available for conducting farmer interviews.

Sample for 2006 Survey

The primary objective for sampling in 2006 was to provide a greater ability to make regional-level assessments (rather than just national), particularly by Major River Basin. Stratified sampling techniques were used to concentrate on fields in the most environmentally sensitive (or vulnerable) areas in order to provide more precise estimates of the effects of conservation in areas where the impacts of conservation are the greatest; sampling in 2003, 2004, and 2005 provided appropriate representation for predominant situations that covered 90% of the cropland base. Funding existed to conduct approximately 6,000 farmer interviews for cultivated cropland fields; no additional tracts of CRP land were selected.

Each county was ranked relative to its potential for soil and nutrient loss from cropland, by using the National Nutrient Loss and Soil Carbon (NNLSC) database which contains estimates based

upon EPIC model runs for 1997 NRI cropland sample points [see Potter et al (2006)]. The NNLSC database used general information on farming practices that was imputed onto the NRI cropland sample points. County level estimates were derived for: wind erosion, waterborne sediment, nitrogen loss in sediment, phosphorus loss dissolved in runoff, nitrogen loss dissolved in runoff, and nitrogen loss dissolved in leachate. County vulnerability rankings were derived using these seven factors as follows:

- A county was classified with vulnerability rank 1 if it had an estimated value for at least one factor in the top 10%; for wind erosion, the factor needed to be in the top 3% of all counties because 85% of all counties do not have significant cropland wind erosion. This category contained 658 counties.
- A county was classified with vulnerability rank 2 if it was not classified as vulnerability rank 1 but had an estimated value for at least one factor in the top 20% [top 5% for wind erosion]. This category contained 385 counties.
- A county was given a vulnerability rank 3 if its vulnerability could not be estimated from the NNLSC database and it contained at least 20,000 acres of cultivated cropland. This category included 70 counties.
- Counties with low and very low vulnerability according to these seven factors were given vulnerability ranks 4 and 5 respectively. There were 736 counties with rank 4 and 1,255 counties with rank 5.

The sample for the 2006 NRI-CEAP Farmer Survey came from 2003 Annual NRI sample points that had not been selected for previous farmer surveys. Each state and county had a different assortment of available cultivated cropland sample points relative to the county vulnerability rankings described above. The 2006 sample is not a stand-alone sample as are the samples for the three previous years. Some areas had no probability of selection for the 2006 survey; the 2006 results can only be used in conjunction with data collected for previous survey years.

For the 2003, 2004, and 2005 NRI-CEAP Farmer Surveys, sample points were spread out across states and counties as much as possible given the nature of the 2002 and 2003 Annual NRI samples. For example, only one cultivated cropland point per sample segment was selected for the farmer surveys; this spread out the sample and also greatly reduced the chance that the same farmer or operator was included in the sample more than one time in a given year. This was a restriction put in place following discussions with USDA-NASS and the Office of Management and Budget (OMB) in an effort to reduce respondent burden. For the 2006 sample, it was necessary to select some sample points in sample segments that had been used for the 2004 or 2005 sample.

One of the basic methods of sample selection for 2006 was as follows:

- determine which segments in P00 and P03 had at least two points classified as cultivated cropland in 2003
- if the segment had two points classified as cultivated cropland in 2003 and the county had vulnerability rank less than 4, select the sample point not used for either the 2004 or 2005 survey
- if the segment had three points classified as cultivated cropland in 2003 and the county had vulnerability rank less than 4, randomly select one of the two sample points not used for either the 2004 or 2005 survey
- no sample points were selected in counties with vulnerability rank 4 or 5.

This procedure was used for Alabama, Arizona, California, Colorado, Kentucky, Michigan, Mississippi, New Jersey, North Carolina, Oklahoma, Oregon, Tennessee, Utah, Virginia, and

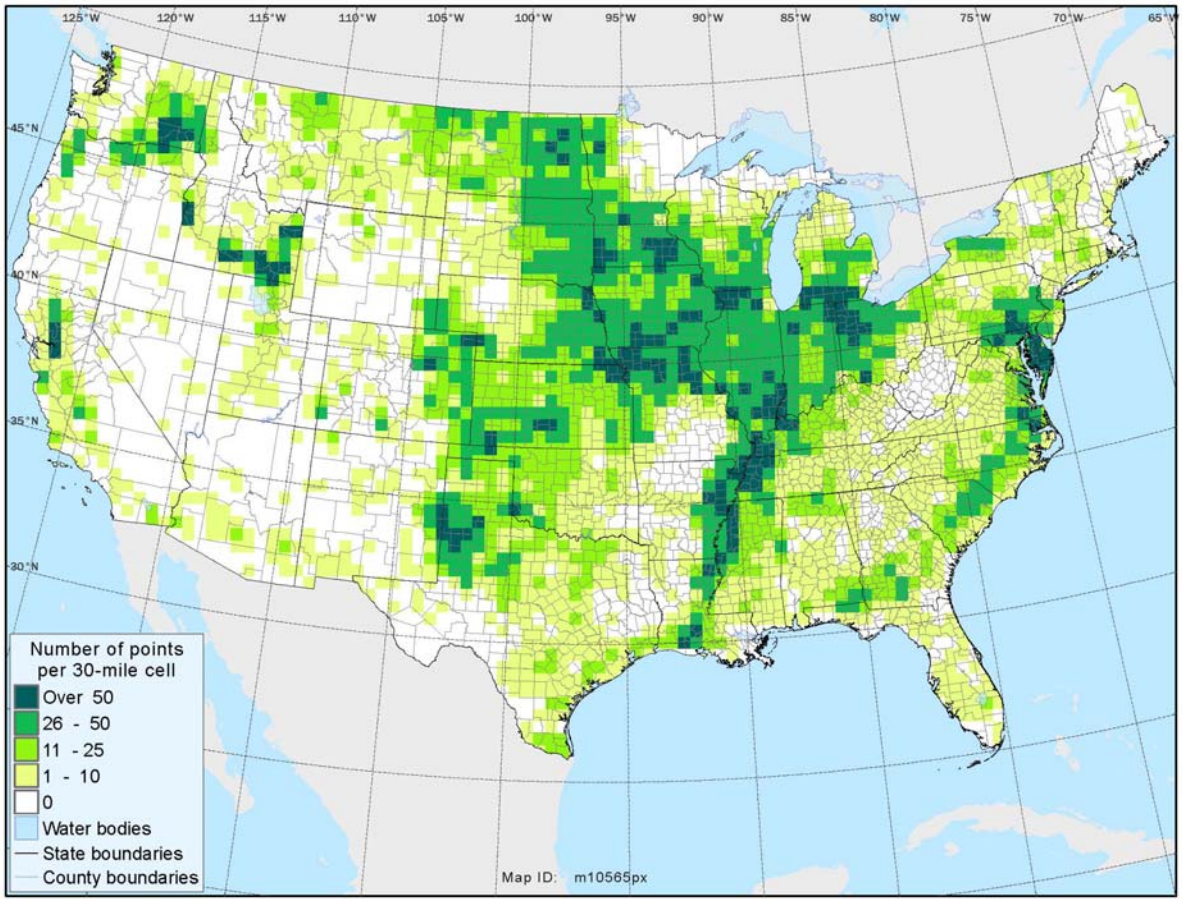
Washington. The modified procedure used for Arkansas, Georgia, Idaho, Louisiana, Maryland, Pennsylvania, and South Carolina was that only sample points from P03 were used.

Florida, Maine, Massachusetts, Nevada, New Mexico, Vermont, and West Virginia used sample points in all P00 segments not used for the 2005 survey. For Indiana, Iowa, and Nebraska, sample points were selected from all P00 segments not used for 2005 for counties with rank 1, and half in counties with rank 2; for Delaware, Missouri, North Dakota, Wisconsin, and Wyoming, all eligible P00 points were selected except only half in counties with rank >3. For Kansas and Texas, sample points were selected from all P00 segments not used for 2005 for counties with rank < 4; and sample points were selected from all eligible P03 segments in counties with rank 1, and half were selected for counties with rank 2 or 3. For Minnesota and South Dakota, all eligible sample points in counties with rank 1 or 2 were selected, and half of the P00 rank 4 or 5 sample points. For Connecticut, half of the P00 points were selected. For Illinois, sample points in P00 segments not used for 2005 were used in counties with rank 1; sample points were selected for half of the segments in counties with rank > 1. For Montana, all eligible sample points in counties with rank 3 were selected; sample points were selected from segments in half of the eligible P03 counties with rank 4 or 5. For Ohio, all eligible points in segments in counties with rank 1 and 2 were selected, except for half of the P03 segments with rank 2. For New York, all eligible points in segments in counties with rank 1 and 2 were selected, except for P00 segments with rank 2. No sample points were selected in New Hampshire and Rhode Island.

Table 6. Number of Sample Points Selected for NRI-CEAP Surveys, 2003 – 2006

	Cultivated Cropland					Land in CRP			
	2003	2004	2005	2006	Total	2003	2004	2005	Totals
Alabama	115	90	108	65	378	21	13	34	68
Arizona	35	30	41	43	149	0	0	0	0
Arkansas	184	194	264	150	792	17	6	13	36
California	166	166	149	181	662	10	3	6	19
Colorado	165	167	215	189	736	68	75	218	361
Connecticut	8	9	0	10	27	0	0	0	0
Delaware	54	61	110	60	285	1	0	1	2
Florida	55	45	0	65	165	4	5	0	9
Georgia	139	130	141	39	449	14	13	14	41
Idaho	235	245	275	143	898	90	109	182	381
Illinois	736	785	327	434	2,282	79	89	129	297
Indiana	484	484	198	201	1,367	25	38	49	112
Iowa	620	609	259	264	1,752	146	132	236	514
Kansas	527	500	442	164	1,633	196	195	322	713
Kentucky	163	194	188	138	683	33	19	60	112
Louisiana	186	184	193	130	693	26	5	16	47
Maine	4	4	0	14	22	1	0	0	1
Maryland	154	148	172	46	520	3	1	7	11
Massachusetts	5	11	0	12	28	0	0	0	0
Michigan	309	308	371	60	1,048	29	21	34	84
Minnesota	568	594	527	172	1,861	95	121	192	408
Mississippi	176	157	191	210	734	87	80	97	264
Missouri	388	400	243	218	1,249	159	235	293	687
Montana	160	197	215	83	655	101	95	221	417
Nebraska	463	497	211	249	1,420	99	121	166	386
Nevada	8	7	0	6	21	0	0	0	0
New Hampshire	6	5	0	0	11	0	0	0	0
New Jersey	42	38	46	7	133	0	0	0	0
New Mexico	80	85	0	100	265	79	50	106	235
New York	120	116	143	46	425	2	1	12	15
North Carolina	202	183	156	189	730	4	4	11	19
North Dakota	284	551	0	371	1,206	255	253	410	918
Ohio	375	370	325	167	1,237	31	18	29	78
Oklahoma	218	188	249	102	757	27	51	76	154
Oregon	118	140	108	124	490	34	36	47	117
Pennsylvania	230	184	244	59	717	2	2	7	11
Rhode Island	2	1	0	0	3	0	0	0	0
South Carolina	118	126	142	35	421	22	9	33	64
South Dakota	242	470	289	185	1,186	92	100	146	338
Tennessee	165	153	187	164	669	23	28	32	83
Texas	544	541	398	474	1,957	184	166	392	742
Utah	55	59	39	23	176	18	26	17	61
Vermont	19	30	0	29	78	0	0	0	0
Virginia	125	119	166	79	489	2	0	3	5
Washington	149	155	157	167	628	79	84	130	293
West Virginia	5	9	0	13	27	0	0	0	0
Wisconsin	326	355	0	301	982	49	40	105	194
Wyoming	48	54	0	51	153	29	24	47	100
All States	9,580	10,148	7,489	6,032	33,249	2,236	2,268	3,893	8,397

Figure 3. Density of Sample Points Selected for NRI-CEAP Surveys
[includes Cultivated Cropland and Land in CRP]



VI. Estimation Procedure

Introduction

The Annual NRI estimation procedure combines information from several sources to produce a final data set composed of records containing information for the years 1982, 1987, 1992, 1997, 2000, and annually thereafter. Each record represents data elements for a sample point; an estimation weight is attached to each record. For each NRI survey year, data are collected at both the segment level and at the point level. The areas measured for small water features, roads and railroads, and urban and built-up lands are converted to point data during the estimation process. Each of these created points is given an initial weight based on the area in the segment and the probability that the segment is included in the sample; imputation is used for unobserved data elements in order to complete the data record for these created points. Initial weights for created points and for observed points are adjusted during the estimation process using ratio adjustments and small area estimation. Control totals for surface area, federal land, and large water areas, derived from GIS databases, are maintained throughout the process. Finally, the weights are adjusted using iterative proportional scaling (raking) so that the new data base produces acreage estimates for broad cover/use categories for historical years that closely match previously published estimates [see Fuller (1999)].

Development of Estimation Weights for NRI-CEAP

Estimation weights for the NRI-CEAP cultivated cropland sample points in the Upper Mississippi River Basin (UMRB) were developed in a manner consistent with development of weights for the Annual NRI. Weights for other river basins will be developed in a similar fashion although some additional ratio adjustment procedures may be utilized, for example, for irrigated conditions. Estimation weights for points identified as "land in CRP" were basically those derived for the Annual NRI data base.

The procedure for points identified as cultivated cropland follows:

- Calculate initial weights, where $W_{Init,q,k,j}$ is the initial weight for point j, where point j falls within 6-digit hydrologic unit q and has cropping system k

  $W_{Init,q,k,j} = A_{q,k,,j} / (p_{q,k,j} * m_{q,k,j})$, where:

  $A_{q,k,j}$ = size of segment (q,k,j) in acres,
  $p_{q,k,j}$ = probability that segment (q,k,j) is in the sample,
  $m_{q,k,j}$ = number of sample points in segment (q,k,j)

- Make the first adjustment to the initial weights

  $W_{Adj1,\,q,k,j} = (W_{Init,\,q,k,j}) * (Y_k / X_k)$, where:

  Y_k = estimated acres of cultivated cropland in cropping system k for the UMRB area, based upon 2003 Annual NRI
  $X_k = \sum_{q,j} W_{Init,\,q,k,j}$

- Make the second adjustment to the initial weights

$W_{Adj2,\,q,k,j} = (W_{Adj1,\,q,k,j}) * (T_q / Z_{1,q})$, where:

T_q = estimated acres of cultivated cropland in 6-digit hydrologic unit q, based upon 2003 Annual NRI

$Z_{1,q} = \sum_{k,j} W_{Adj1,\,q,k,j}$

- Make the third adjustment to the initial weights

$W_{Adj3,\,q,k,j} = (W_{Adj2,\,q,k,j}) * (Y_k / X_{2,k})$, where:

$X_{2,k} = \sum_{q,j} W_{Adj2,\,q,k,j}$

- Make the fourth adjustment to the initial weights

$W_{Adj4,\,q,k,j} = (W_{Adj3,\,q,k,j}) * (T_q / Z_{3,q})$, where:

$Z_{3,q} = \sum_{k,j} W_{Adj3,\,q,k,j}$

- Make further iterations to force the adjusted weights to sum closer to the controls – for the UMRB, there were four additional adjustments using $\{Y_k\}$ and $\{T_q\}$
- Designate the final adjusted weight for point (q,k,j) to be the estimation weight, $W_{0,\,q,k,j}$

Development of Replicate Weights for Estimating Variances

A form of jackknife variance estimation is utilized for the Annual NRI because of the rather complex nature of the estimation procedure. The Annual NRI survey process is a type of two phase sampling, since the samples represent a subsample of segments selected from the 1997 NRI sample. The replication method used for the NRI is a form of the "delete-a-group jackknife" [see Kott (2001)]. The goal of the variance estimation procedure for an Annual NRI data set is to construct a set of H modified weights for each observation, which allows computation of H replicate estimates for a variable y. A variance estimate can then be calculated for an NRI estimate, say $\hat{Y}$, as follows:

$\text{var}(\hat{Y}) = \sum_h c_h * (\hat{Y}_h - \ddot{Y})^2$, where

c_h is a constant determined by the replication procedure

$\hat{Y}_h$ is the h^{th} replicate estimate for Y, and

$\ddot{Y} = H^{-1} \sum_h \hat{Y}_h$

For the 2003 Annual NRI and the NRI-CEAP cropland survey, H = 29 is used. To define the replicates, a form of systematic sampling was used with the 1997 NRI sample units to create 29 groups of samples of approximately equal size. The same set of replicates is used for both the 2003 Annual NRI and the NRI-CEAP cropland database. This means that an estimation process can be established so that variance estimates based upon the larger sample can be retained within the smaller data base, if certain regression and/or ratio techniques are utilized.

The first set of replicate weights for the NRI-CEAP data set is derived as follows:

- Calculate initial weights for the point (q,k,j) by modifying the estimation weight, $W_{0,\,q,k,j}$, as follows:

$$W_{Init,1,q,k,j} = 0 , \quad \text{if point (q,k,j) is in replicate \#1}$$
$$= (29/28) * W_{0,\,q,k,j} , \quad \text{otherwise}$$

- Make the first Adjustment to the Initial Weights

$W_{Adj1,1,q,k,j} = (W_{Init,\,1,q,k,j}) * (Y_k / X_{1,k})$, where:

Y_k = estimated acres of cultivated cropland in cropping system k for the UMRB area, based upon 2003 Annual NRI

$X_{1,,k} = \sum_{q,j} W_{Init,\,1,q,k,j}$

- Make the second adjustment to the initial weights

$W_{Adj2,\,1,q,k,j} = (W_{Adj1,1,\,q,k,j}) * (T_q / Z_{1,q})$, where:

T_q = estimated acres of cultivated cropland in 6-digit hydrologic unit q, based upon 2003 Annual NRI

$Z_{1,q} = \sum_{k,j} W_{Adj1,\,1,q,k,j}$

- Make the third adjustment to the initial weights

$W_{Adj3,1,\,q,k,j} = (W_{Adj2,1,\,q,k,j}) * (Y_k / X_{1,2,k})$, where:

$X_{1,2,k} = \sum_{q,j} W_{Adj2,1,\,q,k,j}$

- Make the fourth adjustment to the initial weights

$W_{Adj4,\,1,q,k,j} = (W_{Adj3,\,1,q,k,j}) * (T_q / Z_{1,3,,q})$, where:

$Z_{1,3,q} = \sum_{k,j} W_{Adj1,1,\,q,k,j}$

- Make further iterations to force the adjusted weights to sum closer to the controls – for the UMRB, there were four additional adjustments using $\{Y_k\}$ and $\{T_q\}$

- Designate the final adjusted value for point (q,k,j) to be the first replicate weight, $W_{1,\,q,k,j}$

A similar process is used for each of the remaining 28 replicates. Each point (q,k,j) then has an estimation weight, $W_{0,\,q,k,j}$, and a set of 29 replicate weights, $\{ W_{h,\,q,k,j} : h=1,2, \ldots, 29 \}$, that are used for variance estimation.

VII. Other Considerations

Preliminary results for the NRI-CEAP National Cropland Assessment are based upon data from less than 60 percent of the sample sites originally selected for the NRI-CEAP Cropland Survey. NASS enumerators were unable to complete questionnaires for about 25 % of the sample sites; data from additional sample sites contained inconsistencies that could not be resolved for inclusion in the preliminary results. A series of statistical analyses will be performed to test the effects of these missing observations and to determine if a modified weighting process should be employed. Other tests will be developed to analyze the effects of small sample sizes for relatively small but influential and/or sensitive areas. Additional work is also needed to develop methods for quantifying uncertainty due to the modeling employed for this project.

The expectation is that follow-up farmer surveys will be conducted starting in the year 2011, in order to account for ongoing changes in cropping and conservation systems. In preparation for this new series of surveys, several aspects of the 2003 – 2006 survey operations will be examined. This includes development of an automated survey instrument as part of the quality assurance process, utilization of improved sample location materials, and examination of response burden including finding methods to shorten the length of each farmer interview.

References

Breidt, F.J. & W.A. Fuller (1999) Design of supplemented panel surveys with application to the National Resources Inventory, *Journal of Agricultural, Biological, and Environmental Statistics*, 4(4): 391 – 403.

Fuller, W.A. (1999) Estimation procedures for the United States National Resources Inventory, *Proceedings of the Survey Methods Section of the Statistical Society of Canada*, 39 – 44.

Fuller, W.A. & F.J. Breidt (1998) Estimation for supplemented panels, Sankhya: The Indian Journal of Statistics, 61: 58 – 70.

Goebel, J. J. (1998) The National Resources Inventory and its role in U.S. agriculture, *Agricultural Statistics 2000*, International Statistical Institute, Voorburg, The Netherlands, 181-192.

Goebel, J.J. & H.D. Baker (1987) *The 1982 National Resources Inventory Sample Design and Estimation Procedures,* Statistical Laboratory, Iowa State University, Ames, IA.

Goebel, J.J. & R.L. Kellogg (2002) Using survey data and modeling to assist the development of agri-environmental policy, *Conference on Agricultural and Environmental Statistical Applications in Rome,* National Statistical Institute of Italy, Rome, Italy, 695–704.

Kim, J.K., A. Navarro & W.A.Fuller (2006) Replication variance estimation for two-phase stratified sampling, *Journal of American Statistical Association*, 101: 312- 320.

Kott, P.S. (2001) The delete-a-group jackknife, *Journal of Official Statistics*, 17: 521 – 526.

McDonald, T.L. (2003) Review of environmental monitoring methods: survey designs, *Environmental Monitoring and Assessment*, 85: 277 – 292.

Nusser, S.M., F.J. Breidt & W.A. Fuller (1998) Design and estimation for investigating the dynamics of natural resources, *Ecological Applications*, 8(2):234-245.

Nusser, S.M. & J.J. Goebel (1997) The National Resources Inventory: a long-term multi-resource monitoring programme, *Environmental and Ecological Statistics*, 4(3):181- 204.

Potter, S.R., S. Andrews, J.D. Atwood, R.L. Kellogg, J. Lemunyon, L. Norfleet, D. Oman (2006) *Model Simulation of Soil Loss, Nutrient Loss, and Change in Soil Organic Carbon Associated with Crop Production*, Natural Resources Conservation Service, USDA, Washington, D.C.