United States Department of Agriculture

NRCS Natural Resources Conservation Service

1997 National Resources Inventory
(revised December 2000)

A GUIDE FOR USERS OF 1997 NRI DATA FILES

CD-ROM Version 1

(December 2001)

NRI

NRCS is committed to making its data products accessible to all USDA customers and employees. If you are unable to read any of the documentation included on this CD, please refer to **Appendix E** for help.

1997 National Resources Inventory
(revised December 2000)

A GUIDE FOR USERS OF 1997 NRI DATA FILES
Version 1
(December 2001)

Table of Contents

Page

Appendices

Sections in **bold** (blue) in Table of Contents contain important information regarding statistical reliability.

1. Overview

1.1. Introduction

The National Resources Inventory (NRI) is a scientifically-designed, longitudinal panel survey of the Nation's soil, water, and related resources designed to assess conditions and trends every five years. The NRI is conducted by the Natural Resources Conservation Service (NRCS), an agency of the U.S. Department of Agriculture (USDA), in cooperation with the Iowa State University (ISU) Statistical Laboratory. The 1997 NRI contains data on nonfederal lands and water areas within the 48 conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands. The 1997 NRI provides results that are nationally consistent for all nonfederal lands for four points in time – 1982, 1987, 1992, and 1997.

The NRI data can be quite useful in developing statistical estimates of natural resource conditions, and in conducting geospatial and temporal analyses of these conditions. The data come from a sample survey and are different in nature than most data used in geographic information system (GIS) work. This statistical nature of the data is important to remember when using the NRI to produce tabulations and analyze resource conditions and trends.

This document serves as an initial compilation of material developed to assist and enhance use of NRI data. It provides users of NRI data with basic information regarding the inventory process, the database, and proper utilization of the data.

This user guide is referenced as Version 1. A second version of this document will be developed if it is found that additional documentation will enhance utilization of the data.

Users are encouraged to provide feedback as outlined in Appendix E.

1.2. Basic Concepts

The NRI database is a statistical database. The data do **not** come in a summarized or aggregated form. Users of the database decide which portions of the data are needed, and how to aggregate and summarize the data most effectively. The data must be aggregated to develop meaningful statistics and analyses. To use the individual sample point data as other than a part of the statistical database is inappropriate; for example, it is a misuse to use the NRI database to attempt to identify a specific farm and field where certain conservation actions should be initiated.

Tabulations made from the NRI database are estimates. These tabulations produce estimates rather than facts because they are based upon sample data derived from a survey – rather than data coming from a census, or complete measurement. Each estimate has some degree of statistical uncertainty associated with it; this statistical uncertainty affects analyses of the data and interpretation of results. Statistical considerations are discussed in Sections 1.4, 4.2, and 6.5.

All data summaries and analyses must take into account the database weights (expansion factors) and at least some type of geographical factors. Proper weighting of NRI data is discussed in Sections 4.4 and 6.5. The statistical weights are the data elements (variables) *xfact* found in the table **point** (see Section 5.2 and Appendix A). Geographic factors include State, county, hydrologic unit area (HUA), and Major Land Resource Area (MLRA); various combinations of these can be used.

The database has been structured so that it is easy to statistically analyze the data. All NRI data were consolidated into a database with a straightforward point-type format; this is discussed in Chapter 4, Construction of the NRI Database. Many characteristics have been associated with each sample point. This means that the statistical nature of the database allows aggregation and cross-tabulation of the specific point data over meaningful entities (domains). Also, it is easy to link or associate information from other databases with the NRI data because of this structure.

The NRI data are organized into nine logical subject-matter tables (data files). Almost all data tabulations and analyses will require data from two or more of these tables. Data from the various tables are linked (matched up) using the unique key variable *recordid* (see Section 5.5). Every point has a unique *recordid* associated with it, and each table contains the *recordid* variable (field).

- As an example, consider estimation of the amount of nonfederal Forest Land that has become Developed Land between 1982 and 1997 in the Northwest (say, Oregon and Washington). This requires using the data fields *fips* (geographic factor) and *xfact* (statistical weight) from the table **point**, as well as *yr* (year) and *broad* (broad land cover/use) from the table **trend**. Conceptually, data would be extracted for all points matching the query that:

  (1) *fips* corresponds to a county in either Oregon or Washington,
  (2) *broad* is the Forest Land code when *yr* is 1982 in the table **trend**, and
  (3) *broad* is one of the Developed Land codes when *yr* is 1997 in the table **trend**.
  The *xfact* variable from the table **point** would be extracted and summed for all points meeting all three of these criteria.

It is strongly recommended that users of the NRI database structure their tabulations in a way that makes it easy to check out new results with "established" results. This is because it is easy to mis-specify or incorrectly structure tabulations and analyses using NRI data. Tabulations can initially be checked against estimates released by NRCS in the *Summary Report, 1997 National Resources Inventory (Revised December 2000)* (USDA, 2000) (hereafter referred to as the *1997 NRI Summary Report*) and on the NRI Web site (see Appendix B). As a data user's queries become more sophisticated, this checking can be against results that were previously checked against already established results.

- Consider again the above example – estimating the amount of nonfederal Forest Land that has become Developed Land between 1982 and 1997, for the States of Oregon and Washington. Rather than just extracting data for certain points as outlined above, it would be best to do a fairly complete cross-tabulation – this would allow replication and checking with estimates that were already derived from the 1997 NRI database (see, for example, the State-level figures for Oregon and Washington in Tables 1 and 2).

Tabulations and analyses are typically performed in an iterative fashion. The iterations involve forming different aggregations of the data, both by changing combinations of element categories and by using additional factors to better define and refine the results.

- Consider again the above example – estimating the amount of nonfederal Forest Land that has become Developed Land between 1982 and 1997, for the States of Oregon and Washington. After looking at several cross-tabulations that provide aggregate estimates of these conversions, it would be appropriate to make further subdivisions and find useful/interesting "messages" in the data. Examples include: types of Forest Land being converted, comparisons between the three 5-year time periods (1982 – 1987, 1987 – 1992, and 1992 – 1997), quality of the land being converted, geographical distribution of the converted lands as well as those not converted, etc. Analysis might involve looking at several of these factors simultaneously, in addition to making comparisons with changes in human population over these periods of time.

The 1997 NRI database contains data for "imputed" points, or "pseudo points." The pseudo points should be treated like all other points when performing tabulations and analyses with the NRI data – they need to be included because of the way the database was constructed. Imputation procedures and construction of pseudo points are discussed in Sections 4.3, 4.4, and 4.5.

The 1997 NRI database contains nationally consistent data for four points in time: 1982, 1987, 1992, and 1997. All comparisons for two points in time need to be made using this current NRI database. Comparisons using data published in 1982, 1987, and/or 1992 can produce incorrect inferences, because of changes in the statistical estimation procedure, and because all data were simultaneously reviewed (edited) as 1997 NRI data were collected.

- As an example, consider estimated average annual sheet and rill erosion rates for Cultivated Cropland in Wisconsin. Based upon the 1992 NRI, the estimates for the years 1982 and 1992 were 5.2 and 4.1 tons per acre per year, respectively. Based upon revisions made to the raw data during 1997 NRI data collection, the corresponding estimates derived from the 1997 NRI database are 4.7 and 3.8 tons per acre per year. These revisions result from updated technical guidance regarding use of the Universal Soil Loss Equation (USLE) – these revisions are applied to historical data so that trends reported using NRI data reflect true changes in resource conditions and not just changes in inventory methodology.

It is important to understand the full nature of each data element used in tabulations and analyses. Section 1.3 provides insights and guidance regarding several important components of the NRI database. Data collection methodology is discussed in Section 3.2; detailed data collection instructions and a glossary are available in other files on the CD (see Appendix A). Definitions, protocols, and instructions differ in some cases from those used by other agencies. These differences need to be considered when analyzing and interpreting the data.

1.3. SPECIAL CONSIDERATIONS

Land Cover/Use: All tabulations and analyses need to account for "Land Cover/Use." NRI uses this term to identify a classification system that accounts for all the surface area of the U.S., using mutually exclusive categories. As the name implies, Land Cover/Use is a hybrid classification system that takes into account both land cover and land use. The NRI database includes both Specific Cover/Use and Broad Cover/Use as variables in the database; they are the data elements *landuse* and *broad* in the table **trend**. There are 12 Broad Cover/Use categories:

(1)	Cultivated Cropland	(7)	Urban & Built-up Land
(2)	Noncultivated Cropland	(8)	Rural Transportation Land
(3)	Pastureland	(9)	Small Water Areas
(4)	Rangeland	(10)	Census Water
(5)	Forest Land	(11)	Federal Land
(6)	Other Rural Land	(12)	Conservation Reserve Program (CRP) Land

There are 64 Specific Land Cover/Use categories. The NRI data collectors determined the Specific Cover/Use category; the Broad Cover/Use category was derived from the Specific Cover/Use category. The system is hierarchical except for some types of Hayland and Pastureland; if the hay or pasture is in rotation with cultivated crops, the point is classified as "Cultivated Cropland" rather than "Noncultivated Cropland" or "Pastureland." The 1997 NRI database contains both a Specific Cover/Use and Broad Cover/Use for each sample point, for each of the four inventory years – 1982, 1987, 1992, 1997.

Note that "Federal Land" is one of the Broad Cover/Use categories. Points classified as "Federal Land" are included in the database; they have the statistical weight *xfact* and various geographical codes but almost no other attributes. They are included in the database for completeness – to make a complete accounting of all U.S. surface area. Note also that there are numerous areas in the U.S. that have either become Federal Land or gone from Federal to private during the period 1982 to 1997. The 1997 NRI database accounts for all of these dynamic changes.

It is recommended that tabulations and analyses start with Broad Cover/Use as one of the baseline variables. That serves several purposes. It will make it possible to check new results against previous (established) tabulations of the data, to make sure that the data are being queried and tabulated properly. It is also helpful to look at the data this way to uncover idiosyncrasies that may not have been apparent during previous analysis of the data. It is easy to misuse the data without taking into account the portion of the landscape that is Federal Land, or Water Area, or often Developed Land. Certain data elements (like Prime Farmland designation) do not apply to certain areas – this needs to be accounted for when tabulating the data, when making certain inferences, and when presenting results to policy makers, scientists, and the general public.

Developed Land: Developed Land contains three Land Cover/Use categories:

(1) Large Urban & Built-up Areas (greater than 10 acres in size)
(2) Small Built-up Areas (less than 10 acres in size)
(3) Rural Transportation Land (roads, railroads, and associated rights-of-way)

As discussed in Section 1.2, the 1997 NRI database contains data that are nationally consistent for 1982, 1987, 1992, and 1997. Only the 1997 NRI database should be used to analyze issues dealing with conversion of Rural Land to Developed Land. It is confusing (and incorrect) to compare the 1997 NRI results with results published for the earlier inventories. For example: (1) Urban & Built-up Land figures published for the 1982 NRI only included the category Large Urban & Built-up Areas; Small Built-up Areas were included in a category called Minor Land Cover/Use; (2) the 1987 NRI Developed Land category included Rural Transportation Land in addition to Large Urban & Built-up Areas; Small Built-up Areas were still included under Minor Land Cover/Use; (3) for the 1992 NRI, Developed Land also included the Small Built-up Areas; (4) for the 1997 NRI, inventory protocols were changed for measurement of Rural Transportation Land, which required statistical adjustment of figures for the years 1982, 1987, and 1992. (*Note:*

Each time a change was made either in reporting categories or inventory protocols, appropriate changes were made in the historical data so that the newest inventory contained data that were comparable over all inventory years.)

For the NRI, Developed Land areas are considered (permanently) removed from the Rural Land base. Other agencies and groups that collect seemingly similar data have different reasons for collecting data. This means that these other sources of data will show both differences and similarities with NRI data. Thorough examination and analysis of the data from two differing sources can be beneficial and strengthen understanding of the issues, as long as both studies use credible inventory procedures.

> *Additional documentation regarding Developed Land is provided on the NRI web site,* http://www.nhq.nrcs.usda.gov/NRI/1997/national_results.html. *Included are discussions on: the amount of uncertainty associated with various Developed Land estimates, how to properly use and interpret these figures, and how to minimize misuse of these estimates. The documentation also provides thousands of estimates and their associated margins of error.*

Erosion Estimates: Erosion data are a unique and important part of the NRI. The erosion data can not be used to compute the erosion that actually occurred during a particular year; instead, the database contains "factors" that are used as inputs into two different erosion prediction equations (models):

(1) Universal Soil Loss Equation (USLE) – an erosion model designed to predict long-term average annual soil loss (due to sheet and rill erosion) from specific field areas in specified cropping and management systems. *Sheet and rill erosion* is the removal of layers of soil from the land surface by the action of rainfall and runoff; it is the first stage in water erosion. The equation is: A = RKLSCP, where:

- A = Computed (estimated) soil loss per unit area
- R = *Rainfall and runoff* factor
- K = *Soil erodibility* factor
- L = *Slope-length* factor
- S = *Slope-steepness* factor
- C = *Cover and management* factor
- P = Support *practice* factor

(2) Wind Erosion Equation (WEQ) – an erosion model designed to predict long-term average annual soil losses (due to wind erosion) from a field having specific characteristics. *Wind erosion* is the process of detachment, transport, and deposition of soil by wind. The equation is: E = f(I,K,C,L,V), where:

- E = Computed (estimated) soil loss per unit area
- I = *Soil erodibility index*
- K = Soil *ridge roughness* factor
- C = *Climatic* factor
- L = Equivalent *unsheltered distance*
- V = Equivalent *vegetative cover*

Erosion rates computed from the data are estimated average annual (or expected) rates based upon the cropping practices, management practices, and inherent resource conditions. Climatic factors for both USLE and WEQ are based upon long-term average conditions and not upon one year's actual events.

The 1997 NRI database contains both computed (estimated) soil loss and the individual factors, for both the USLE and WEQ, for all points that are Cropland, Pastureland, or CRP land in a given year. Erosion data are not given for points that are any other land cover/use. If a sample point changes land cover/use between two points in time, it has erosion equation factors for the years it is Cropland, Pastureland, or CRP land – but not for any years that is some other land cover/use. This is an important factor to keep in mind when trying to estimate erosion rates for a particular area – to only account for those sample points with a land cover/use of Cropland, Pastureland, or CRP land. It is incorrect to average USLE rates over the land area of an entire State, rather than just some portion of the agricultural land.

Calculating average annual estimated erosion rates for particular areas and types of land can be confusing. The proper method is to take "weighted averages" over the appropriate set of NRI sample points (see Section 6.5). It is also confusing when making comparisons between two points in time – to take into account those sample points with changes in land cover/use, and calculate these averages over a slightly different set of sample points for the two years.

The NRI database contains all of the individual USLE and WEQ factors because special use can be made of these factors. Erodibility Index (EI) is a good example – this index is a numerical expression of the potential of a soil to erode, considering climatic factors and the physical and chemical properties of the soil (the higher the index, the greater is the investment needed to maintain the sustainability of the soil resource base if intensively cropped). These factors are also useful, for example, for analyses of how erosion rates might change if certain cropping practices are changed. Both of these examples relate to use of the NRI database for formulation of provisions of the 1985 Food Security Act.

1.4. STATISTICAL CONSIDERATIONS

Results derived from the NRI database are estimates – not absolute facts. All NRI results have some degree of uncertainty associated with them. Interpretation of NRI results requires an understanding of the amount of this uncertainty associated with each estimate, as well as an understanding of the inventory procedures. Since the NRI employs recognized statistical methodology, it is possible to quantify this statistical uncertainty.

NRI data were collected at more than 800,000 sample sites nationwide (see Sections 3.1 and 3.2). This is a very large sample, which means that the data can be legitimately used to analyze issues at many geographic levels — national, regional, State, and sub-State (multi-county). However, the NRI was **not** designed for analyzing issues at the county level (see discussions below).

The precision of NRI estimates depends upon the number of samples within the region of interest, the distribution of the resource characteristics across the region, and the sampling procedure. Characteristics that are common and spread fairly uniformly over an area can be estimated more precisely than characteristics that are rare or unevenly distributed.

The NRI was originally designed (in 1982) to obtain natural resource data usable for analysis at a sub-State (multi-county) level. Of interest were estimates that could be developed for such sub-State entities as Major Land Resource Areas (MLRA), Soil Conservation Service Administrative Areas, 4-digit Hydrologic Unit Areas (HUA), and Water Resources Council aggregated sub-areas (ASA). The sample was selected giving particular attention to MLRA portions of States. Note, however, that it was usually too expensive to sample an MLRA of much less than one million acres at a rate sufficient to meet the NRI standards.

> *A "Major Land Resource Area" (MLRA) is a broad geographic area characterized by a particular pattern of soils, climate, water resources, vegetation, and land use.*
>
> *A "Hydrologic Unit Area" (HUA) is a drainage area defined by a hierarchical classification system developed by the U.S. Geological Survey. The system divides the area of the U.S. and Caribbean into 21 major regions, 222 subregions, 352 accounting units, and 2,150 cataloging units that delineate river basins (watersheds) having drainage areas usually greater than 700 square miles. Major regions and subregions are referred to as 2-digit and 4-digit HUAs, respectively; accounting and cataloging units are called 6-digit and 8-digit HUAs, respectively.*

This design criterion can also be expressed quantitatively. The sample was selected so that the margin of error was less than 20 percent for any estimate of a resource condition that comprised at least 10 percent of the MLRA land area of the State. Most items could be estimated more precisely. This criterion also holds for estimates derived from the 1997 NRI database (Nusser and Goebel 1997, Goebel 1998). Note that estimates of change between two points in time will be less precise (relatively) because the changes will be occurring on a smaller fraction of the landscape.

The "margin of error" is a commonly used measure of reliability and can be used to construct a 95 percent confidence interval for an estimate. The lower bound of the confidence interval is obtained by subtracting the margin of error from the estimate; adding the margin of error to the estimate forms the upper bound. The margin of error is calculated by multiplying the standard error by the factor 1.96.

These measures of uncertainty (margins of error, standard errors, or confidence intervals) should be taken into consideration in all data analyses. Data users who have limited experience in using sample survey data are advised to seek professional assistance regarding the statistical implications of the data. As discussed in Section 6.7, it is advisable to work in teams when utilizing the NRI for a scientific project. This analysis team should include someone who has statistical analysis skills, can explain and interpret statistical uncertainty and sampling error, and can offer guidance for the generation and interpretation of the NRI estimates.

The NRI design objectives should not be thought of as a restriction to full analytical use of the NRI database. Examining data at various geographic levels is a legitimate analytical technique – it is a particularly valuable way to gain additional insights into the database, and in discovering "messages" that can be pulled from the data. Users of the 1997 NRI database can initially analyze the data at a fairly localized level, paying close attention to the margins of error of the estimates at those localized levels. Data can then be aggregated to higher levels for display or presentation purposes; estimates at these aggregated levels would have sufficiently narrow confidence intervals so that results would be appropriate for presentation and discussion.

1.5. CONFIDENCE INTERVALS – EXAMPLES

Consider the following example.

- The estimated amount of Cultivated Cropland for Kentucky in 1997 is 3,514,000 acres. The estimated standard error for this estimate is 66,600 acres. This error multiplied by a factor of 1.96 is 130,500, which is the estimated margin of error. (Note: the margins of error rather than the standard errors are given in Table 1.) A 95 percent confidence interval for this estimated acreage is constructed as:

  Lower bound = (3,514,000 – 130,500) acres = 3,383,500 acres
  Upper bound = (3,514,000 + 130,500) acres = 3,644,500 acres

  This means a 95 percent certainty exists that the true amount of Cultivated Cropland is between 3,383,500 and 3,644,500 acres.

As mentioned above, estimates can be derived for many different geographical units. Notice in the following table how the "reliability" of estimates varies across geographic units in a predictable manner.

Reliability of Cultivated Cropland Acreage Estimates, 1997

Region	Estimated acres	Margin of error (acres)	Margin of error (%)
U.S.	326,783,700	1,838,300	0.6
Kentucky	3,514,000	130,500	3.7
MLRA 122 (KY)	1,075,900	76,200	7.1
Grayson County, KY	45,500	21,000	46.1

Examination of confidence intervals is necessary to interpret the data correctly. Misleading conclusions can be reached if confidence intervals are not considered. As an illustration of how confidence intervals affect what statements can be made, consider Table 1 at the end of this chapter. There were significant net changes (gains and losses) in Cultivated Cropland for many States, for the period 1982 to 1997. For some States, the estimated gains or losses are negligible, i.e., not statistically significant. For example, for Illinois, there is an estimated net change of + 60,600 acres, but the estimated margin of error is ± 95,600 acres; and for Maine the estimated net change is -5,600 acres, with an estimated margin of error of ± 29,800 acres. It is appropriate to describe the net changes in these two States by stating there were *"no net changes in Cultivated Cropland in Illinois and Maine."* This statement is true because the 95 percent confidence intervals are -35,000 to 156,200 for Illinois and -35,400 to 24,200 for Maine. This means that the observed changes were so close to zero that we can not be certain if there were small gains or small losses in those two cases.

The precision of all statistical results should be derived and examined. The 1997 NRI database has been constructed in a manner that provides methods for deriving estimated margins of error (or standard errors, or confidence intervals) for any estimate derived from the database.

Table 1. Estimated acres of Cultivated Cropland and Developed Land, with estimated margins of error in parentheses

State	Cultivated Cropland, 1997		Change in Cultivated Cropland, 1992 to 1997		Developed Land, 1997		Change in Developed Land, 1992 to 1997	
	----- 1,000 acres -----							
Alabama	2,614.1	(155.6)	-189.8	(64.3)	2,252.3	(139.7)	315.3	(31.6)
Arizona	982.2	(123.7)	-34.5	(37.2)	1,491.4	(286.4)	113.8	(35.9)
Arkansas	7,362.4	(257.0)	-74.1	(58.0)	1,409.1	(130.7)	168.9	(34.7)
California	6,219.7	(558.8)	-337.5	(183.3)	5,456.1	(303.0)	553.4	(76.6)
Colorado	7,567.4	(443.7)	-107.6	(128.0)	1,651.7	(149.7)	112.5	(28.2)
Connecticut	81.0	(20.2)	-12.9	(13.3)	873.9	(46.6)	39.4	(7.3)
Delaware	478.4	(35.5)	-14.7	(9.0)	225.5	(22.7)	23.1	(7.4)
Florida	1,480.3	(184.8)	-213.4	(87.2)	5,184.8	(224.8)	825.2	(91.3)
Georgia	4,174.8	(186.8)	-539.4	(91.1)	3,957.3	(145.4)	851.9	(58.4)
Hawaii	198.0	(35.7)	-30.6	(16.9)	179.7	(35.3)	6.8	(2.9)
Idaho	4,541.3	(254.8)	-113.9	(145.0)	754.9	(64.3)	91.9	(21.6)
Illinois	23,563.4	(266.4)	**60.6**	**(95.6)**	3,180.9	(125.8)	246.5	(37.8)
Indiana	12,761.4	(214.6)	-162.4	(85.1)	2,260.4	(97.0)	195.3	(31.9)
Iowa	24,216.5	(283.8)	239.4	(124.7)	1,702.1	(86.8)	69.1	(17.1)
Kansas	24,793.6	(421.6)	-293.2	(123.9)	1,939.9	(97.4)	96.5	(20.0)
Kentucky	**3,514.0**	**(130.5)**	-109.6	(103.9)	1,737.5	(94.5)	237.1	(24.5)
Louisiana	5,470.1	(172.7)	-323.1	(67.0)	1,623.8	(61.5)	133.6	(16.9)
Maine	154.7	(43.5)	**-5.6**	**(29.8)**	712.0	(76.2)	111.1	(17.8)
Maryland	1,412.8	(70.6)	-85.8	(28.4)	1,235.7	(52.7)	177.6	(19.0)
Massachusetts	64.7	(20.4)	-11.7	(14.5)	1,479.2	(79.2)	211.8	(19.2)
Michigan	6,558.8	(200.9)	-558.1	(109.4)	3,545.5	(134.1)	364.1	(36.8)
Minnesota	19,731.2	(395.3)	111.8	(173.9)	2,185.5	(112.9)	231.8	(39.8)
Mississippi	4,931.5	(161.5)	-550.2	(79.4)	1,474.0	(101.5)	206.4	(26.5)
Missouri	10,517.2	(236.0)	-477.8	(114.5)	2,517.4	(115.1)	224.2	(25.9)
Montana	12,526.7	(789.5)	44.4	(248.1)	1,032.3	(107.8)	76.3	(20.8)
Nebraska	17,983.5	(357.3)	173.1	(127.8)	1,205.9	(90.7)	55.1	(12.7)
Nevada	121.2	(75.5)	-89.9	(48.6)	381.4	(65.7)	26.7	(9.6)
New Hampshire	20.7	(10.4)	0.0	(4.5)	588.6	(47.6)	62.6	(9.4)
New Jersey	426.7	(48.4)	-60.9	(21.0)	1,778.2	(60.8)	213.6	(19.2)
New Mexico	1,388.1	(125.8)	-72.9	(59.4)	1,152.7	(114.7)	217.2	(74.5)
New York	2,751.6	(146.2)	-123.7	(97.2)	3,183.6	(117.6)	317.6	(27.2)
North Carolina	5,065.5	(206.2)	-483.6	(85.5)	3,856.4	(178.2)	506.6	(60.6)
North Dakota	22,820.7	(378.7)	-36.0	(142.1)	991.8	(72.1)	32.8	(10.6)
Ohio	10,254.6	(193.3)	-598.6	(103.5)	3,611.3	(136.0)	364.8	(34.3)
Oklahoma	9,345.4	(288.7)	-289.9	(107.4)	1,926.3	(110.2)	176.7	(31.4)
Oregon	2,676.8	(178.6)	-104.4	(55.7)	1,222.3	(112.7)	103.9	(22.7)
Pennsylvania	3,629.1	(156.4)	-104.4	(100.4)	3,983.2	(126.8)	545.1	(33.7)
Rhode Island	4.6	(3.1)	-2.6	(2.4)	200.6	(16.3)	6.6	(2.0)
South Carolina	2,319.3	(122.5)	-436.2	(56.4)	2,097.3	(104.3)	362.0	(38.6)
South Dakota	14,340.0	(345.4)	-64.6	(149.7)	959.7	(89.2)	57.8	(20.8)
Tennessee	3,261.7	(132.5)	-380.6	(86.4)	2,370.6	(108.2)	401.9	(37.0)
Texas	26,330.0	(591.1)	-1,225.3	(250.1)	8,567.0	(287.3)	893.5	(78.2)
Utah	705.3	(147.0)	-195.4	(159.9)	661.6	(74.7)	81.3	(24.3)
Vermont	138.7	(25.3)	-6.7	(20.6)	317.5	(29.2)	11.5	(3.9)
Virginia	1,636.6	(112.3)	-169.9	(54.7)	2,625.8	(102.3)	343.5	(29.8)
Washington	5,577.1	(363.0)	-182.8	(119.6)	2,065.0	(174.4)	240.8	(46.6)
West Virginia	166.1	(37.0)	-48.3	(19.2)	873.6	(65.9)	176.8	(21.6)
Wisconsin	8,752.0	(241.3)	-57.7	(153.9)	2,417.9	(127.4)	188.2	(26.7)
Wyoming	978.0	(181.3)	3.0	(82.3)	643.7	(98.0)	34.4	(15.1)
Caribbean	174.2	(22.7)	-108.5	(21.2)	506.8	(29.6)	112.4	(10.2)
Total	**326,783.7**	**(1,838.3)**	-8,456.5	(741.1)	98,251.7	(884.4)	11,217.0	(247.9)

Table 2. Estimated average annual sheet and rill erosion for 1997 Cultivated Cropland, with estimated margins of error in parentheses (in tons/acre/year)

State	Erosion rate	Margin of error
Alabama	6.7	(0.32)
Arizona	0.7	(0.07)
Arkansas	3.5	(0.09)
California	0.7	(0.17)
Colorado	1.7	(0.12)
Connecticut	5.6	(1.53)
Delaware	2.0	(0.19)
Florida	1.8	(0.28)
Georgia	5.9	(0.30)
Hawaii	2.5	(0.47)
Idaho	3.3	(0.31)
Illinois	4.1	(0.10)
Indiana	3.0	(0.10)
Iowa	4.9	(0.13)
Kansas	2.2	(0.04)
Kentucky	4.4	(0.23)
Louisiana	3.3	(0.09)
Maine	3.9	(0.93)
Maryland	4.4	(0.41)
Massachusetts	4.5	(1.61)
Michigan	2.0	(0.09)
Minnesota	2.1	(0.07)
Mississippi	5.3	(0.23)
Missouri	5.6	(0.23)
Montana	1.9	(0.13)
Nebraska	2.9	(0.10)
Nevada	0.2	(0.05)
New Hampshire	3.5	(2.07)
New Jersey	5.6	(0.63)
New Mexico	0.9	(0.08)
New York	3.8	(0.26)
North Carolina	5.0	(0.37)
North Dakota	1.4	(0.04)
Ohio	2.6	(0.10)
Oklahoma	2.8	(0.10)
Oregon	3.0	(0.33)
Pennsylvania	5.1	(0.26)
Rhode Island	3.5	(0.99)
South Carolina	3.2	(0.19)
South Dakota	2.0	(0.07)
Tennessee	7.7	(0.41)
Texas	2.6	(0.05)
Utah	1.6	(0.75)
Vermont	3.1	(0.67)
Virginia	5.9	(0.42)
Washington	4.7	(0.27)
West Virginia	4.3	(1.19)
Wisconsin	3.7	(0.19)
Wyoming	1.1	(0.17)
Caribbean	12.2	(2.71)
National	3.1	(0.03)

2. The National Resources Inventory Program

2.1. THE NATURAL RESOURCES CONSERVATION SERVICE

The Natural Resources Conservation Service, formerly known as the Soil Conservation Service, is an agency of the United States Department of Agriculture. Its mission is “to provide leadership in a partnership effort to help people conserve, maintain, and improve our natural resources and environment.” The Soil Conservation Service was established through the Soil Conservation Act of 1935 after the Dust Bowl catastrophe. This legislation was based, in part, upon results from the 1934 National Erosion Reconnaissance Survey, which was the first formal study of erosion conducted in the Nation. Hugh Hammond Bennett, the agency founder and first administrator, convinced the U.S. Congress that soil erosion was a national menace and that a permanent agency within USDA was needed to call landowners' attention to their land stewardship opportunities and responsibilities. With the establishment of the Soil Conservation Service, a nationwide partnership of Federal agencies, local conservation districts, and communities was developed to provide assistance to the rural and urban sectors in the conservation of natural resources. Today, more than 60 years later, NRCS champions the vitality of the land as USDA's lead conservation agency. No other Federal agency speaks for the health of America's private land.

2.2. HISTORICAL INVENTORIES

Throughout its history, NRCS has conducted periodic inventories of the Nation's natural resources. After the Erosion Reconnaissance Survey of 1934, the agency performed the 1945 Conservation Needs Inventory, which also was based on reconnaissance methods and the gathering of information from a variety of sources.

In 1956, NRCS's predecessor, the Soil Conservation Service, was assigned responsibility to lead a cooperative endeavor of 8 Federal agricultural agencies to develop estimates of the magnitude and urgency of the various conservation measures needed to maintain and improve the country’s productive capacity. This study used statistical sampling methodology and was called the 1958 National Inventory of Soil and Water Conservation Needs, or the 1958 Conservation Needs Inventory (CNI). The National Inventory of Soil and Water Conservation Needs, 1967, or the 1967 CNI, was conducted primarily to update the 1958 data. To facilitate data collection and processing while reducing field costs, a two-stage sample design was implemented. These inventories dealt mostly with land use, soil type, and conservation treatment needs. The Potential Cropland Study of 1975 was conducted to determine the quantity, locations, and characteristics of lands that could be easily converted from noncropland to cropland uses. This study was motivated by increased global demands for agricultural products. A national-level sample was easily obtained by selecting a subset of the 1967 CNI sample points.

2.3. INITIATION OF THE NATIONAL RESOURCES INVENTORY

The Rural Development Act of 1972 played a critical role in the initiation of the National Resources Inventory. The Act specified that information on urban spread and rural economic survival, prime farmland, flood plains, and conservation practices was needed to devise community guidance for balanced rural-urban growth. In addition, the Act directed NRCS to assess the status, condition, and trends of soil, water, and related resources on the Nation's nonfederal lands at intervals of five years or less. The Soil and Water Resources Conservation Act of 1977 and other supporting legislation augmented the statutory mandate for periodic assessment of the Nation's natural resources. These demands for up-to-date information and for additional items related to soil and water resources led to development of the 1977 National Resources Inventory. New data elements included water and wind erosion, conservation practices, incidence of wetlands, and flooding propensity. For the first time, there was a scientific collection of national quantitative erosion data based upon the Universal Soil Loss Equation and the Wind Erosion Equation. The 1977 NRI introduced a number of new methodologies and is considered the starting point for the current NRI program.

2.4. THE NATIONAL RESOURCES INVENTORY PROGRAM

To adhere to the mandate set forth in the Rural Development Act of 1972, NRCS began conducting the NRI on a five-year cycle. After the 1977 NRI, the sample size and scope were significantly expanded for the 1982 NRI, in response to increased demand for data emphasizing environmental and ecological concerns. Additional data were collected on soils, irrigation practices, wetlands, windbreaks, riparian vegetation, wildlife habitat diversity, critical eroding areas, and many facets of vegetative cover. Many far-reaching decisions were based upon the 1982 NRI data because of its comprehensiveness. However, because land use definitions and erosion estimation protocols were significantly different between the 1977 NRI and the 1982 NRI, the 1982 NRI has the crucial role of the initial time-point for the monitoring conducted in more recent inventories.

The scope and size of the 1987 NRI were reduced because of staffing and funding constraints. Trained data collection teams gathered approximately 30 percent of the 1987 sample data using remote sensing and entered the data via computer. Data collection protocols, quality assurance, estimation techniques, and database construction were all designed to ensure that the 1987 NRI provided meaningful and legitimate trending information.

The 1992 and 1997 NRIs are comparable to the 1982 NRI in size and extent. Some data elements that were collected for the 1982 and/or 1987 NRIs have been dropped, while a few new items were added. However, data elements, definitions, and protocols were kept as similar as possible because the main objective has been to develop a time series database for the years 1982, 1987, 1992, and 1997. Data collection for both the 1992 and 1997 NRIs relied heavily on remote sensing and computer-based technologies. In 1997, data gatherers recorded information with personal digital assistants (PDAs) and computer-assisted survey instruments (CASIs).

Concurrently, there have been several small-scale topical inventories conducted from 1995 to 1999. These special inventories were designed to update trends in erosion on highly erodible cropland and investigate effects of the Federal Agriculture Improvement and Reform Act of 1996 and high commodity prices on conservation practices and cropping patterns.

2.5. NRCS DATA COLLECTION ACTIVITIES

In addition to these recurrent NRI inventories, NRCS also collects large quantities of field-level natural resources data in support of conservation planning activities and the Soil Survey Program. Thousands of NRCS technical specialists, including soil scientists, soil conservationists, range conservationists, foresters, wildlife biologists, and agronomists, collect data at the field and farm level in order to provide conservation assistance to farmers and ranchers in the development of conservation systems uniquely tailored to the land and their individual way of doing business. Assistance is also provided to rural and urban communities to help reduce erosion, conserve and protect water resources, and solve other resource-related problems. The information that NRCS collects about natural resources in the United States is critical for sustaining agriculture, promoting the conservation and stewardship ethic, and preserving the health of the Nation's natural resources and environment.

2.6. MEETING FUTURE NEEDS FOR INFORMATION

The NRI program is continuing to evolve as cost-effective methods are developed to collect more timely and relevant information that addresses emerging agri-environmental issues. These new inventory approaches are built upon past experiences as well as by incorporating new tools, methodologies, and technologies. NRCS has developed a new continuous inventory process, wherein data are collected annually. This process better addresses priority issues, provides many results on an annual basis, and more fully utilizes capabilities of highly trained staff and sophisticated tools and technologies. Efforts are underway to develop a multi-agency integrated inventory approach from the many Federal activities that currently attempt to monitor the environment. Other initiatives include the incorporation of various tools for assessing resource health and the utilization of inventory data for modeling and policy analysis.

3. Inventory Procedures

3.1. INVENTORY DESIGN

Current NRI statistical procedures are a result of decades of collaborative research between NRCS and the Iowa State University Statistical Laboratory. The basic sampling procedures were established in 1977 and 1982, based upon research conducted in the 1970s. Estimation techniques have evolved with each successive survey.

The national sample is a stratified two-stage unequal-probability area sample that can be modified for specific national survey objectives and used as a frame for special studies. Stratification was developed county-by-county, utilizing the grid of sections and townships defined by the Public Land Survey System (PLSS), where possible. A section is a one-mile square parcel of land, and a township is a 6-mile square area consisting of 36 sections. Each township was subdivided into three 2-mile by 6-mile strata for sampling purposes. For counties not covered by the PLSS, strata were developed by: utilizing latitude and longitude, utilizing the Universal Transverse Mercator grid system, or superimposing lines analogous to townships and sections over a county highway map.

Two-stage area samples were selected within each stratum. The first stage sample unit, or Primary Sampling Unit (PSU), was an area of land; at the second stage of sampling, one or more sample points were selected within each sample PSU for observation. Most PSUs corresponded to quarter-sections and were half-mile squares; three sample points were selected within most PSUs. Sampling rates varied across strata, typically being between 2 percent and 6 percent; these PSUs constitute about a 3 percent sample of all the total land and water area of the 48 conterminous States, Hawaii, and the Caribbean area. There are instances throughout the U.S. where components of the sample design deviate from these standard rules. The inventory design for Alaska and the Pacific Basin is significantly different and analysis for these two areas is handled separately from the 48 conterminous States, Hawaii, and the Caribbean. NRI data for Alaska and Pacific Basin are not included in the CD-ROM release.

Further details are provided in Appendix C, Understanding the NRI Design. Nusser and Goebel (1997) and Goebel (1998) provide more details on the specifics of the sampling design, and on historical perspectives.

The national sample consisted of approximately 300,000 PSUs and 800,000 sample points for the 1992 and 1997 National Resources Inventories; almost all of these were also part of the 1982 NRI. Experience has shown that it is necessary to return to specific sample points that were in previous inventories in order to obtain needed data on the dynamics of change in land use and various natural resource parameters.

3.2. DATA GATHERING

Data for the 1997 NRI were collected for PSUs and points using photo-interpretation (PI) and other remote sensing methods and standards. For the most part, analog PI techniques were employed but GIS technologies were also tested and evaluated at selected sites. The agency contracted for the acquisition of aerial photography or obtained necessary imagery in cooperation with other USDA agencies and partners. Data gathering for the 1997 NRI occurred from July 1997 through October 1998. This schedule took into account that some aerial photography needed to be flown during a time period that highlighted late growing season conditions.

Inventory procedures were developed to ensure that data reflected 1997 growing season conditions, that inventory results are nationally consistent, and that data recorded for the 1997 NRI are consistent with 1982, 1987, and 1992 determinations. The 1997 NRI protocols specified that data collected from previous inventory cycles (1982, 1987, and 1992) be reviewed, evaluated, and edited concurrently with the 1997 NRI data collection effort.

Detailed soils information is essential for the analysis and interpretation of NRI data. Soils data were provided by the NRCS Soil Survey Program and were obtained from the NRCS Soil Interpretation Record database maintained at the Iowa State University Statistical Laboratory. For the 1992 NRI extensive work was done to match individual State Soil Survey Databases (SSSD) with each point in the NRI. This process was designed to verify the accuracy and completeness of the NRI soils database. This work also provided accurate soils data for use in the 1997 NRI. Published soil surveys, advanced (pre-publication) soil mapping field sheets, State-level databases, and ancillary lists of soils information maintained in field offices were used to provide critical soils data. Information on soil properties related to soil erosion and other soil-dependent interpretations (i.e., prime farmland) were linked to the NRI database.

County Base data is another critical component of the design and development of the NRI database. This dataset requires periodic editing and updating in order to establish NRI statistical controls and to generate certain types of estimates. See Section 3.4 for a discussion of the design, construction, and purpose of County Base data.

USDA field office records and local NRCS personnel provided information pertaining to historical cropping and management systems for calculating long-term erosion rates induced by wind or water, and to determine if the field at the sample point was enrolled in the Conservation Reserve Program (CRP). Data gathering protocols incorporated NRCS technical standards and procedures, records and maps in local USDA offices, and various Federal publications and standards, primarily the following:

- NRCS national and field office technical guide (FOTG) publications and standards relating to the Universal Soil Loss Equation and Wind Erosion Equation (WEQ),
- NRCS-published or -correlated soil surveys,

- NRCS information relating to provisions of the 1985 Food Security Act and subsequent farm bills,
- U.S. Fish and Wildlife Service wetland maps,
- U.S. Fish and Wildlife Service Cowardin wetlands classification system,
- Society of American Foresters forest classification,
- U.S. Geological Survey (USGS) hydrologic and topographic maps, and
- Bureau of the Census TIGER files and auxiliary information.

Field visits were conducted when available materials were not suitable for making determinations for one or more data elements. Field visits were also made for training purposes and other facets of the quality assurance process. All NRI sample sites were visited on-site for the 1982 NRI. Subsequent on-site visits of selected PSUs also occurred in 1987, 1991, 1992, or 1995.

Twenty-one Inventory Collection and Coordination Sites (ICCS) were established and assigned oversight and management authorities for data gathering. These sites were:

Ames, IA	Amherst, MA	Anchorage, AK
Auburn, AL	Bismarck, ND	Boise, ID
Bozeman, MT	Davis, CA	East Lansing, MI
Lakewood, CO	Lexington, KY	Little Rock, AR
Madison, WI	Morgantown, WV	Phoenix, AZ
Portland, OR	Raleigh, NC	Reno, NV
Salina, KS	Spokane, WA	Temple, TX

The ICCSs were the front-line management structures responsible for coordinating the day-to-day activities associated with the collection of data for the 1997 NRI. The ICCS leader trained subordinate staffs, provided technical support, and managed quality reviews during the operational phase of data gathering. Full-time, part-time, and temporary NRCS employees – and in a few places, volunteers – gathered 1997 NRI information. The organization of data gathering varied with regional land use and State staffing patterns. Geographic boundaries of ICCS organizations ranged from one State to all or portions of several States. Some collection sites assembled staffs at multiple office locations, while others assembled staff at one central location. The State Conservationists assigned various technical specialists to provide overall support and to work with the ICCS data gathering teams.

3.3. QUALITY ASSURANCE

Intricate quality assurance procedures were developed to make sure that year-to-year differences reflect actual changes in resource conditions, rather than differences in the perspectives of two different data collection specialists or changes in technologies and protocols.

Data gatherers entered all sample data directly into hand-held computers called personal digital assistants. All subsequent data quality checking and evaluations were similarly based on computer forms of survey information. The PDAs (Apple Computer's Newton

MessagePad™ models 130, 2000, and 2100) were programmed to provide an intelligent survey questionnaire with historic information, procedural logic, and single or multi-variate checking for data completion and consistency. The PDAs uploaded and downloaded sample records via Internet protocols from a centralized database server at Iowa State University. The server controlled and monitored access by "client" instruments and protected survey data from loss, unauthorized access, or accidental disclosure. A secure Web site allowed database access for purposes of survey management, review of progress, and data quality evaluation.

Data gatherers received the protocols as well as training and technical reference materials in printed and electronic forms. Cross-indexed data gathering instructions, training, and the survey instrument were specifically designed to foster consistent data gathering standards, practices, and procedures. Instructions and training articulated the requirement that trending information be gathered by the same protocols and for the same locations used in prior studies. The 1997 instrument included algorithms to evaluate trending data for consistency.

Quality assurance procedures included the use of data validation software packages on the personal digital assistants. Hundreds of data collection rules comparing multiple data elements were run on the PDAs prior to submittal of data to the ISU Statistical Laboratory. The Statistical Laboratory performed additional data validation and consistency checks on all data received. After statistical estimation procedures were completed, tables were generated and sent to technical specialists at the ICCSs for further review and comment.

Quality assurance of NRI data was monitored at all organizational levels within the NRI program and was accomplished by several procedures and protocols. These included consistent training of data collectors in data collection processes; standardized formalized written data gathering instructions, documentation, and definitions of data elements; consistent national rules and methods for data collection; and a national Help Desk to resolve data collection issues.

National staffs conducted multiple training sessions to support core work groups. NRCS national program staff in Washington (DC), Fort Collins (CO), Ames (IA), and Fort Worth (TX) guided and supported the data gathering activities. State or regional oversight authorities supported local data gathering staffs in matters relating to quality assurance. A single point of contact ("Help Desk") on the national staff responded to questions from data gatherers and coordinated technical responses provided by subject-matter experts from NRCS and the Iowa State University Statistical Laboratory.

Quality assurance for the NRI process was guided by instructions, coordinated training, and ongoing technical support. Data quality checking procedures were embedded in every step, stage, and phase of the data collection process for the 1997 NRI.

3.4. COUNTY BASE DATA

County Base data are geostatistical data used as surface area control totals in the NRI statistical weighting process. County Base data for the 1997 NRI were developed in a different manner than for previous NRI surveys. New statistical estimation procedures were developed for the 1997 NRI to take advantage of the increased detail available from this new geostatistical control information.

County Base data for the 1997 NRI were obtained from a Geographic Information System developed specifically to provide these control data. Geostatistical information was extracted from this GIS for each “HUCCO” in the NRI sampling universe. A HUCCO is a polygon formed by the intersection of 4-digit Hydrologic Unit Areas and county boundaries. There are about 4,900 HUCCOs in the conterminous U.S., Puerto Rico, and the Virgin Islands. However, 213 of these were too small to be used in the NRI estimation process and were combined (collapsed) with an adjacent HUCCO within the same county. (Note: The data element *mhydro* in the table **point** specifies this collapsed HUCCO code – it is provided in the database for documentation only and would not be used when creating tabulations from the data. The data element *hydro* in the table **point** specifies the 8-digit hydrologic unit code that is used when tabulating the data.)

County boundaries were obtained from U.S. Bureau of the Census TIGER digital line files released in 1995, corresponding to a 1:100,000 map scale. The TIGER files were augmented with shoreline data from the National Oceanic and Atmospheric Administration (NOAA); several modifications were needed to match NRI land universe definitions. Spatial data for the 204 4-digit HUAs within the conterminous U.S. were based upon a 1:250,000 U.S. Geological Survey digital line file; boundaries were adjusted to conform to the shoreline created for the county boundary data.

The GIS also included data layers identifying Large Water and Federal Land areas. The TIGER data were used as a starting point in delineating water areas; modifications were made using 1992 NRI County Base data, USGS quadrangle sheets, and other materials. Federal lands were initially based upon a 1996 USGS 1:2,000,000 scale digital layer (released in 1998) and the TIGER data files. Modifications were made to the Large Water and Federal Land data based upon several editing and checking stages, including NRCS State office reviews. Federal Land data for several western States were based mostly on databases available to the State offices. (Note: The NRI definition of Large Water is that used through 1980 by the Census Bureau, and consists of water bodies ≥ 40 acres and streams ≥ 660 feet wide at normal pool.)

Specific data extracted from this GIS were:

(1) Total surface area of each HUCCO
(2) Area of Large Water within each HUCCO, for 1992 and 1997
(3) Federal Land area within each HUCCO, for 1992 and 1997 (Large Water is not part of the Federal Land area)

The County Base data used for the 1997 NRI estimation process also used geostatistical control data for the years 1982 and 1987 (for Large Water and Federal Land area within HUCCO). These figures were derived from the 1992 NRI County Base data, using statistical adjustments based upon differences between the 1992 control acres used in 1992 and the 1992 control acres obtained from the GIS. The 1992 NRI County Base data (and also those for the

1982 and 1987 NRIs) had been supplied by NRCS State and field offices, were not based upon GIS procedures, and were obtained for Major Land Resource Area portions of counties rather than HUA portions.

Use of County Base data within the 1997 NRI estimation process is briefly discussed in Chapter 4; detailed procedures are given by Fuller, Dodd, Wang, and Peterson (2001). Development of these geostatistical estimation data is discussed by Nusser, Kienzler, and Fuller (1999); they state that the general objective of these new procedures was to "provide a more reliable and efficient source of control information, and to improve the geospatial properties of imputed data." Updating and maintaining the accuracy of County Base data is a critical component of NRI statistical controls and estimation processes.

3.5. CONFIDENTIALITY AND INTEGRITY OF SAMPLE LOCATIONS

It is USDA and NRCS policy that actual locations of NRI sample sites are confidential information and are not to be released to the public. As per policy: *"Proper security and confidentiality of information and materials pertaining to locations of data gathering sites must be maintained. Site locations (identified by coordinate systems, depicted by maps or photographs, or described by direct observation of survey location conditions, and other materials assembled for inventories) are not public information and shall be used only for official inventory purposes. This critical requirement must be met in order to ensure the statistical integrity and validity of the NRI survey design, dataset, and estimates, to prevent intended or unintended sampling bias, and to maintain the confidentiality and cooperation of landowners and operators."*

There is much at stake regarding the continued confidentiality and integrity of the NRI sample sites. Many staff years and dollars have been invested to develop this sophisticated and extensive set of samples. A great deal is now known about these hundreds of thousands of samples. Maintaining the confidentiality and integrity of these sample locations is vital to the continued success of the NRI program.

4. Construction of the NRI Database

4.1. INTRODUCTION

The NRI database contains millions of pieces of information. It is a powerful and unique tool, if it is understood and used properly. This chapter presents information on how the NRI database is constructed – it provides an overview of various features of the database and of selected statistical estimation procedures used during creation of the database. Further information describing the database is contained in Chapter 5.

4.2. STATISTICAL ESTIMATION PROCEDURE

The NRI statistical estimation procedure is fairly complex, because of the complexity of the data and because a number of requirements have been established regarding construction of the database. The goal of the NRI statistical estimation procedure is to create a final NRI database that meets the following requirements:

- *Contains all information collected for the survey* **–** The dataset must contain all segment (PSU) and point sample data (for 1982, 1987, 1992, and 1997). Also, the estimation procedure must incorporate certain county and sub-county ancillary data as controls; these inputs come from a geostatistical database and are called County Base data. (See Section 3.4.)

- *Produces estimates for previous years that are similar to those previously published* **–** The estimation procedure incorporates thousands of estimates published in 1992 (for the years 1982, 1987, and 1992) as controls. However, there are several complications. For example, all historical raw data were reviewed and edited as part of the 1997 data collection protocol. Also, county and sub-county ancillary data used for controls in 1997 were developed using improved (geostatistical) technologies. (See Section 3.4.)

- *Is easy to use* **–** Users of the data should be able to produce tabulations from the database without needing to understand the estimation procedure. All data are transformed into analysis variables for the public database. There is a unique weight (expansion factor) assigned to each observation.

- *Is complete* (for 1982/1987/1992/1997) **–** Imputation procedures were used to make sure the database is “fully populated” and has no missing values. Many other databases, whether large or small, do have missing values, which makes the data more difficult to analyze.

- *Produces estimates that agree with known data* – For example, the sum of the weights in County *S* must equal the total area of County *S.*

- *Contains linkages to other databases* **–** Linkage to NRCS's extensive soils database is done directly; many linkages can be made geographically (spatially); and other themes, such as land cover/use or forest cover type, serve as ways to link with (or tap into) other databases.

- *Produces reasonable estimates for small areas* **–** The NRI was designed to support analysis of issues at National, regional, State, and sub-State levels – but **not** at the county level. However, estimates for counties and 8-digit HUAs can be derived from the database; such estimates need to be "reasonable," even though they do not meet NRI reliability standards because of the small sample sizes for geographic units of that size. This characteristic of the database makes users of the database more comfortable, and serves to provide more statistically robust results when the data are aggregated to higher levels for display, presentation, and discussion purposes.

- *Allows users of the data to estimate standard errors* – The database contains factors identifying the sampling strata and clusters. Variance estimates are based upon a number of approximations because of the multiple sources of data and relatively complex estimation procedure.

- *Accounts for rare occurrences* **–** This is important for large-scale surveys; users of the data expect that the survey will provide such estimates (at least at the national or regional level). For agri-environmental surveys**,** the extreme or rare situations are often of most interest.

The different types of data collected for the NRI are combined using a series of statistical estimation procedures. The result is a set of data tables (or files) that are easy to use and manipulate and represent consolidated point-type information. This database structure makes it possible for different users of the data to obtain the same results when asking the same questions or making the same queries of the data.

Another feature is that many different pieces of information or many different data themes can be looked at concurrently. In other words, many pieces of site-specific information describing conditions at each of the specific sampling sites are available. These hundreds of pieces of information are site-specific data — not generalized map and tabular information. This greatly enhances the user's ability to analyze and correlate the data properly. Overlaying many layers of generalized data can cause problems when trying to make correlations and construct models or scenarios.

The point-style data also make it possible to associate non-NRI data with the NRI database. This can be done much more rigorously, even though non-NRI data are usually much more generalized; examples would be county figures on production, chemical use, or precipitation. Analysts will often treat these data as they do specific NRI data items — they will associate or impute these factors onto the specific data points. The structure of the NRI dataset makes this process more direct and feasible to implement.

An important requisite of the entire inventory process is that data for the four points in time (1982, 1987, 1992, and 1997) must be completely comparable and consistent from year to year so that the data reflect true trends and temporal analyses can be made. When making comparisons with 1982, 1987, and/or 1992 data, the new (1997) NRI database must be used. Originally released 1982, 1987, and 1992 data may not be consistent with the new database — the historical 1982, 1987, and 1992 data were updated to reflect 1997 technical criteria and protocols as part of the 1997 NRI inventory process. The 1997 NRI database should be considered self-contained in that respect.

The NRI statistical estimation procedures are discussed by Fuller (1999); extensive details are given by Fuller, Dodd, Wang, and Peterson (2001). These papers discuss the use of several statistical techniques, including: small area estimation, imputation, pseudo point generation, hot deck procedures, ratio adjustment, raking, and rounding procedures. All statistical techniques have been carefully developed (over many decades of cooperative research and application) and are accepted by the scientific community.

4.3. IMPUTATION

Data collection protocols, quality assurance processes, estimation techniques, and database construction were all designed to ensure that the 1997 NRI provides meaningful and legitimate trending information (Goebel 1998). Imputation procedures were used during construction of the NRI database to:

(1) statistically combine information contained in the County Base data, PSU data, and point data to produce a complete point dataset with one weight for each point; and
(2) statistically fill in any missing data for sample units not observed in 1987 and for those new to the sample in 1987 and/or 1992.

Imputation is a statistical estimation technique. Imputation processes use information that was actually collected to predict the values of missing data elements within the dataset. Because the NRI database is used in many different ways by many different individuals and groups, it is important that the database be complete and "fully populated" – there are no missing values, and every data field for every database record contains a legitimate value. Statistical imputation techniques were used in developing the 1992 and 1997 NRI databases as a means of enhancing the quality of the database.

Data for the 1997 NRI were collected for the same sample sites that were used for the 1992 NRI. However, only one-third of these was sampled for the 1987 NRI, because of staffing and funding constraints (Also, some 1992 NRI samples were not part of the 1982 NRI.). For the two-thirds of the samples that were not inventoried in 1987, data were obtained during the 1992 NRI using the following process:

(1) If suitable imagery or adequate ancillary information, or both, were available for 1987, data gatherers recorded key selected 1987 data elements—but not every data element.
(2) If no suitable imagery or adequate ancillary information could be found for 1987, data gatherers recorded no data for the 1987 inventory year; this occurred infrequently.

The key data elements that were determined for those 1992 NRI samples that were not part of the 1987 NRI were: ownership category, prime farmland, land cover/use, and irrigation for 1987 for each sample point; and all PSU data items except for those describing Large Water.

Missing 1987 data were imputed during construction of the 1992 NRI database. The 1997 NRI database also contains these imputed values. The following is an example of how the 1992 NRI imputation process handled missing 1987 data elements dealing with erosion estimation.

Consider first the USLE C-factor, which for Cropland is based upon the cropping system, particularly the crop rotation, tillage, and residue management. This is the factor in the USLE that accounts for ground cover present during the critical eroding period.

The imputation procedure for 1987 USLE C-factor can be thought of as a stochastic search procedure. This procedure searched for a sample point that was:

(1) inventoried in 1987
(2) geographically close
(3) similar with regard to 1987 land cover/use history, land capability class and subclass, irrigation status, and wind erosion

The search procedure for this "donor point" used a hierarchical scale in weighing closeness and similarity, and in placing iterative limitations upon the number of times given points could serve as donors. The search for a donor first focused on points in the same MLRA and county, then on points within the same county but in a different MLRA, and then on points located in adjacent counties. Steps were developed for extreme situations where no sufficiently similar points existed.

In most cases, the search procedure also determined the imputed 1987 values for the following data elements:

- cropping history
- conservation treatment needs
- P-factor, for USLE
- K-factors and V-factors, for the Wind Erosion Equation
- length of rotation, for the Wind Erosion Equation
- types of conservation treatment needed
- whether double cropping is present
- conservation practices

– range condition
– apparent trend in range condition

Many regression correlations were examined to determine which variables should be included and in what order of importance.

> *WARNING: This type of estimation procedure does not produce individual records that are perfect. It is like any prediction or modeling technique. Any single predicted value is subject to an estimation error. However, the imputation process significantly improves the overall quality of the database.*
>
> *Analyses focusing on certain types of changes involving imputed data will NOT be legitimate; most of the imputed 1987 data elements dealt with erosion estimation, conservation practices, and conservation treatment needed. For example, it is not legitimate to estimate number of acres where erosion increased by at least two tons/acre/year from 1987 to 1992 by looking at specific point data; however it is legitimate to examine the overall change in erosion for some set of sample points.*
>
> *The imputation process provides defensible scientific estimates for overall or aggregated changes — but not where individual records are required to be exact.*

It is necessary to use imputed data values when constructing estimates from NRI data. For example, imputed data must be used to properly construct an estimate of 1987 erosion rates. If imputed values are omitted, estimates will be biased and inconclusive. The 1997 NRI database contains a data element that indicates which values, if any, were imputed for the sample point.

The second primary use of imputation for NRI statistical estimation is generation of "pseudo points" that are needed to transfer the PSU and County Base data to point-level data. (County Base data are discussed in Section 3.4.) A goal of NRI is creation of a complete point dataset that combines the information contained in the County Base data, PSU data, and point data to produce a complete point dataset with one weight for each point. A dataset of this form is user-friendly and easy to tabulate.

The procedures for generating pseudo-points are required because County Base data and PSU data may contain patterns of change over time that are not observed at an actual NRI sample point. For example, the County Base data may indicate a slight increase in Federal land from 1992 to 1997 for a particular county. However, there may be no points in the county showing a change from nonfederal ownership in 1992 to Federal ownership in 1997. The approach used by the NRI is to create "pseudo points" within the county to represent the observed change in Federal ownership.

This imputation approach is used in the NRI to match point estimates for county total area of Large Water and Federal Land to those values given in the County Base data. Other examples include area estimation and change for relatively small geographic areas of interest such as roads, urban areas, or farmsteads (Fuller 1999).

In Section 4.4, development of weights for point-level data is defined and examined. In Section 4.5, an example of how pseudo points are constructed is presented.

4.4. DEVELOPMENT OF NRI WEIGHTS

One of the data fields in the NRI database is the weight, or expansion factor; it is the data element *xfact* in the table **point**. This factor specifies the number of acres that a sample point represents when statistical estimates are derived using the NRI database. This weight must be used for all tabulations and analyses – whether estimating average erosion rates, acreages, percentage figures, or margins of error; otherwise, results will be biased.

These weights have many desirable features. For example, the weights for all points in a county sum to the total surface area of the county. Also, only one set of weights exists within the NRI database. This one prepared set of weights is used to construct any type of estimate, regardless of the theme. (Some databases have multiple sets of weights or do not even have weights constructed. Data analysts then face confusion and extra work in determining how to properly use the database.)

A discussion of the development and construction of the NRI weights follows. It begins with an introduction to the basic concepts and then describes how other factors affect the calculations. The actual process used for the 1997 NRI is fairly complex; see Fuller, Dodd, Wang, and Peterson (2001) for a detailed description.

Basic concept

Consider first the simplest situation:

- Only have one year's worth of data
- Only have point data (and no PSU data)
- Constant sampling rate across the county (for the first stage sampling, or PSUs)
- Each PSU contains three points

The straightforward way to calculate the weights for this county would be to assign the same value (acres) to each point. In other words, each point would represent the same number of acres when using the data to estimate acres, erosion, and other statistics.

- For example, if there are 48 PSUs, each with three points, and the county is 368,640 acres, then the weight for each of the 144 points is: 368,640 acres ÷ 144 = 2,560 acres per point.

  Note that for this example sampling was one out of every 48 quarter-sections, which is a nominal 2 percent sampling rate. Each PSU then represents a total of 48 quarter-sections; therefore:
  one PSU represents: 48 x 160 acres = 7,680 acres
  and each point represents: 7,680 acres ÷ 3 = 2,560 acres.

This is the basic concept – there are five factors that necessitate modification of this concept:

(1) The sampling rate is not constant for the entire county
(2) There are PSU-level data that need to be incorporated into the point data
(3) County Base data are used as controls
(4) Data for four years (1982, 1987, 1992, and 1997) are included in this one database
(5) Several ancillary data sources are incorporated into the weight generation process

It is not surprising that there is not a simple closed-form equation for a weight. Instead, the weights are developed through a multi-step process, including several steps that are iterative.

Modification – varying sampling intensities

In simple terms, the value (or size) of the weight for a point is inversely related to the intensity of the sample. Consider the earlier example. Suppose the sample had been twice as dense as described. In other words, the sampling rate was 4 percent rather than 2 percent over the entire county. Then each weight would be one-half as large, or 1,280 acres per point, rather than 2,560 acres as derived earlier.

Consider next how weights would be determined if exactly half of the county was sampled at the original 2 percent density and the other half at double the intensity, or at a 4 percent rate. The points within the 2 percent sample area would have weights of 2,560 acres each. The points within the 4 percent sample area would have weights of 1,280 acres each.

Note that these weights would sum to the county total –

$$\begin{aligned} 72 \times 2{,}560 \text{ acres} + 144 \times 1{,}280 \text{ acres} &= 184{,}320 \text{ acres} + 184{,}320 \text{ acres} \\ &= 368{,}640 \text{ acres} \end{aligned}$$

Adjustments for PSUs larger or smaller than 160 acres follow logically from the above discussion.

- For example, consider the basic concept for the nominal 2 percent sample with 640-acre PSUs and three points per PSU. Each point's weight would be four times what it is for 160-acre PSUs, because each PSU (and hence point) represents four times as much area:

$$\text{weight} = 4 \times 2{,}560 \text{ acres} = 10{,}240 \text{ acres}$$

- For 40-acre PSUs containing two points each and for a straight 2 percent rate, the basic concept gives a weight for each point equal to:

$$48 \times 40 \text{ acres} \div 2 = 960 \text{ acres per point}$$

Similar adjustments can be made for PSUs in the Northeastern States. Those PSUs are 30 seconds of longitude by 20 seconds of latitude, varying from 97 acres in Maine to 113 acres in Virginia.

Modification – use of PSU-level data

The discussion of the basic concept of a weight omitted several important factors that are taken into account during the development of the actual NRI weights. The next consideration is how PSU-level data affect weights, still omitting the impacts of different years of data, Geostatistical Control Data, and other ancillary sources.

Area measurements for seven Land Cover/uses are collected at the PSU level:

(1) Farmsteads
(2) Small Built-up Areas
(3) Large Urban and Built-up Areas
(4) Small Water Bodies (<40 acres)
(5) Small Streams (<660 feet wide)
(6) Rural Transportation Land
(7) Census Water

- Consider the following example for a PSU that is 160 acres in size –

PSU Area Data

- 5 acres of Farmsteads
- 55 acres of Large Urban & Built-up
- 0 acres of the other four categories

Point Data

- 1 point falls on Large Urban & Built-up
- 2 points fall on Cropland

For this PSU, this means:

(1) 60 acres are in Land Cover/Uses collected at the PSU level
(2) 100 acres are in Land Cover/Uses not covered by PSU-level data
(3) 2 points (out of three) fall on Land Cover/Uses not covered by PSU-level data

The logical way to construct weights for the two points falling on Cropland is to equally allocate the 100 acres. The weight for each of these points, assuming the nominal 2 percent sampling rate, is:

$$(100 \text{ acres} \div 2) \times 48 = 2{,}400 \text{ acres}$$

(Note that the 48 can be thought of as the PSU expansion factor. Recall that a nominal 2 percent sample means that 1/48 of the county is covered by the selected PSUs — and weights are inversely proportional to sampling rates.)

The Urban and Built-up point would be given an initial weight of: (55 acres) x 48 = 2,640 acres.

The acres of Farmsteads are accounted for by creating a pseudo point, as discussed in Section 4.5. Note that Census Water areas are also part of the NRI County Base data and are handled differently than Small Water areas during construction of weights.

4.5. CONSTRUCTION OF PSEUDO POINTS

The concept of a pseudo point was introduced in the previous example. The PSU-level data indicated there were 5 acres of Farmsteads on the PSU, but none of the three real sample points fell on a Farmstead. To account for these Farmstead acres within the NRI database, a pseudo point is constructed.

To more fully consider construction of pseudo points, it is necessary to take into account that all levels of data (County Base, PSU-level, and point data) are collected for four years (1982, 1987, 1992, 1997) – and that the NRI database incorporates all three levels of data for all four years.

- For example, consider a PSU where the following data were recorded —

PSU Data (in acres)

	1982	1987	1992	1997
Small Built-up	2.8	5.0	5.0	7.4
Large Urban	12	12	26	26

Point Data – Land Cover/Use

	1982	1987	1992	1997
Point # 1	Corn	Corn	Corn	Corn
Point # 2	Pasture	Pasture	Pasture	Pasture
Point # 3	Corn	Soybeans	Corn	Soybeans

Note the following —

(1) The PSU contains both Small Built-up and Large Urban and Built-up areas
(2) Both categories of Built-up are increased during the 15-year period
(3) None of the three points fell on either type of Built-up area

Pseudo points are created to reflect these occurrences within the 1997 NRI database. Five pseudo points need to be created for this PSU – two to reflect the acres of Built-Up Land that existed in 1982, and three to reflect the changes. If the PSU is 160 acres, there are 126.6 acres that are not in a PSU-level category as of 1997 – this means the three points are each given one-third of these acres, or 42.2 acres, as an initial weight (before applying the PSU expansion factor). This PSU requires a total of eight "points" (three real points and five pseudo points), as follows:

Augmented Point Data – Land Cover/Use & Acres

	1982	1987	1992	1997	Acres
Point # 1 - *real*	Corn	Corn	Corn	Corn	42.2
Point # 2 - *real*	Pasture	Pasture	Pasture	Pasture	42.2
Point # 3 - *real*	Corn	Soybeans	Corn	Soybeans	42.2
Point # 4 - *pseudo*	Small B-U	Small B-U	Small B-U	Small B-U	2.8
Point # 5 - *pseudo*	Pasture	Small B-U	Small B-U	Small B-U	2.2
Point # 6 - *pseudo*	Corn	Soybeans	Corn	Small B-U	2.4
Point # 7 - *pseudo*	L Urban	L Urban	L Urban	L Urban	12.0
Point #8 - *pseudo*	Corn	Soybeans	L Urban	L Urban	14.0

The 1982 Land Cover/Use for point 5, the 1982-1992 Land Cover/Uses for point 6, and the 1982 and 1987 Land Cover/Uses for point 8 were selected through a randomized search process that is part of the statistical estimation procedure.

Pseudo points were also created based upon the NRI County Base data. Recall that these base acreages were developed for each HUCCO portion of a county.

Consider the following situation:

(1) Federal Land area increased by 200 acres from 1992 to 1997 (for a HUCCO portion of a county), and
(2) None of the sample points within this HUCCO changed ownership from nonfederal in 1992 to Federal in 1997

A pseudo point with weight of 200 acres would be created to reflect this change. This pseudo point would be similar to those discussed earlier that reflect changes in Urban and Built-up Land for a PSU. Statistical procedures would be used to impute missing data elements for 1982, 1987, and 1992.

4.6. ADDITIONAL FEATURES

The NRI database weights were constructed by taking several additional factors into consideration Most important are controls – coming from both the NRI County Base data and several ancillary sources. An example of a control is the stipulation that the weights for all of the points within a county must sum to the pre-established total acres for the county. In other words, a control is placed upon the weights to ensure that they do sum to the desired total.

The statistical estimation procedure developed weights that met these controls, to the extent possible. To produce weights that simultaneously meet all desired controls is not possible. However, by using an iterative process, it is possible to meet some controls exactly and come close to meeting most others.

This iterative procedure incorporated a series of regressions and ratio adjustments. This statistical methodology is the result of over 20 years of research and development conducted collaboratively by the Iowa State University Statistical Laboratory and USDA.

The weights were developed so that the sample points meet as closely as possible the following controls:

(1) Total surface area figures for States, counties, and HUCCO portions of counties
(2) Census Water area figures for States, counties, and HUCCO portions of counties — for 1982, 1987, 1992, and 1997
(3) Federal Land area figures for States, counties, and HUCCO portions of counties — for 1982, 1987,1992, and 1997
(4) Nonfederal Land area figures for States, counties, and HUCCO portions of counties — for 1982, 1987, 1992, and 1997
(5) Acres contracted to the Conservation Reserve Program as of the 1992 and 1997 growing seasons, by county or groupings of counties, and by groupings of sign-up periods
(6) Broad cover/use figures developed for the 1982 and 1992 NRI, by State

Several other features of the estimation procedure should be mentioned –

(1) A statistical test was performed to determine if estimates for 1992 broad cover/uses were significantly different based upon 1997 NRI data (compared to the estimates originally published for the 1992 NRI)
(2) A small area estimation procedure was used to provide better estimates of Developed Land cover/uses
(3) Weights in the NRI database are in terms of 100s of acres – rounding (to 100s) was done as part of the weight generation process so that users of the data will obtain identical results, as long as they handle the data appropriately and do no further rounding

5. NRI Database Structure and Characteristics

5.1. Introduction

The NRI database is a statistical database. The data do not come in an aggregated or summarized form. The data must be aggregated or summarized in order to be utilized properly and effectively. This chapter describes the organizational structure of the NRI point data on the CD-ROM and various characteristics of the database. The user is provided references to additional documentation and tools on the CD to assist in using the data.

5.2. Database Structure on CD-ROM

All NRI data are combined into a dataset containing point-type information. The data are placed in a normalized or relational database structure so that the data can be efficiently managed and accessed. The relational database is partitioned into nine logical subject matter tables as described below. The variables (fields) within each table are comma delimited.

Every record in each table includes a unique key variable named *recordid* that is used to match or link data across tables. Every record in all tables except the table **point** includes a *yr* variable to indicate the particular inventory year. See an example of linking records across tables in Section 5.5.

See the "Point Record Layout and Database Design" for specific variables and associated codes, **CD-ROM filename: Documentation/point_layout.pdf**. A crosswalk from the layout of the 1992 NRI CD point records is provided for users of the previous NRI CD **at CD-ROM filename: Documentation/crosswalk92_97.pdf**.

1. Table: **point**

    Data that are not year-specific. This table serves as the "master table" relative to the relational structure. This master table includes two required variables for any data summary: the sample point weights or expansion factors (*xfact) and* a geographical factor (*fips, hydro, or mlra*). Almost all data tabulations and analyses will require data from this master table and one or more of the other tables, often the table **trend.**

2. Table: **trend**

    Trending data that are year-specific and are available for at least two inventory years (usually for all four inventory years: 1982, 1987, 1992, 1997). Refer to the Point Record Layout and Database Design "Field Name" column to see what years are available for a particular variable. [**CD-ROM filename: Documentation/point_layout.pdf**] (Includes Average Annual Wind Erosion Equation as opposed to yearly equation factors in table **weq**.)

3. Table: **1997only**

   Data that are year-specific, but are only available for 1997.

4. Table: **crophist**

   Data on cropping history for the three calendar years prior to the inventory years (only for those points that have Broad Cover/Uses of Cropland, Pastureland, or CRP in a given inventory year).

5. Table: **weq**

   Wind erosion equation factors for each inventory year plus the three prior calendar years (only for those points that have Broad Cover/Uses of Cropland, Pastureland, or CRP in a given inventory year). The Average Annual Wind Erosion Equation (aaweq) is included in table **trend**. "aaweq" is the estimated soil loss due to wind erosion averaged for the rotation during the four-year period.

6. Table: **habitatindices**

   Wildlife habitat diversity indices and metrics computed from the raw habitat composition and configuration data collected for transects associated with the point, for 1997 only. (See "Applications of Habitat Composition and Configuration Data**,"** **CD-ROM filename: Documentation/habexamp.pdf**.)

7. Table: **habitatraw**

   Raw habitat composition and configuration data collected for transects associated with the point, for 1997 only. There are at least four records for each point. A unique record consists of: *recordid*, *yr*, *hctransno*, and *hctrsegno*.

8. Table: **overlandraw**

   Raw overland flow data, for 1997 only (only for those points that have Broad Cover/Uses of Cropland, Pastureland, or CRP for 1997). There usually are multiple records for each point. A unique record consists of: *recordid*, *yr,* and *ofsegno*.

9. Table: **wetland**

   Wetland and aquatic habitat data for 1997 only.

5.3. Database Characteristics

For the 1997 NRI database on CD-ROM, the table **point** serves as the "master table" relative to the relational structure. The table **point** contains 1,314,726 point-type records; this includes 814,962 real points. The other records represent pseudo points, except for

42,690 records added to the dataset to improve estimation of margins of error for several categories of estimates. These are records identical to 42,690 real points, except their *recordid* is modified (they are identified by the split code, element *split*). It is interesting to note that over half of the real points fell on either Forest Land or Cropland in 1997— 224,615 on Forest Land and 216,335 on Cropland.

The number of data records per State varies considerably. Most States have between 20,000 and 35,000 records. The extremes are Delaware and Rhode Island with 1,998 and 2,015 records, respectively, and Texas and California with 88,215 and 46,057 records, respectively. [See **CD-ROM filename: Documentation/record_counts.pdf**, Number of Records for 1997 NRI Dataset on CD.]

5.4. DATABASE ELEMENTS

In general terms, the various data elements (variables/fields) will contain one of the following types of information —

(1) Data about the location of the point
(2) Data about natural conditions at the point
(3) Data about use and management of the site (describing human activities)
(4) Several factors needed to produce statistical estimates from these data
(5) A key, or pointer, to the soil interpretations database
(6) Variables needed to provide record uniqueness within the relational database structure

All database fields are identified in 1997 NRI – Point Record Layout and Database Design [**CD-ROM filename: Documentation/point_layout.pdf**]. The NRI database contains several data elements that have been derived, in addition to those items recorded directly by the NRI data gatherers. The following data fields were derived rather than directly observed:

- Weight
- Stratum
- Cluster
- Split Code
- Imputation Code
- Imputation 87 Code
- Broad Land Cover/Use [1982/1987/1992/1997]
  (derived from *Specific* Land Cover/Use)
- Pointer to Soils Database
- USLE 'K' Factor
- USLE Flag
- USLE Soil Loss [1982/1987/1992/1997]
- WEQ 'I' Factor
- WEQ Soil Loss, Individual Years [1979, 1980, 1981, 1982; and 1984, 1985, 1986, 1987; and 1989, 1990, 1991, 1992; and 1994, 1995, 1996, 1997]
- Average Annual Soil Loss, WEQ [1982/1987/1992/1997]

- Erodibility Index, EI, Water [1982/1987/1992/1997]
- Erodibility Index, EI, Wind [1982/1987/1992/1997]
- Erodibility Index, EI [1982/1987/1992/1997]
- Habitat Cover Percent [1997]
- Habitat Cover at the Point [1997]
- Habitat Cover Patch [1997]
- Interspersion Index [1997]
- Shannon Diversity Index [1997]
- Shannon Index of Equitability [1997]
- Simple Dominance [1997]
- Dominant Cover Type [1997]
- Cover Type Dominance [1997]
- Mean Length of Segment 1 [1997]
- Segments Per Unit [1997]
- Aquatic Habitat [1997]
- Cowardin System [1997]

A number of database elements exist that affect the proper use of the data. These include the following:

Federal Land — The NRI database does not contain information for Federal lands. It does contain records for sample points identified as Federal, but these contain no attributes for years that they are Federal. These records must be omitted when calculating averages and other statistics. They do need consideration when looking at land use changes and other issues involving multiple years of data, because points do change between Federal and nonfederal. Federal ownership is determined through examination of the Broad Land Cover/Use (see below).

Land Cover/Use — NRI uses this term to identify a classification system that accounts for all the surface area of the U.S., using mutually exclusive categories. The database includes variables for both a Specific Cover/Use (*landuse)* and Broad Cover/Use (*broad*) in the table **trend**. There are 12 Broad Cover/Use categories:

(1) Cultivated Cropland
(2) Noncultivated Cropland
(3) Pastureland
(4) Rangeland
(5) Forest Land
(6) Other Rural Land
(7) Urban and Built-up Land
(8) Rural Transportation
(9) Small Water Areas
(10) Census Water
(11) Federal Land
(12) Conservation Reserve Program (CRP) Land

There are 64 Specific Land Cover/Use categories. The NRI data collectors determined the Specific Cover/Use category; the Broad Cover/Use category was derived from the Specific Cover/Use category. The system is hierarchical except for some types of Hayland and Pastureland; if the hay or pasture is in rotation with cultivated crops, the point is classified as "Cultivated Cropland" rather than "Noncultivated Cropland" or "Pastureland."

Note that "Federal Land" is one of the Broad Cover/Use categories. Points classified as "Federal Land" are included in the database; they have the statistical weight *xfact* and various geographical codes but almost no other attributes. They are included in the database for completeness – to make a complete accounting of all U.S. surface area.

The category Other Rural Land contains farmsteads, other land in farms, marshland, barren lands, and other rural lands not covered by the first five categories.

Developed Land — Developed Land contains three Land Cover/Use categories:

(1) Large Urban & Built-up Areas (greater than 10 acres in size)
(2) Small Built-up Areas (less than 10 acres in size)
(3) Rural Transportation Land (roads, railroads, and associated rights-of-way)

Important special considerations regarding Developed Land are discussed in Section 1.3.

Soils Data — The soil identified for sample points is determined by standard soil mapping conventions. This means that sample points that fall on minor soil components are included as part of the primary mapping unit or mapped soil. The NRI database contains certain soil data elements, interpretations, and a key or pointer (*nriptr* in table **point**) that can be used to link the point to additional soil interpretations information. Note that these additional soil interpretations are <u>not</u> provided on the CD-ROM. (Interested users may request these data by contacting the NRCS Resources Inventory Division at the address given in Appendix E.) Efforts are currently underway to link NRI data to a new soil database management system, the National Soils Information System (NASIS), which is under development by NRCS.

Erosion — NRI erosion data are discussed in Section 1.3. Note that the database does not contain erosion data for Broad Cover/Uses other than Cultivated Cropland, Noncultivated Cropland, Pastureland, and Conservation Reserve Program lands.

Erodibility Index — The Erodibility Index (EI) expresses the inherent erodibility at a site. It is only calculated for years when a point is Cropland. The EI data field can be used to identify highly erodible Cropland (EI equal to or greater than 8). EI is calculated from erosion equation factors determined for a transect crossing the sample point.

Irrigation — For the NRI a site is considered irrigated if irrigation occurs during the inventory year or for two or more of the last four years. Consider this when making comparisons with other data sources. For example, previous Census of Agriculture irrigation figures were lower than NRI numbers because actual application of water is required for the year of the census. Historically, problems have existed with undercounting irrigated acres in the Census of Agriculture.

Wetlands — The Cowardin classification system covers both wetlands and aquatic deepwater habitats. Wetland and deepwater habitat determinations are included in the 1997 NRI database for all nonfederal points for inventory year 1997 only. FSA wetland classifications are also included. Use the "Cowardin System" data element to replicate estimates found in Table 16 of the 1997 NRI Summary Report. The "Cowardin Specific" data element provides a more specific Cowardin system category.

Practices — Conservation Practices are identified if they were applied in the area in which the point falls, or the portion of the field surrounding the point that is considered in conservation planning. The point does not necessarily fall directly on the practice. There are 22 practices.

Habitat — Measurements of habitat composition and configuration are new in the 1997 NRI. The makeup of the general cover categories and their arrangement about a point are useful in analyzing diversity of habitat in defined geographical areas. Several diversity indices and metrics can be calculated from these data based upon the needs of the user. See: "Applications of Habitat Composition and Configuration Data**" [CD-ROM filename: Documentation/habexamp.pdf**].

Overland Flow — Overland flow and delivery to water data were collected only for points classified as a Land Cover/Use category of Cropland, Pastureland, and CRP land for 1997. Information related to distance from the point to nearest receiving water, type of receiving water, cover categories that surface water flows through, and length of each cover within the water flow route were collected for these points.

Stratum/Cluster — These data fields are needed to properly estimate margins of error or any other measure of the statistical uncertainty of an NRI estimate. They were constructed using the NRI sampling scheme, which is discussed in Chapter 4, Construction of the NRI Database. SAS programs provide examples of methods used to properly estimate margins of error **[CD-ROM filenames: Tools/SAS/table1.sas . . . table14.sas]**.

Imputation Code — This code identifies which points are real points and which are "pseudo points." For pseudo points, this code identifies the type of imputation. As discussed in Sections 4.3 and 4.5, these pseudo points are required because County Base data and PSU data may contain patterns of change over time that are not observed at an actual NRI sample point.

Imputation Code for 1987 Data — This code identifies situations where it was necessary to impute 1987 data, either partially or completely; situations requiring imputation, and methodology, are discussed in Section 4.3.

5.5. Example of Linking Data Across Tables

- This example illustrates how to link data across subject matter tables using the *recordid* as the unique key. In order to get an estimate of acres of nonfederal Cropland in Alabama for all four inventory years (Table 2, *1997 NRI Summary Report*, column 1), one must query data from two tables.

Table **point**:		Table **trend**:
recordid	←--------------------------→	*recordid*
fips		*yr*
xfact		*broad*

The database or statistical software query must be written to sum the expansion factors/weights (*xfact*) for all the points that have a *fips* that begins with 01 (for Alabama) and a *broad* cover/use code of either 1 or 2 (for Cultivated and Noncultivated Cropland), where the *recordid* in table **point** equals or matches the *recordid* in table **trend**. In order to summarize the data for each of the inventory years, the data would be grouped by the *yr*.

For SAS® users, there are examples of SAS programs that provide estimates for a sample state for several of the *1997 NRI Summary Report* tables, including standard errors. [**CD-ROM filename: Tools/ SAS/table1.sas . . . table14.sas**]

5.6. Importing Raw Data

Raw data must be imported into database and/or statistical software * in order to perform the aggregation mentioned in the preceding sections. Users may choose from three different options for accessing the raw data. Further details are provided in the readme file at **CD-ROM filename: readme.txt**.

1. Import data for certain table(s) for certain States *

   Users may choose to import only those table(s) (file(s)) required for particular analysis needs. One or multiple States of data for the chosen table(s) may be imported.

   The CD contains a zip archive of data for each State. Each State archive (st.zip) consists of the following comma-delimited files for the subject matter tables. ("st" is replaced by the two-letter State abbreviation.)

   1. st_point.txt
   2. st_trend.txt
   3. st_1997only.txt
   4. st_crophist.txt
   5. st_weq.txt
   6. st_habitatindices.txt

7. st_habitatraw.txt
8. st_overlandraw.txt
9. st_wetland.txt

Data header files are also provided.

(A sample MS® ACCESS database software unpopulated database framework is provided on the CD.
Sample SAS® programs for importing the raw data are provided on the CD.)*

2. Import data for certain table(s) for entire country *

   A user may decide to import only those table(s) or file(s) required for particular analysis needs for all the States.

   The CD contains a zip archive of the entire Nation (50 States), with filenames as listed above in option 1, but with "national" as the prefix. Data header files are also provided.

   (Sample MS® ACCESS and SAS® tools as described above in option 1 are provided.)*

3. Extract populated MS® Access database *

   Users who own MS® Access and expect to use all the tables of data for most or all of the States may choose to extract the populated Access database provided on the CD. The database is a zipped archive that contains four files:
   1997NRI_1.mdb
   1997NRI_2.mdb
   1997NRI_3.mdb
   National.mdb

   National.mdb links to the three 1997*.mdb files and provides relationships between tables. This implementation solves the one-gigabyte size limitation of Access at the expense of performance.

* The MS® Access and SAS® tools are provided as a convenience to users who already own this software and may be modified to suit individual needs. Use of product names (or trade names) does not imply endorsement of products (or software systems) by the USDA Natural Resources Conservation Service.

5.7. RECOMMENDATIONS TO USERS AFTER DATA IMPORT

To make effective use of the NRI database, one must understand NRI concepts, methods, definitions, and survey design. It is helpful to have an understanding of the material in the previous chapters of this user guide. Additionally, it is helpful to continue to use and explore the data, to compare analyses with results from other databases, and to discuss and present new results to other data analysts.

After loading the raw data into a database or statistical software application, users should compare the number of records for each table(s) and each State(s) against the spreadsheet provided in **CD-ROM filename: Documentation/record_counts.pdf**.

Users are advised and encouraged to replicate various estimates given in Tables 1-17 and Appendix Tables 1 and 2 of the *1997 NRI Summary Report* as a data check [**CD-ROM filename: Documentation/summary_rpt.pdf**]. This replication will verify whether (1) the data loaded properly, (2) the data were aggregated correctly, and (3) the margins of error were generated properly.

5.8. DERIVING MARGINS OF ERROR

The precision of all statistical results should be derived and examined. The 1997 NRI database has been constructed in a manner that provides methods for deriving estimated margins of error (or standard errors, or confidence intervals) for any estimate derived from the database.

- The data elements *stratum* and *cluster* in the table **point** must be used, in order to account for the stratification and two-stage sampling used for the NRI, as described in Section 3.1 and Appendix C, Understanding the NRI Design.

- Several examples of computer programs that appropriately calculate estimated standard errors using the statistical software package SAS® are included on the CD **[CD-ROM filename: Tools/SAS/table1.sas . . . table14.sas**].

- Such algorithms and procedures provide suitable estimates and should be used for all NRI estimates. However, there are some items for which these procedures actually provide an overestimate of the standard error, because of some of the specialized statistical (estimation) techniques used to construct the 1997 NRI database (see Chapter 4 for a discussion of these techniques).

- Several commercial statistical packages provide the functionality of estimating standard errors based upon the NRI stratified cluster sampling design: **
  - SAS® (Release 8.2) (http://www.sas.com/)
  - SUDAAN ® (http://www.rti.org/units/shsp/sud1.cfm)
  - WesVar (http://www.westat.com/wesvar/)
  - PC CARP and SUPERCARP – DOS software available from: Statistical Laboratory, Iowa State University, 219 Snedecor Hall, Ames, Iowa 50011; telephone (515) 294-9773.

** Most common statistical packages will not provide appropriate estimates for the NRI stratified cluster sampling design. However, the packages listed above provide this functionality.

5.9. Citing Data from the CD-ROM and the 1997 NRI Summary Report

<u>CD-ROM</u>:

U.S. Department of Agriculture. 2001. 1997 National Resources Inventory (revised December 2000), Natural Resources Conservation Service, Washington, DC, and Statistical Laboratory, Iowa State University, Ames, Iowa, CD-ROM, Version 1.

<u>*1997 NRI Summary Report*</u>:

U.S. Department of Agriculture. 2000. *Summary Report: 1997 National Resources Inventory (revised December 2000*), Natural Resources Conservation Service, Washington, DC, and Statistical Laboratory, Iowa State University, Ames, Iowa, 89 pages.

.

6. NRI Data Analysis

6.1. INTRODUCTION

This chapter provides an introduction to the use of the NRI database from an analyst's perspective. Typical analytical processes and routines followed to produce NRI estimates are outlined and examples are presented. The information in this chapter was originally developed to accompany classroom instruction to internal NRCS staff on the 1992 NRI. The chapter is written in a tutorial format and some of the information has been emphasized in other parts of this user guide. The training module is included in its entirety in this chapter.

6.2. THE ANALYTICAL PROCESS

Data analysis using the NRI—or any other database—must be done carefully to avoid misusing the data and misrepresenting results to decision-makers. To help avoid these potential problems, the following basic steps should be taken in conducting the analyses.

1. Identify the problem or question.
   All NRI applications are in response to a question or problem that requires analysis. The first step in analysis is to identify the question as specifically as possible.

2. Determine if NRI data can be used to address the question.
   The NRI database can be used to address many questions related to agriculture and natural resources. However, the NRI database will not always be the best information set to use. For example, application of analyses to small areas—such as a single county—will not be appropriate in most cases.

   For some applications satellite imagery data or county-level Census of Agriculture data may be better suited to the specific question being asked. In other cases the definitions used in the collection of the NRI data may not be specific enough to properly address some questions. Evaluate the feasibility of using NRI data to address a specific question early in the analytical process.

3. Determine what other kinds of data are needed.
   For many applications, the NRI database will need to be augmented with information from other sources. GIS technology provides one means for integrating geographic data from other sources. There are also tabular databases available that contain geographic codes or descriptions, such as Census of Agriculture.

4. Involve the relevant discipline specialists.
   Involve soil scientists, agronomists, GIS specialists, or other discipline specialists as needed to aid in preparing and evaluating the analysis.

5. Define the analytical product precisely.
   After determining that the NRI can be used to address the question and the appropriate discipline specialists have been consulted, define the precise analytical product. This can range from a simple calculation, such as percent area, to complex analysis involving models.

6. Caveats—know your data.
   Caveats—assumptions—always underlie the analysis. Keep a list of them. For example, percent area of Cropland depends specifically upon how Cropland was defined. Prevent the audience from misunderstanding how the results were derived and from misunderstanding what they actually mean.

7. Prepare tabular and graphical representations of results.
   Use both tables and maps whenever possible. Specifically plan to produce these products, as they are the most effective formats for communicating findings.

8. Document methods and results in writing.
   The importance of documentation cannot be overstated.

9. Present results and documentation to discipline specialists for feedback.

10. Adjust analysis and presentation as needed.

11. Present results in written and oral form to decision-makers.

12. Follow up regarding further questions and concerns.
    Often decision-makers ask new questions after receiving answers to old questions. Be prepared to follow up with additional analysis.

6.3. Overview of NRI Applications

The NRI was designed to provide support for natural resources and environmental policy development and program implementation. The questions that the NRI can be used to answer typically fall into one of the following four categories:

1. Describe the status and trends of natural resources and agricultural activity.

   - This analytical function is, for the most part, a direct representation of the data elements in the NRI.
   - Examples are the geographic extent of a particular land use, the spatial variability in land use, or changes in land use over time.

2. Evaluate the condition of natural resources, using environmental and ecological indicators.
   - This function involves more sophisticated analysis than the first.
   - An example is generating projected environmental impact indices utilizing NRI data and other potential data sources.

3. Assess environmental and economic implications of changes in resource use, including expected impacts associated with shifts in government policies.
   - This function is very complex. It will almost always involve data from other sources, will typically be conducted using mathematical models, and will involve experts from several disciplines.

4. Plan and manage the Nation's conservation programs.
   - This analytical function actually incorporates elements from each of the previous three functions, but is focused sharply on a narrow objective -- planning and managing programs.

6.4. Know Your Database

The key to successfully applying the NRI to answer questions is thorough knowledge of the database.

Recognize what is in the NRI, and what is not.

- Is rangeland condition included in the database? (No.)
- Are there erodibility index values for all land? (No—just Cropland, Pastureland, and CRP.)

Definitions are important.

- What is Cropland in the NRI? Does it include Hayland? Idle land? Set-aside acres? (See the Glossary in the *1997 NRI Summary Report*.)
- What is Developed Land? Are farmsteads and barns included as Developed Land? Are airports? Or abandoned railroads? (See the Glossary.)
- What is prime farmland? How can acres of prime farmland increase? Can prime farmland in Cropland change over time to nonprime farmland in Cropland? (Yes—over one million acres did just that between 1982 and 1992.)

Understand how the data were collected.

- You should know which attributes were obtained from handbooks and which were obtained from a site visit (and thus are actual data). (See "Instructions for Collecting 1997 National Resources Inventory Data" at **CD-ROM filenames: Documentation/collection_instructions/content.pdf . . . through . . . ptmod11.pdf** or at **http://www.nhq.nrcs.usda.gov/NRI/1997/data_gathering.html**.)
- Understand how PSU data were converted to point data, and which data were imputed.

For example, pseudo points were used to distribute PSU data to a point database. Can you ignore or exclude pseudo points in your analysis? (No. The expansion factors are adjusted so that they sum to control totals only when pseudo points are included in the sum.)

- How do you handle imputed points for 1987? Can they be excluded from your analysis? (No.) Why not? (Same answer as above.)
- Lastly, an important aspect of the NRI is that it is a statistical database. That means that there is some uncertainty about all estimates. The extent of the uncertainty is estimated by the variance, or the margin of error. Margins of error should be calculated for all estimates. To make this point again, look at the distribution of NRI PSUs in a typical Midwest situation. Estimates are made based on this sample. The greater the sample size, the smaller the margin of error.

These topics have all been covered in previous chapters.

6.5. Understanding the Statistical Nature of the NRI

Important Note: Statistical Reliability

NRI data were collected at more than 800,000 sample sites nationwide. The data, which are gathered using recognized statistical sampling methods, are statistically reliable for national, regional, State, and some multi-county sub-State analyses. The standard error, margin of error, or coefficient of variation should be taken into consideration in all data analyses. (See Sections 1.4 and 1.5.)

Information on statistical reliability and limitations is required for:

(1) Reporting results to decision-makers
(2) Comparing results with estimates from other sources
(3) Identifying real differences
  – Area to area
  – Time period to time period
(4) Presentation of results
  – Use of error bars
  – Parsimonious use of classes in maps and tables
(5) Analysis—Aggregation (summing of NRI expansion factor acreages by some geographic code, e.g., Hydrologic Unit Areas, MLRAs, States) is almost always required

Three kinds of calculations are necessary to properly use NRI data:
(1) Total acres
(2) Weighted totals
(3) Weighted averages

Total acres

- Sum of Weight over the collection points of interest
- Examples:
  – Acres of corn in Iowa
  – Acres of Cropland in Nation

- SPECIFIC EXAMPLES:

The Weight or expansion factor is used to represent the sample point. It is ALWAYS used to aggregate.
(The Weight has been defined and discussed previously.)

Example A1-- Total Acres of Cropland in County A

Sample Point	1997 Land Cover/Use	Weight
1	cropland	3,000
2	cropland	2,500
3	cropland	4,000
4	cropland	3,500
5	cropland	3,300
6	cropland	6,000
7	cropland	4,600
8	cropland	2,500
9	cropland	4,000
10	cropland	3,500
11	cropland	3,300
12	cropland	6,000
13	cropland	4,600
	Sum =	50,800 acres

➢ A percentage as a ratio of two Total Acres estimates can be calculated.

Example A2-- Portion of Total Acres of Cropland that are Prime Farmland in County A

Sample Point	1997 Land Cover/Use	Prime Farmland	Weight
1	cropland	yes	3,000
2	cropland	no	2,500
3	cropland	yes	4,000
4	cropland	yes	3,500
5	cropland	yes	3,300
6	cropland	no	6,000
7	cropland	no	4,600
8	cropland	yes	2,500
9	cropland	no	4,000
10	cropland	no	3,500
11	cropland	yes	3,300
12	cropland	no	6,000
13	cropland	no	4,600

Sum of prime = 19,600 acres

$$\text{Prime Farmland Fraction} \quad \frac{19{,}600}{50{,}800} = 0.386$$

Weighted totals

- Per acre attributes multiplied times Weight (NRI expansion factor acres) and then summed
- Example: tons of soil loss from sheet and rill erosion

  - SPECIFIC EXAMPLE:

    It is not used frequently, but it is important. It is appropriate when the attribute is defined in per acre units. Here, it is used to calculate soil loss.

Example B -- Average Rate of Soil loss from Sheet and Rill Erosion on Cropland in County A

Sample Point	1997 Land Cover/Use	USLE (t/a/y)	Weight (acres)	Weighed Total (t/y)
1	cropland	4.0	3,000	12,000
2	cropland	5.0	2,500	12,500
3	cropland	12.1	4,000	48,400
4	cropland	8.5	3,500	29,750
5	cropland	6.8	3,300	22,440
6	cropland	10.1	6,000	60,600
7	cropland	6.1	4,600	28,060
8	cropland	3.0	2,500	7.500
9	cropland	4.0	4,000	16,000
10	cropland	9.2	3,500	32,200
11	cropland	11.3	3,300	37,290
12	cropland	8.4	6,000	50,400
13	cropland	3.9	4,600	17,940
			Sum =	375,080 Tons/year

Weighted average

- Multiply weight (NRI expansion factor acreages) times the attribute, sum over the sample points of interest, and then divide by the sum of Weight for all sample points in the domain (a geographic area unit such as a watershed).
- Examples:
  - Average erosion rate, tons per acre per year, for a watershed in Texas
  - Average K-factor for the State of Iowa

  ➢ SPECIFIC EXAMPLE:

  Calculation of a weighted average is used frequently.

Example C --Average Erosion Rate for Acres of Cropland that are Prime Farmland in County A

$$\text{Average USLE} = \frac{\text{Weighted Total}}{\text{Sum of Weight}}$$

$$\text{Average USLE} = \frac{375{,}080}{50{,}800} = 7.38 \text{ tons/acre/year}$$

6.6. INFORMATION PRESENTATION

Once the question has been asked, and the calculation completed, the challenge is to correctly present what has been found and in a manner that accurately communicates the findings to the decision-maker.

Some analysts believe this phase of the process is more important than all the others.

Various analytical presentation forms are possible. The presentation format should be considered even before the analysis is run.

Often, the presentation medium will dictate aspects of the analysis. It certainly dictates the kind of numbers that need to be obtained from the NRI database.

Presentation format is also important because it is the final step in providing answers to the questions that prompted the analysis. When the presentation has been planned, the specific objectives of the analysis become apparent, such as what numbers need to be calculated, and whether they are total acres, weighted totals, or weighted averages.

Three main presentation formats can be used:

(1) Tables
- Summary tables for reference
  - Tables containing large amounts of information usually are not helpful in directly answering questions
- Tables that make a point
  - A simple table directly answers a specific question

(2) Charts
- Pie charts
- Bar charts
- Text
  - A simple text format avoids confusion

(3) Maps
- Spatial patterns
- Maps cannot stand alone
- Easy to mislead -- consider all aspects carefully
- Colors are important. How color is used can greatly influence how the map reader will interpret the map. Indeed, by altering colors the map creator can cause the map reader to reach different conclusions.

Summary:

(1) Whenever possible, use the format that conveys the results visually
(2) Keep it as simple as possible and still answer the question
(3) Always proof your work

6.7. WORK IN TEAMS

Work in teams -- do not try to do everything individually.

- The NRI is a spatial database. GIS tools and expertise are required to do some of the analyses and to display the results. A GIS expert should be part of your team.

- Involve discipline specialists (such as soil scientist, cartographer, and statistician) to provide scientific credibility to your study.

- Combine efforts across State boundaries whenever possible.

6.8. NRI ANALYSIS SUBJECT AREAS

Display of percentage data aggregated by different geographic units

Maps can be used to communicate NRI data differently depending in part on how the data are geographically aggregated. The NRI database includes several codes that enable the user to aggregate data by different types of geographic units. These include States, counties, Major Land Resource Areas (MLRA), Land Resource Regions (LRR), Bailey's Eco-Regions, and 2-, 4-, 6-, and 8-digit hydrologic unit areas (HUA).

Boundaries such as hydrologic units and MLRAs are often chosen for portraying NRI data, since they better reflect geographic trends in the natural environment over broad areas than do political entities such as counties and States. In many cases, their polygon area sizes encompass a sufficient number of NRI points to generate acceptable statistical errors for many common purposes (but be careful about over-interpretation), and their numbers are such that they communicate more information than maps with a fewer number of polygons, such as States.

The 8-digit hydrologic units are usually smaller than MLRAs. They usually provide more geographic information but error estimates are generally higher for individual HUAs than for MLRAs.

Data can be aggregated at a 4-digit HUA level, an area that would typically equate to several 8-digit HUAs. The contrast between maps produced using 8-digit HUAs and maps produced using the larger 4-digit HUAs or MLRAs can be striking. When comparing the maps side-by-side, one may notice that the data properties of a single 8-digit HUA, if it is highly skewed relative to surrounding HUAs, can skew the visual impression for the entire 4-digit HUA in which it is contained. In any case, the choice of data aggregation unit can significantly impact the impression left on the map reader.

Although smaller areas (8-digit HUAs, MLRAs, or counties) usually have higher error estimates than larger area analyses, their message is sometimes less misleading when compared to larger area aggregations of data.

Multi-county aggregation in the form of Resource Conservation and Development areas or administrative areas may be used. Various groupings of this sort are useful if high error estimates are a concern because larger area aggregations tend to lower the error estimates.

Nevertheless, one should keep in mind that, although the 4-digit HUA estimates tend to lower the error estimates, they are a concern because larger data aggregations do not always yield results that more closely approximate the truth in the geospatial context. Large area data aggregations, such as States or 2-digit HUAs, while more correctly generalizing the data (as indicated by lower statistical error estimates), often fail to reflect legitimate variations within each unit.

When showing NRI data on polygon maps, it is better to represent the data as a *percentage of total acres* within each polygon that meet a specific set of criteria rather than as the *actual acres* that meet the criteria within each polygon. The disadvantage of mapping actual acres within a polygon is that acreages have a direct relationship with the sizes of the polygons. This will obscure whatever point you are trying to make with your map theme. Actual acreage totals are better represented through the use of dot maps, where each dot is randomly distributed within a geographic unit (such as a hydrologic unit) and each dot represents a specific number of acres.

When mapping NRI data by polygon, it is also reasonable to use smaller universes than total acres in the denominator of a percentage calculation, such as the percent of total nonfederal acres, percent of total Cropland acres, etc. However, the smaller the universe, the less accurately it will visually represent the area of each unit as it appears on a map. Likewise, some units may contain few NRI points for a specific universe, and the calculation of a percentage in these cases can create a very misleading impression on a map. For example, few Cropland points exist in some regions of the West where Rangeland dominates. Aside from the fact that the statistical error estimate may be unacceptable in these regions when Cropland is the denominator, it is probably misleading to represent that Cropland has much significance in these areas when the map theme emphasizes some more specific Cropland criteria.

Effective use of mapping polygon acreages

Although polygon maps are better used for representing percentages of area rather than acreages, representing acres is possible in ways that may be useful. One way is in the analysis of workload. Workload within a county may relate to the actual count of acres of Cropland or a variety of other data elements, or both. For example, the total acres of Cropland in a geographic area may be directly related to the number of conservation plans that need to be developed or the number of hours required for applying pesticides. The sum of the acres from the various data components, whatever they may be, could be used to make a composite workload map.

The legend category breaks can be set up as a linear increase in the width of each class. For example, eight is added to the class width of each subsequent class so that the first class has a width of 8, the second has a width of 16, and so on. The cartographer should delineate class breaks that best highlight the distribution of the data.

Choosing equal-sized class widths is just one of many legitimate methods that may be employed to establish legend class boundaries. Other commonly used methods include setting class boundaries at naturally occurring breaks in the data range or setting boundaries so that approximately equal numbers of occurrences will occur in each legend class. How legend breaks are selected will always introduce some bias on the part of the mapmaker, just as selection of colors and patterns do.

As suggested before, depending on the intended use, it may be more appropriate to aggregate to multi-county levels to reduce the values of the error estimates (i.e., reduce the width of the resulting certainty interval). The trade-off in aggregating to multi-county levels is that although you may have greater confidence in the data estimates, a map becomes a less useful means of representing the data. This occurs especially when the geographic extent of that map is only one State and is not in the context of all 50 States. When data are aggregated to a small number of geographic areas, that is less than 10 or 20, tables may be an equally or more effective means of communication.

For maps with few polygons, the benefit over tables may be marginal. When considering the trade-off between mapping small polygons and larger ones, the analyst needs to consider how much statistical variability is acceptable in the estimates, and what is the cost of being wrong. If you are just looking for general trends in the data and are not as concerned with the reliability of the estimate for any given polygon, smaller polygons are often more useful. If, however, having a high level of confidence in estimates for individual polygons is critical, then larger data aggregations are required. The format chosen for display and analysis will depend upon the question being addressed.

6.9. Mapping Results of Data Analysis

Because of its multi-resource approach and consistency across State borders, the NRI has numerous applications for addressing regional issues.

For example, resource managers in the Chesapeake Bay watershed are confronted with decisions that affect water quality of the Bay. The use of NRI data has increased their ability to provide assistance and programs, and create new policy that addresses critical issues in this area.

- The following examples illustrate a few NRI products that can assist managers to better understand natural resource conditions and trends in the Chesapeake Bay watershed.

  Understanding change in the cultivation of prime farmland by 8-digit hydrologic unit area could be valuable for many reasons. Increases in cultivation of prime farmland may indicate a shift towards using the best soil resource for this purpose. Such changes may also be indicators of long-term impacts of agricultural legislation. Decreases in cultivation of prime farmland may be an alert to investigate these areas to determine the effect on water quality from these land use shifts.

  The distribution of cultivated land that is classified as highly erodible may indicate areas where special treatment, practices, programs, or policies need to be implemented. Such information can be useful in assessing conservation workloads and identifying logical interstate partnerships to address natural resource concerns.

  Mapping the 8-digit HUAs by percentage of Cropland planted to a specific crop, such as corn, could be both interesting and useful. Because specific pesticides are sometimes linked to specific crops, adverse effects on water quality can be predicted. A map locating areas where corn is being grown on highly erodible soils can easily be linked to a preceding map for visual comparison or combined using GIS. Such analysis can be useful for assessing the potential use of specific practices such as conservation tillage.

Some water quality problems may be related to specific land uses, such as Pastureland. Such analysis could be useful in understanding issues concerning nutrient loading of streams, streambank erosion, or similar issues.

The links in the NRI to the Soil Interpretations Records database are useful for regional analysis. NRI users have the unique ability to ask questions that correlate specific land management practices with the soils resources upon which they exist. For example, understanding the distribution on a regional basis of soil hydrologic groups and its impact on ground water recharge can contribute to understanding water quality conditions and planning broad land management strategies.

Citations

Fuller, W.A., K.W. Dodd, J. Wang, and C. Peterson. 2001. Estimation for the 1997 National Resources Inventory. Unpublished manuscript, Statistical Laboratory, Iowa State University, 161pp. http://www.statlab.iastate.edu/survey/nri/est971nr.pdf

Fuller, W.A. 1999. Estimation procedures for the United States National Resources Inventory. Proceedings of the Survey Methods Section, Statistical Society of Canada Annual Meeting, June 1999, pp. 39-44. http://www.statlab.iastate.edu/survey/nri/Fullerpr.html

Goebel, J.J. 1998. The National Resources Inventory and its role in U.S. agriculture. *In:* Agricultural Statistics 2000, International Statistical Institute, Voorburg, The Netherlands, pp. 181-192. http://www.nhq.nrcs.usda.gov/NRI/1997/stat_design.html

Nusser, S.M., J.M. Kienzler, and W.A. Fuller. 1999. Geostatistical estimation data for the 1997 National Resources Inventory. USDA-Natural Resources Conservation Service, Washington, D.C. http://www.statlab.iastate.edu/survey/nri

Nusser, S.M., F.J. Breidt, and W.A. Fuller. 1998. Design and estimation for investigating the dynamics of natural resources. Ecological Applications. 8(2):234-245. (coming soon to: http://www.nhq.nrcs.usda.gov/NRI

Nusser, S.M., and J.J. Goebel. 1997. The National Resources Inventory: a long-term multi-resource monitoring programme. Environmental and Ecological Statistics. 4:181-204. http://www.nhq.nrcs.usda.gov/NRI/1997/stat_design.html

APPENDIX A

CD-ROM Contents:
Data, Documentation, and Tools

This appendix describes the organization of the files on the CD_ROM. The CD filename: readme.txt contains further technical details about the data and some associated tools provided as a courtesy to users who own particular commercial software packages.

The CD top level directory structure is:

- Data
- Documentation
- readme.txt (plain ASCII text file)
- Tools

CD-ROM filename:

Data:

- Nationally aggregated State files (50 States in each of nine subject matter table files)

Data/National.zip

- **National_point.txt**
- **National_trend.txt**
- **National_1997only.txt**
- **National_crophist.txt**
- **National_weq.txt**
- **National_habitatindices.txt**
- **National_habitatraw.txt**
- **National_overlandraw.txt**
- **National_wetland.txt**

- State files (50 files)

Data/st.zip

(where "st" is two-letter State abbreviation)

- **st_point.txt**
- **st_trend.txt**
- **st_1997only.txt**
- **st_crophist.txt**
- **st_weq.txt**
- **st_habitatindices.txt**

st_habitatraw.txt
st_overlandraw.txt
st_wetland.txt

- Header files **Data/header.zip**
  - **hdr_point.txt**
  - **hdr_trend.txt**
  - **hdr_1997only.txt**
  - **hdr_crophist.txt**
  - **hdr_weq.txt**
  - **hdr_habitatindices.txt**
  - **hdr_habitatraw.txt**
  - **hdr_overlandraw.txt**
  - **hdr_wetland.txt**

Documentation:

- *A Guide for Users of 1997 NRI Data Files, CD-ROM Version 1 (December 2001)* [this document] **Documentation/user_guide.pdf**

- *Summary Report, 1997 National Resources Inventory (Revised December 2000)* (includes **Glossary**) **Documentation/summary_rpt.pdf**

- 1997 Data Collection Instructions **Documentation/collection_instructions/**

Contents	**content.pdf**
Introduction	**nrintr.pdf**
General Inventory Guidelines	**nrigen.pdf**
PSU Module I General Information	**psumod1.pdf**
PSU Module II Farmsteads and Developed Areas	**psumod2.pdf**
PSU Module III Land Use	**psumod3.pdf**
Point Module I Ownership	**ptmod1.pdf**
Point Module II Conservation Reserve Program	**ptmod2.pdf**
Point Module III Land Use	**ptmod3.pdf**
Point Module IV Soils Information	**ptmod4.pdf**
Point Module V Habitat Composition and Configuration	**ptmod5.pdf**
Point Module VI Overland Flow/Delivery to Water	**ptmod6.pdf**
Point Module VII Irrigation	**ptmod7.pdf**
Point Module VIII Erosion Data	**ptmod8.pdf**
PT Module 8 Erosion Data (Exhibit 1)	**ptmod8ex.pdf**
Point Module IX Wetlands	**ptmod9.pdf**
Plates 1-10	**pt9ex1-10.pdf**
Plates 11-20	**pt9ex11-20.pdf**
Plates 21-30	**pt9ex21-30.pdf**
Plates 31-40	**pt9ex31-40.pdf**

Plates 41-50	**pt9ex41-50.pdf**
Plates 51-60	**pt9ex51-60.pdf**
Plates 61-70	**pt9ex61-70.pdf**
Plates 71-80	**pt9ex71-80.pdf**
Plates 81-86	**pt9ex81-90.pdf**
Point Module X Conservation Practices	**ptmod10.pdf**
Point Module XI Saline Deposits on Agricultural Land	**ptmod11.pdf**

- Applications of Habitat Composition and Configuration Data
  Documentation/habexamp.pdf

- Point Record Layout and Database Design
  Documentation/point_layout.pdf

- File Sizes
  Spreadsheet of file sizes to estimate computer system space requirements
  Documentation/filesize.pdf

- Number of Records for 1997 NRI Dataset on CD
  Documentation/record_counts.pdf

- Data Codes Files
  For reference and/or for user to load into a software application if desired.

	Documentation/codes/
FIPS	**county_codes.txt**
2-Digit Hydrologic Unit Area	**hydro2_codes.txt**
4-Digit Hydrologic Unit Area	**hydro4_codes.txt**
6-Digit Hydrologic Unit Area	**hydro6_codes.txt**
8-Digit Hydrologic Unit Area	**hydro8_codes.txt**
Major Land Resource Area	**mlra_codes.txt**
State	**state_codes.txt**
Variable**s**	**variable_codes.txt**

- Crosswalk Between 1992 and 1997 NRI Record Layouts
  (for users of 1992 NRI CD-ROM)
  Documentation/crosswalk_92_97.pdf

Tools:

- Sample Import Programs (SAS® *)
  (further details in **CD-ROM filename: readme.txt**)

	Tools/SAS/
Import Table **1997only**	**rd97only.sas**
Import Table **crophist**	**rdchist.sas**

Import Table **habitatindices**	**rdhabind.sas**
Import Table **habitatraw**	**rdhabraw.sas**
Import Table **overlandraw**	**rdolraw.sas**
Import Table **point**	**rdpoint.sas**
Import Table **trend**	**rdtrend.sas.**
Import Table **wetland**	**rdwetland.sas**
Import Table **weq**	**rdweq.sas**

- Sample Analysis Programs (SAS®*)
  (further details in **CD-ROM filename: readme.txt**)
  Tools/SAS/

Surface area	**table1.sas**
Land cover/use	**table2.sas**
Changes in land cover/use	**table5sas**
Average annual sheet and rill erosion	**table10.sas**
Average annual wind erosion	**table11.sas**
Average annual sheet and rill erosion in relation to T value	**table12.sas**
Average annual wind erosion in relation to T value	**table13.sas**
Erodibility index for cropland	**table14.sas**

- User Notes for MS® Access * Implementation
  (Use of **1997NRI.mdb)** **Tools/ readme1997NRI.pdf**

- User Notes for MS® Access * Implementation
  (Use of **NationalNRI.mdb)** **Tools/ readmeNationalNRI.pdf**

- Sample Unpopulated Database Framework (MS® Access *)
  (further details in **CD-ROM filename: readme.txt**)
  Tools/1997NRI.mdb

- Populated National Database (MS® Access *)
  (further details in **CD-ROM filename: readme.txt**)
  Tools/NationalNRI.zip
  National_Master.mdb
  1997NRI_1.mdb
  1997NRI_2.mdb
  1997NRI_3.mdb

*There are some limited MS® Access database and SAS® tools provided as a courtesy to users who already own this software, but the software itself is not provided.

Use of product names (or trade names) does not imply endorsement of products (or software systems) by the USDA Natural Resources Conservation Service.)

APPENDIX B
Important NRI Reference Materials on NRI Home Page, World Wide Web

http://www.nhq.nrcs.usda.gov/NRI *

There are several technical documents on the statistical design of the NRI and statistical estimation procedures used in the NRI that may be helpful in optimizing the utilization of NRI data. Users of NRI data are encouraged to review these documents and become familiar with the statistical procedures applied in the NRI. Because of copyright restrictions, these documents have not been included on this CD-ROM, but can be found on the NRI Web site or by using the URLs listed below. Other pertinent documents to which the user may wish to refer are the *1997 NRI Summary Report* and *Instructions for Collecting 1997 National Resources Inventory Data*. Both of these items are part of the documentation on this CD and also can be found on the NRI Web site shown above.

We will periodically generate additional documentation, further analyses of NRI data, and/or announcements related to the 1997 NRI. Users are encouraged to monitor the NRI Web site for updated information.

Statistical Design of the NRI

Goebel, J. J. (1998) "The National Resources Inventory and Its Role in U.S. Agriculture," *In:* Agricultural Statistics 2000, International Statistical Institute, Voorburg, The Netherlands, 181-192.

http://www.statlab.iastate.edu/survey/nri/goebel.pdf

Nusser, S. M., and J. J. Goebel (1998) "The National Resources Inventory: A Long-Term Multi-Resource Monitoring Programme," Environmental and Ecological Statistics, 4(3):181-204.

http://www.statlab.iastate.edu/survey/nri/nussergoebel.pdf

Statistical Estimation for the NRI

Fuller, W. A. (1999) "Estimation procedures for the United States National Resources Inventory," 1999 Proceedings of the Survey Methods Section, Statistical Society of Canada (in press).

http://www.statlab.iastate.edu/survey/nri/fuller.pdf

Nusser, S. M., J. M. Kienzler, and W. A. Fuller (1999) "Geostatistical Estimation Data for the 1997 National Resources Inventory." Working paper.

http://www.statlab.iastate.edu/survey/nri/Summary%20of%20geostat%20info.PDF

Fuller, W. A., K. W. Dodd, J. Wang, and C. Peterson (2001) "Estimation for the 1997 National Resources Inventory."

http://www.statlab.iastate.edu/survey/nri/est971nr.pdf

Nusser, S. M., F. J. Breidt, and W. A. Fuller (1998) "Design and Estimation for Investigating the Dynamics of Natural Resources," Ecological Applications, 8(2):234-245.

(Web address not yet available)

* Refer to Appendix E if help is needed regarding Web addresses.

APPENDIX C
Understanding the NRI Design

C.1. INTRODUCTION

The information in this appendix was originally developed to accompany classroom instruction to internal NRCS staff on the 1992 NRI. The appendix is written in a tutorial format and some of the information has been emphasized in other parts of this user guide. The training module is included in its entirety in this appendix.

The NRI serves as the Federal Government's principal source of information on the status, condition, and trends of soil, water, and related resources in the United States. The assessment of the Nation's natural resources and environmental condition provided by the NRI requires a nationally consistent database that allows for manipulation of various parameters in order to make accurate assessments. For these manipulations and subsequent analyses to be useful for public policy considerations, the NRI database —

- has a high degree of reliability,
- looks at natural resources over a period of time, and
- permits analyses at national, regional, and local levels.

The NRI is unique because—

- it provides a nationally consistent database for all nonfederal lands,
- it features data gathered and monitored in 1982, 1987, 1992, and 1997 by thousands of technical and natural resource data collection experts, and
- it has a direct correlation with soils data, which permits analysis of numerous resources in relation to the capability of the land and in terms of soil resources and conditions.

Informed and valid decisions are best made when based upon scientifically derived data. The NRI has been developed to supply this type of information. The NRI database is powerful, credible, and scientifically constructed. The present NRI is the result of many decades of development.

Factors influencing the NRI

Numerous factors influence the methodology and design of each NRI. Some of the important factors considered when developing an NRI include:

- Mandates--information required/requested by the Rural Development Act and the Resources Conservation Act, information needed to support recent Farm Bill legislation, and various laws dealing with environmental issues.

- Data needs--resource information needed by NRCS, USDA, and legislative bodies; information needed by other agencies to examine agricultural, natural resource, and environmental issues; modifications to assist the research community; requirements for educators and others who disseminate information.

- Timing or schedules--deadlines for product completion, data delivery to properly support decision-making.

- Available funding and staffing--resources and staff hours available for data collection, individual staff available.

- Properly trained and experienced staff.

- Impact on staff--modification of sample to minimize travel by data collection staff.

- Previous inventories--lessons learned about data collection procedures, training, data capture, and processing from both NRIs and Conservation Needs Inventories; matching the existing sample to objectives; modifying the existing sample rather than developing a new national sample; providing historical data for existing sample locations for the trending database.

- Available technology--technologies that are available, suitable for the data to be collected, and feasible.

- Vision--new emerging issues, data needs, modifications to the NRI.

Goals for the 1997 NRI effort

The first step in the design process was to identify goals, or reliability levels for the data. For the 1997 NRI, four general goals were set:

(1) Provide estimates of 1997 conditions

Obtain sufficient data to permit analysis of natural resource issues at the multi-county or sub-State level such as:

- areas comparable in size to Major Land Resource Areas,
- 2- or 4-digit hydrologic units such as sub-basins or aggregated sub-areas,
- NRCS Administrative Areas as they existed in the early 1980s, and
- comparably sized areas.

Historically, this level has been referred to as MLRA level for the NRI.

(2) Provide estimates of change between 1982 and 1997

For many items, estimates of the change from 1982 to 1997 are statistically significant at the State and some multi-county sub-State levels. Examples include: changes in erosion rates for class II cropland, loss of prime farmland to urban and built-up development uses.

Some items are rare, occur infrequently, or change slowly over time. Detecting changes in these features is difficult unless quite different inventory techniques are used. Examples include: changes in wetland type and changes from Forest Land to Pastureland for class W land. For rare items, significant changes can only be detected at regional levels.

(3) Continue to develop the multi-resource concept

To support ecosystem based planning and assistance, and to properly analyze environmental and ecological issues, dozens of site-specific pieces of information are collected for each sample site. For the NRI, data items covering diverse landscapes and features are included, and linkages to databases such as soils, climate, and forest productivity can be established.

(4) Enhance opportunities to examine spatial trends

Analysis using NRI data takes into account the geographical or spatial nature of issues related to land use, natural resources, and the environment. Spatial analysis is facilitated by geographic information systems. Spatial linkage to databases is provided. The use of NRI data within a GIS is discussed in Chapter 6, NRI Data Analysis.

C.2. PROBABILITY SAMPLING

Alternative data gathering techniques

Before examining specific NRI techniques, discussing data collection methods in a more general way is useful. Many factors influence how this is usually accomplished. Consider the following common methods for collecting data:

- Complete enumeration—Census or wall-to-wall data collection; used in past years for watershed studies, and by the Census of Agriculture.

- Scientifically designed probability sample.

- Subjective sampling methods—In this method selected representative fields are chosen for observation. A certain amount of bias is introduced into the system from the beginning because a somewhat arbitrary decision has already been made as to what is a representative field. This method will rarely hold up when challenged, but research does show there are a few special situations where this methodology can be useful and efficient. For example, consider determining the average weight of assorted pieces of gravel within a bucket. Selecting one or two pieces subjectively can provide better results than a completely random selection. However, as the chosen sample size increases (becomes five or six pieces), random selection can produce better results because it more accurately reflects the size ranges within the bucket.

Also, soil scientists have determined ways to effectively use subjective sampling, for example, through the use of typical sites or pedons.

- Assimilation of available information — compile whatever information is available. This involves using information collected with varying techniques, definitions, and time frames.

  Examples are available where creative analysts have been able to do a good job of developing plausible scenarios, but there are cases where questionable conclusions have been reached. For example, the Economic Research Service has developed a reputable land use series for several decades, using data from various sources. On the other hand, the National Agricultural Lands Study developed questionable conclusions about the rate of conversions of agricultural land occurring during the 1970s.

The NRI employs probability sampling to determine where to collect data. This means data are collected for only a portion of the possible locations.

Samples are selected in a scientific rather than subjective manner that uses scientifically developed and broadly accepted methodology based in statistical theory.

Advantages of sampling

There are many reasons for using sampling techniques rather than complete enumeration when collecting resource data. Several of the obvious and less obvious advantages of sampling are:

- Costs less to conduct
  Data are collected for only a fraction of the universe or population, such as farms, tracts, people, and cities.
- Involves fewer people
  A smaller number of data gatherers is required because less data are collected.
  Fewer supervisors and total staff are required. Controlling data-gathering activities that involve smaller numbers of staff is easier.
- Takes less time
  A smaller quantity of data is collected, so all steps of the process can be accomplished in less time. This means final results are produced and published in a more timely manner.
- Produces more accurate and reliable results
  Samples can produce higher quality data because fewer staff members are involved; this means training, supervision, and quality control activities are more directly managed and easier to implement.
- Allows estimation of accuracy

Properly conducted sample surveys provide a means for evaluating the accuracy of the results. Sampling variability and measurement error can be estimated with less effort than for a census-type operation.

- Is statistically defensible
  Statistical sampling methods used in the NRI are accepted in the scientific community as sound inventory methodology.

C.3. NRI SAMPLING DESIGN

Background

NRI data are collected at scientifically selected sample sites located in every county and parish throughout the United States. The sample design is the result of more than 40 years of research and application.

The goals of the agency's Resources Inventory program have evolved over decades of data use. As a result, modifications or changes were made to the sampling procedure, as well as to the inventory procedures and protocols to meet this evolving need. Data from previous inventories have been used to adjust and continually refine the survey sample.

It is critical to have a basic understanding of NRI data in order to fully utilize this extensive data source when examining land use, natural resources, and related environmental issues.

Knowledge of the principles behind the NRI sampling scheme, as well as where and how the data were collected, is essential to proper use of the data. This knowledge enhances an analyst's ability to explain the use of NRI data for examining specific natural resource related issues.

Basic concepts

Samples for the NRI are selected using a stratified, two-stage, area sampling scheme.

Stratification

Stratification serves to make sampling more efficient by subdividing the entire population of interest, or data universe, into non-overlapping portions or layers, or "strata," that are more homogeneous than the population as a whole.

Administration, data collection, and other operational procedures can be affected or improved by using stratification.

Area sampling

For an area sample, the sampling units are areas of land.

When drawing a sample of people, farms, fields, or landscapes you can either:

> Use a list frame—compile a complete list of every person or field, and then select a sample from this list; or
>
> Use an area frame—divide the entire landscape area into pieces or tracts of smaller areas. Then select a sample of the pieces or areas and collect data for those units within these selected areas.
>
> For example, to select a sample of people within a city, it is common to select city blocks as the pieces of the city.

For the NRI, this area sample method is used and described in the next section.

Two-stage sampling

Sample units are selected in two stages. This allows clustering of sample sites, which usually reduces data collection costs.

For example, to select students for a survey, it would be common to select schools as a first stage and to then take a sample of students from each selected school.

Stratification of schools and of students within schools would make sampling more efficient.

The NRI two-stage selection procedure is discussed in the following sections.

Selection of the sample

Sample design and selection for the NRI are handled on a county-by-county basis.

The steps for sample selection include:

(1) Determine design -- specify the parameters needed to perform steps (2) to (4).
(2) Stratify total surface area of the county.
(3) Select first stage sampling units -- within each stratum, select one or more Primary Sampling Units.
(4) Select Sample Points - within each sample PSU, select specific locations (points).

The next three sections describe steps (2), (3), and (4) of sample selection.

Stratification

Each of the 3,100 counties in the United States was first subdivided or stratified into non-overlapping geographical strata.

In parts of the United States covered by the Public Land Survey system, geographic stratification was based upon legally described sections and townships.

Each township was subdivided vertically into three 2- by 6-mile strata for sampling purposes.

A section is nominally a one-mile square segment of land, and a township is typically a 6-mile square area consisting of 36 sections. Correction areas can cause sections and townships to be larger or smaller than the normal sizes.

In the southeastern half of Ohio and for counties in the Southern and Southeastern States not covered by the Public Land Survey System, lines analogous to townships and sections are superimposed on county maps, and geographic strata were formed from 2-mile by 6-mile tracts of land.

In the 13 Northeastern States, strata were formed using latitude and longitude. For Louisiana and the northwestern half of Maine, sampling was based upon the Universal Transverse Mercator grid system.

Some counties were also stratified according to specific resource conditions and general ownership patterns. This occurred most frequently in irrigated regions and urbanizing or suburban fringe areas.

Stratification was also altered in relatively remote homogeneous areas containing large tracts of range, forest or barren land.

In summary -- stratification is based geographically and uses the Public Land Survey System, latitude/longitude, or Universal Transverse Mercator for all counties. Further stratification was employed where it was advantageous such as where irrigation patterns, urbanization, and other conditions affect heterogeneity or homogeneity.

Selection of Primary Sampling Units

The Primary Sampling Units, called PSUs, or first stage sampling units were square plots. In the 34 States that have 2- by 6-mile strata, PSUs were generally 160-acre square parcels measuring 0.5 mile on each side.

In the Western States, PSUs were often 40- or 640-acre square parcels. The 40-acre units were used in most irrigated areas, and the larger PSUs were used in the relatively homogeneous areas containing large tracts of range, forest, or barren land.

PSUs in portions of the country covered by the Public Land Survey system are based upon the township/section grid. Therefore, they are frequently smaller or larger than the nominal sizes of 40, 160, or 640 acres.

In the 13 Northeastern States, PSUs were defined to be 20 seconds of latitude by 30 seconds of longitude, ranging from 97 acres in Maine to 113 acres in southern Virginia. In Louisiana and part of northwestern Maine, PSUs were 0.5 kilometer squares or 61.8 acres.

Within each stratum, one or more PSUs were randomly selected.

Note: For counties with 2-mile by 6-mile strata and 160-acre PSUs, there were 48 possible 160-acre square plots that could have been selected. Selecting one out of the 48 meant the sampling rate was about 2 percent for that stratum. Often two or more PSUs were selected within a stratum. These were color-coded, usually red, blue, green, or purple so that a ready-made smaller sample already existed if needed.

Selection of points within PSUs

The last step in selecting the sample was to locate three sample points within each PSU. There were exceptions - two points were selected from 40-acre PSUs and only one point was selected per PSU in Louisiana and northwestern Maine.

The procedure for selecting the points within a PSU was as follows:

1. A grid consisting of squares formed with three rows and three columns was superimposed on the PSU. Each square was subdivided into four equal blocks. The numbers 1 to 12 were assigned to the blocks in each row with a number appearing once in each row and once in each column No adjoining blocks had the same number.

2. Two numbers between 1 and X were selected at random, where X is the width of the side of the PSU in feet. These two numbers determine the coordinates of sample point #1 in feet north and east from the PSUs southwest corner.

3. Points #2 and #3 were located in the blocks with the same label as the block for point #1. They were positioned in the same relative position within the blocks as point #1. Steps for selection of two sample points within a PSU were similar, except the PSU was divided into four blocks instead of 36.

NRI sample size

The NRI base sample consists of about 300,000 PSUs. These PSUs constitute a sample of about 3 percent of all the total land and water areas of the 48 conterminous States, Hawaii, Puerto Rico, and the U.S. Virgin Islands.

Sampling rates varied across strata, but typically ranged between 2 percent and 6 percent. They generally were consistent across part or all of a county. The rate of sampling for any given stratum depended upon such factors as land use and soils patterns, MLRA and hydrologic unit distribution, and county size.

The sampling scheme was basically a compromise, attempting simultaneously to balance the workload among NRCS field offices and fulfill the design objectives of the study.

Features and summary of the sample

The present NRI sample is the result of over 40 years of research and development performed collaboratively between NRCS and the Statistical Laboratory at Iowa State University. Both organizations have well known experts in probability-based survey sampling and estimation techniques.

In general terms, the NRI samples have been selected using a stratified, two-stage, area sampling scheme. Nationally, there are 800,000 sample sites, or point locations, in the current base-line sample. Probability-based sampling is used to ensure credibility by eliminating bias and to provide the ability to assess the accuracy of results.

The samples have been developed on a county-by-county basis. This was done mostly for operational reasons, but this has also allowed use of information developed and reported by other agencies on a county basis.

Sampling intensity varies from county to county and often within a county, resulting in unequal probability sampling. Higher sampling rates occur where land use and resource patterns are more heterogeneous. This unequal probability sampling easily accommodates special features and interests. Additional features of the NRI sampling process include:

- Process is complete—it is for all landscapes and for the entire United States.
- Design is concurrent and consistent for all data themes and regions. This facilitates a multi-resource- or ecosystem-based approach to analysis and provides consistent national interpretations.
- Flexibility is provided. The sample can be easily modified by augmenting or sub-sampling as emphasis changes over time.

C.4. CONFIDENTIALITY OF SAMPLE LOCATIONS

It is Natural Resources Conservation Service policy that actual locations of NRI sample sites are not public information and are not to be released to the public. The policy was established to maintain the integrity of these sample sites so that they will not be affected or treated differently, and hence lose their randomness. This policy is also in effect to maintain the privacy, confidentiality, and cooperation of landowners and operators.

The U.S. Department of Agriculture, NRCS, and the Iowa State Statistical Laboratory have much at stake regarding the continued confidentiality and integrity of the NRI sample sites.

Many dollars and staff years have been expended to develop this sophisticated and extensive set of samples. A great deal is now known about these hundreds of thousands of samples.

C.5. TYPES OF DATA

There are two types of sampling units -
- The Primary Sampling Units, and
- The specific sample points within the selected PSUs

For NRI data collection, taking advantage of these two stages of sampling units has been beneficial for producing differing types of statistical estimates. Most data items are collected for the sample points, but several are collected for the entire PSU.

The NRI database is constructed using:
- NRI County Base data
- Data collected for each first stage sampling unit (PSU)
- Data collected at the designated specific points within each sample PSU
- Ancillary data

The NRI database is constructed by consolidating all of these data into a convenient format called Point-Type Information.

NRI Geostatistical Control Data

These data were derived from geospatial layers (Fuller et al. 2001, Nusser et al. 1999, Nusser 1997). Acreages are derived for each control unit of the sample for the following:
- The area of each control unit (4-digit HUA intersected with county polygon).
- The area of "Census" water (water bodies 40 acres or more in size, streams at least one-eighth mile wide) within each control unit.
- The area of Federal land within the control unit for 1992 and 1997.

PSU-level data

These types of data are collected on the entire PSU (Fuller 1999):
- Farmsteads and ranch headquarters
- Small built-up areas
- Large urban and built-up areas
- Large streams and water bodies
- Small streams and water bodies
- Rural roads

As part of the data collection process, these features were delineated, labeled, and measured on PSU support maps. The size category for water bodies was recorded. Area measurements, such as acres or dimensions, were needed for all of these features.

These features occur as small areas, have linear features, or possess other unique patterns. Most would be subject to bias if sampled only on a point basis.

Note: If more detailed data are needed, for example, for water areas or urbanizing areas, the NRI PSU-level data provide a powerful sampling frame from which to start.

Major Land Resource Area and 8-digit Hydrologic Unit code are recorded for the PSU. Also recorded are two factors used in erosion equations:

- R-factor that is the Rainfall factor for water erosion
- C-factor that is the Climate factor for wind erosion

Point-specific data

The third level of data collection was at the designated sample points located within each sample PSU. Most NRI data elements are collected on a point basis; this is the most extensive level of data collection. All land cover/use types are recorded at the points, so in effect data for farmsteads, water, urban and built-up, and rural transportation are collected at more than one level.

Ancillary data

Ancillary data are used in the estimation procedures developed and implemented at the Statistical Laboratory. These data come from other sources and are used to calibrate NRI results. Further details are provided in the NRI Database module.

Many types of ancillary materials that were assembled and used during the NRI data collection process are discussed in the next module.

C.6. POINT DATA ELEMENTS

An important facet of inventory design is determining what data collection method to use for each particular data item. This is sometimes referred to as response design, or plot design.

The PSU-level data and NRI County Base data represent two particular response designs. Collection of point-specific data represents several more response designs.

Multiple response designs have evolved for point data because natural resource and land use concepts are not really 0-dimensional or point-type concepts. They need to be considered spatially, in two or three dimensions.

> Collection of point data can be thought of as follows. The NRI needs to characterize the sample point. What area (or polygon) or spatial portion of the landscape does the NRI use?

Three central concepts used by the NRI are:

- For many features, think of the point falling within a field or land unit. Characteristics of this parcel of land are then determined and recorded. Examples include Land Cover/Use, and Conservation Practices.
- Wetlands and soils data items represent a similar response design - except these spatial features typically do not occur relative to fields or land unit. They have natural boundaries determined by a combination of technical definitions and mapping conventions.
- For some features, a transect can help characterize the point. The NRI uses transects as a response design for erosion determinations and rangeland investigation.

Development of appropriate response designs also includes consideration of such factors as:

— accepted and available classification systems
— utility for those planning to use the data
— concepts and systems that can be successfully implemented during data gathering
— replication of measurements and determinations by other data gatherers
— providing consistent, scientifically credible information through monitoring or remeasurement in subsequent inventories

APPENDIX D
Geographic Information System Coverages

Caution for Maps of NRI Data

NRI sample data are generally "reliable" for State and certain broad sub-State area analyses. Generally, analyses that aggregate data points by smaller geographic areas and/or more specific criteria result in fewer data points for each aggregation and therefore less reliable estimates. Users should generate and refer to the estimated standard errors or margins of error to determine suitability of the data for a particular purpose. It is important to remember that values for individual geographic units are often not sufficiently reliable for certain analytical purposes. NRI maps reflect broad national patterns rather than site- specific information.

NRCS has developed some spatial data coverages for mapping 1997 NRI data. Work is in progress to further test and update the coverages that exist as of December 2001. For the latest information on what spatial coverages exist and their status in relation to the 1997 NRI data, please visit our GIS Coverages Web site at:

http://www.nhq.nrcs.usda.gov/land/aboutmaps/coverages.html

GIS coverages include:

- County and FIPS
- Hydrologic Unit
- Major Land Resource Area

Other coverages may become available later in 2002.

APPENDIX E
Where to Get Help and Give Feedback

As new information regarding NRI data is developed, it will be posted to the NRI Web site at:

http://www.nhq.nrcs.usda.gov/NRI *

GET HELP:

If assistance is needed while using this CD-ROM, contact the NRI Help Desk at:

email: **nri@nhq.nrcs.usda.gov**

User messages will be responded to in the order they are received;
response time will vary depending on the volume of requests.

GIVE FEEDBACK:

NRCS is interested in comments and suggestions for improvement of this CD-ROM. In particular, users are encouraged to provide feedback regarding the documentation. A Version 2 may be developed if additional documentation will enhance the utilization of the data.

Please send feedback via: email at **nri@nhq.nrcs.usda.gov** or mail to

NRI Help Desk
Resources Inventory Division
5601 Sunnyside Avenue, Stop Code 5475
Beltsville, MD 20705-5475

* Note regarding Web addresses:

NRCS is in the process of a Web modernization effort to improve all of our Web pages. Efforts are being made to keep current Web addresses from changing. However, if a "file not found" error is received when locating one of the addresses cited on this CD, please try one of the following to find us.

1. **http://192.153.14.2/NRI**
   (if needed, a link will be added to this page to direct user to a new location)
2. **http://www.nrcs.usda.gov**
   Look for "technical assistance," then "National Resources Inventory."
3. Email a message to **nri@nhq.nrcs.usda.gov**