Environmental and Ecological Statistics **4,** 181–204 (1997)

The National Resources Inventory: a long-term multi-resource monitoring programme

S.M. NUSSER[1] and J.J. GOEBEL[2]

[1] *Department of Statistics, Iowa State University, Ames, IA 50011–1210, USA*
[2] *Natural Resources Conservation Service, US Department of Agriculture, Washington, DC 20013–2890, USA*

Received December 1995. Revised January 1997

Interest in natural resources and the environment has led to the development of new federal monitoring efforts, the expansion of existing federal inventory programmes, and discussions of inter-agency collaboration for natural resource assessment data collection. As federal programmes evolve, knowledge gained from existing long-term survey programmes can provide valuable contributions to statistical and operational aspects of survey efforts. This paper describes the National Resources Inventory (NRI), which has been conducted by the US Department of Agriculture's Natural Resources Conservation Service in cooperation with the Iowa State University Statistical Laboratory for several decades. The current NRI is a longitudinal survey of soil, water, and related environmental resources designed to assess conditions and trends every five years on non-federal US lands. An historical overview is provided highlighting the development of the survey programme. Sample design, data collection, and estimation procedures used in the 1992 NRI are described, and statistical issues related to long-term monitoring are discussed.

Keywords: environmental statistics, imputation, natural resource surveys, survey sampling, two-phase estimation

1. Introduction

Interest in the environment has led to an increased demand for credible information on the nation's natural resources. Several new federal data collection initiatives have been proposed, and long-standing inventory and monitoring programmes have been broadened. As a result, the objectives of various efforts have begun to overlap, and there is interest in integrating programmes.

Considerable knowledge and experience have been gained from conducting the surveys associated with existing long-term monitoring programmes. This information can be beneficial in the development of new initiatives for natural resource assessment, and in the broader application and integration of existing programmes. The purpose of this paper is to provide a comprehensive description of the National Resources Inventory (NRI) and to discuss statistical issues associated with long-term multi-resource national monitoring efforts through the example provided by the NRI.

The NRI is a longitudinal survey designed to assess conditions and trends for soil, water, and related natural resources on non-federal lands in the United States. The NRI and related natural resource assessment surveys have been conducted by the US Department of Agriculture's Natural Resources Conservation Service (USDA NRCS, formerly the Soil Conservation Service) in cooperation with the Iowa State University Statistical Laboratory for four decades. Currently, the primary survey is conducted every five years. The procedures used to collect and analyse NRI data have been developed over time under a broad set of political conditions and scientific objectives. 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. Estimation procedures have been developed in response to the increasing complexity associated with repeated surveys, including methods to impute various kinds of missing data and weighting procedures that incorporate controls from other data sources and from previous surveys so that current estimates are consistent with published figures.

Although much effort has been devoted to the design of monitoring programmes, until recently relatively few articles have been published in the statistical literature that describe sample design, estimation procedures, and operational considerations for long-term monitoring studies. In connection with the European Year of the Environment in 1987, McCormack (1988) outlines several statistical challenges induced by spatial correlation in the environmental sciences, including issues in sample design and modelling when estimating changes in land use and environmental status over time. Saeboe (1983) describes systematic point sample designs as applied to Norway's land use and natural resource monitoring programmes. The multi-tiered systematic sampling design and Horvitz–Thompson estimators proposed for the US Environmental Protection Agency's Environmental Monitoring and Assessment Program are discussed by Messer *et al.* (1991), Peck *et al.* (1992), Stehman and Overton (1992, 1994a, 1994b), Stevens and Olsen (1992), Cassell (1993), and Overton and Stehman (1993, 1994). Statistical issues associated with conducting forest inventories, with detailed discussions of the US Forest Service's Forest Inventory and Analysis programme, are presented by Schreuder *et al.* (1993), Schreuder and Czaplewski (1993), Powell *et al.* (1994), and Birdsey *et al.* (1995). A special session at the 1995 Joint Statistical Meetings included presentations on the US Geological Survey's National Water Quality Assessment programme sample design for monitoring surface and ground water quality (Helsel, 1995) and the use of simple sample designs for long-term regional monitoring efforts (Overton and Stehman, 1995). Published statistical literature that describes issues related to the NRI includes a discussion of least squares estimation for panel surveys using NRI as an example (Fuller, 1990), procedures for two-phase estimation through imputation as applied to NRI (McVey *et al.* 1994), and a brief discussion of NRI in relation to broader sampling issues for regional monitoring studies (Nusser, 1995).

This paper provides an overview of the history of the NRI and of the statistical methods used to conduct the 1992 NRI and to develop the 1992 NRI data base, which contains 1982, 1987, and 1992 inventory data. Our interest is in providing a discussion of the statistical aspects of the NRI as impacted by operational considerations. Because space does not permit detailed descriptions of the complex set of statistical procedures associated with the NRI, a more general presentation of these methods is provided. We begin with an historical overview that outlines early Soil Conservation Service resource inventories and factors affecting the development of the NRI programme. Sample design, data collection, and estimation procedures are described for the 1992 NRI, and plans for future surveys are outlined. Statistical issues related to long-term natural resource monitoring are discussed.

2. Historical overview of the National Resources Inventory

USDA natural resource data collection efforts date back to the 1930s, when there was concern over severe soil erosion occurring during several years of drought. The need to gather information on the status of agricultural lands led to the formation of the USDA Soil Conservation Service (SCS) in 1935, now called the Natural Resources Conservation Service (NRCS). The early USDA surveys were based on reconnaissance methods and information gathered from a variety of sources. As the field of survey sampling began to mature, the SCS adopted probability sample designs to support their inventory efforts. Over the years, the objectives and designs for these efforts have been modified in response to increased demand for multi-resource information and as a result of practical knowledge gained from conducting surveys.

A brief historical summary of USDA natural resource surveys is presented in Table 1. Information on the objectives, sample design, data collection, and special features are outlined. Further details are available in USDA (1945, 1962), Goebel and Schmude (1982), Goebel *et al.* (1985), Goebel and Baker (1987), Baker *et al.* (1992), and Goebel and George (1996). In reviewing the evolution of the sample surveys conducted since 1958, several features are notable. First, a constant feature of the sample surveys is the use of numerous small strata. Finely divided strata were initially used to achieve geographic spread throughout the country and to assign unequal selection probabilities to areas. These small strata have also supported repeated alteration of study domains and redefinition of sampling rates as design requirements have changed.

Over time, the NRCS surveys have come to rely on detection of change over time through repeated visits to points. In the 1958 National Inventory of Soil and Water Conservation Needs (also referred to as the 1958 CNI), the sampling unit was an area segment. Points within segments were adopted as sampling units for the 1967 CNI to reduce resources required to obtain data. However, no 1958 data were recorded for the 1967 points, making it difficult to discuss factors influencing net changes in land use. This experience emphasized the need to return to specific points to obtain information on land use dynamics leading to observed net changes in land use.

As the sample design has become increasingly complex, weighting procedures have also become more sophisticated. The early surveys relied on a simple ratio adjustment of inverse selection probabilities so that the weights for points within a county summed to the county's surface area as reported by the Census Bureau. The 1958 survey is an early example of providing summary statistics that are consistent with figures published by other federal agencies.

As surveys have been repeated over time, the need to calculate weights that are consistent with statistical summaries for previous time points has arisen. That is, if historical data for sample units are included in the data base along with the current study year's data, it is often desirable to have the weighted summaries of the historical data contained in the current data base match closely with estimates published from the earlier survey. This is particularly true for special studies, which typically have smaller sample sizes. Weights for these smaller studies can be improved by incorporating more precise estimates obtained from larger surveys conducted during a prior year as controls in the weighting process. Recently, more complex statistical models have been used to improve the accuracy of weights when sample sizes within a region are small. For example, urban surface area controls have been generated using urban population data in small area estimation models.

Table 1. Historical overview of NRCS natural resource surveys.

Year	*Survey*	*Objective*	*Sample design*	*Approximate sample size*	*Estimation*	*Information collected*	*Additional comments*
1934	National Erosion Reconnaissance Survey	Assess extent and effect of soil erosion on agricultural production	Reconnaissance			Anecdotal information on erosion	Led to passage of Soil Conservation Act of 1935
1945	Conservation Needs Inventory	Support programme development and assign conservation priorities	Assimilation of available data			Land capability, land use, and conservation needs, by Land Resource Area	
1958	National Inventory of Soil and Water Conservation Needs	Estimate magnitude and urgency of conservation measures needed to maintain and improve productive capacity	Stratified area sample; 40, 100, 160, or 640-acre square parcels of land used as sample segments	Sampling rates ranged from 1% to 8%, depending upon size of county	Inverse selection probabilities, adjusted by county to known inventory acres (acres of non-federal rural land)	Acreage for soils and land use classes for each sample segment; local committees estimated conservation treatment needs for 1958 and 1975, by county	First use of probability-based sample designs; multi-agency effort; commonly referred to as Conservation Needs Inventory (CNI)
1967	National Inventory of Soil and Water Conservation Needs	Update 1958 survey data	Stratified two-stage area sample; minor modifications to 1958 sample; points systematically selected within each 1958 segment	38 points in a typical 160 acre segment	Similar to 1958 procedure, but weights developed for each sample point	Soil type, land use, and conservation treatment needed	Two-stage design adopted to reduce data collection costs; first tabulation data set produced; only net change could be estimated
1975	Potential Cropland Study	Identify amount, location, characteristics of land that could easily be converted to cropland to support world-wide demand for agricultural products	Stratified multi-stage subsample of 1967 sample points; selected counties, segments within counties, points within segments	506 counties, 5300 segments, 41 000 points	1967 weights multiplied by inverse subsample selection probabilities, adjusted to reflect state inventory acres	1975 land use, factors affecting conversion to cropland, prime farmland status; also had soil type and 1967 land use for each sample point, from 1967 CNI	Use of national sample as frame for special study; paired observations (observations for same sample sites for 1967 and 1975) allowed investigation of land use dynamics

1977	National Resources Inventory	To obtain current information on status, condition, and trends of soil, water, and related resources, as mandated by the 1972 Rural Development Act	Stratified multi-stage design consisting of subsample of 1958 CNI segments; new selection procedure for points within selected segments	70 000 segments (i.e. PSUs); 1, 2, or 3 points per PSU, depending on size of PSU; 3 points in a typical 160 acre PSU	Similar to 1967 CNI procedure, modified to account for acres of farmsteads, built-up areas, small streams, and water bodies collected for sample segments; new procedures to establish inventory (control) acres	Elements similar to those collected in 1967 and 1975, plus wetland incidence, flooding propensity, and parameters needed to estimate water and wind erosion; built-up areas, farmsteads, small streams, and water bodies within the segment	Redesign of basic CNI framework; first nationally consistent survey of erosion, was originally called 1977 Erosion Inventory; provided basis for extensive 1980 appraisal, mandated by Soil and Water Resources Act (RCA) of 1977; established linkage with soil interpretations database
1982	National Resources Inventory	5-year update to obtain current information on soil, water, and related resources, including expanded ecological concerns; substate-level inference	Expanded the 1977 multi-stage area sample; new sample selected for 13 northeastern states using latitude/longitude for frame construction; additional samples included to support special local-area studies	321 000 PSUs to meet national needs; additional 44 000 PSUs for local-area studies; new PSUs in northeast were 20 seconds longitude by 30 seconds latitude and contained 3 sample points	Procedure from 1977 modified; inventory acres (controls) developed for Major Land Resource Area portions of counties; small area estimation used to improve built-up area estimates; see Goebel *et al.* (1985)	Added to 1977 list: soil properties from field visits, critical eroding areas, irrigation practices, wetland class, windbreaks, riparian vegetation, wildlife habitat diversity, vegetative cover, rangeland and pastureland condition	Modification of erosion equations and land use classes invalidated comparison with 1977; survey is initial time point in current NRI series; large increase in sample size and diversity of data elements due to analytical needs discovered during 1980 RCA appraisal
1987	National Resources Inventory	5-year update of 1982 NRI; for state-level inference	Subsample of 1982 NRI samples, using post-stratification based upon 1982 data; augmented 97 counties where analyses indicated the need for additional samples	108 000 PSUs, 4000 of which were new in 1987	Weights controlled so that 1982 estimates derived from 1987 data closely match published figures for 1982; small area estimation for urban change estimates based on data collected in special urban study	A subset of the items collected during 1982 NRI	Reduced resources led to substantial reductions in sample size and number of data elements, and to use of remote sensing for 30% of sample; prior information on PSUs/points and modelling used to increase efficiency and accuracy of estimates from smaller sample

Year	*Survey*	*Objective*	*Sample design*	*Approximate sample size*	*Estimation*	*Information collected*	*Additional comments*
1992	National Resources Inventory	5-year update on NRI series; for substate-level inference	Based on a large subsample of the 1982 NRI, with a small set of additional PSUs to augment sample size in various areas; included all 1987 NRI samples	300 000 PSUs	Imputation used to complete 1987 data not observed in 1987 and not retrospectively gathered in 1992; imputation of change patterns detected at PSU but not observed at point level; raking to account for inventory (control) acres	Similar to that of 1982 and 1987, with addition of new earth cover classification; used photo-imagery to collect 1987 data retrospectively for several variables	Data analysis software developed; many procedures implemented to ensure comparability between 1982, 1987, and 1992 observations
1995	Erosion Update Study	National-level inference	Multistage subsample of 1992 PSUs that were observed in the 1982, 1987 and 1992 NRIs; several states selected with certainty, other states selected pps from remaining list of states; counties selected, then PSUs	34 states, 343 counties, 3016 PSUs	Control acres included surface area and erosion rates from previous years	Land use, cropping history, conservation practices, tillage residue	Special study conducted for rapid assessment of special issue

The goal of creating data sets that make it easy to produce meaningful tabular summaries has been part of inventory efforts since the 1967 survey. As with weight calculations, initial procedures were relatively simple because the survey was based on a single sampling unit (segments or points). As the use of multi-stage designs and the number of observations made on each sampling unit have increased, the methods required to provide an accessible database have grown in complexity.

Finally, the objectives of NRCS surveys have expanded over time. Agricultural science has grown beyond production issues to focus on multiple resources, with increased emphasis on ecological communities and environmental quality. As a result of this shift, data are now being collected on a broader set of characteristics, and the analytical capabilities provided by the current multi-resource trending data base are finally being realized. An important consequence of the multi-faceted nature of the current NRI objectives is the capacity to conduct special studies on a wide variety of topics. In recent years, special studies have been conducted almost annually to investigate topics such as changes in wetlands, erosion rates on highly erodible land, and changes in field practices as a result of economic conditions and farm policies.

The evolution of NRCS surveys has resulted in a complex set of requirements for sample design, weighting, and analysis. The historical information on sample units provides flexibility in designing special studies, but presents challenges when sample units need to be added in specific regions. Inter-agency collaboration increasingly influences the selection of variables to be collected. Weighting procedures are also impacted by a complex set of controls generated by historical summaries and the need to reduce two-stage data to create a point-level data set for easy use by analysts. In the next section, we describe how statistical procedures associated with the most recent comprehensive survey, the 1992 NRI, address these issues.

3. The 1992 National Resources Inventory

3.1 *Overview*

The 1992 NRI represents the third survey in a series that was initiated in 1982. The 1992 NRI database contains data on sample units for the years 1982, 1987, and 1992. In what follows, we describe statistical procedures used during the 1992 survey, including sample selection, data collection, imputation of missing data, weight calculation, and database construction. Initial results from the survey are described in Kellogg *et al.* (1994) and USDA (1994a).

3.2 *NRI sample design*

The basic design of NRI surveys is a stratified two-stage area sample. The land area of the majority of states in the US is divided according to the Public Land Survey (PLS) system, which has provided a convenient structure for developing NRI sample selection procedures and for locating primary sampling units (PSUs) in the field. Counties or analogous units have been used to implement the basic design. We begin by describing design features for a typical PLS county, and then briefly outline modifications for areas based on other political subdivisions.

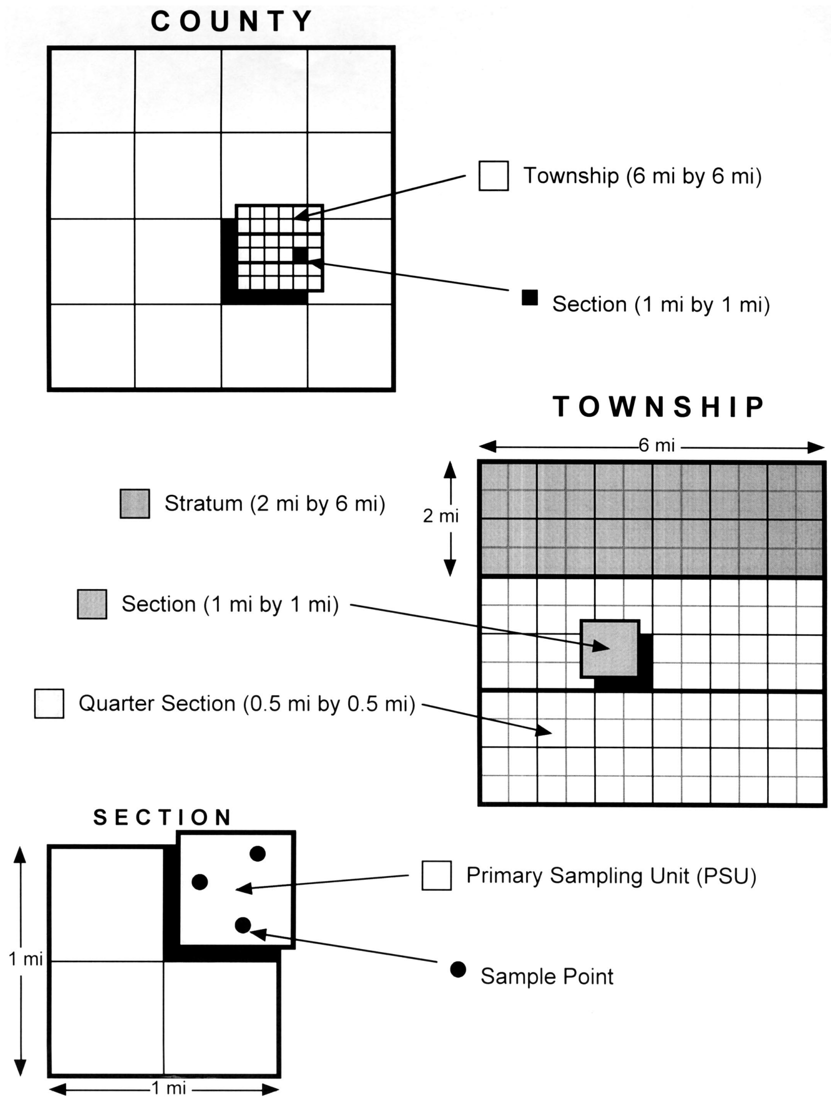

Figure 1. A graphical representation of the political subdivisions of a typical Public Land Survey (PLS) county, and the relationship of PLS subdivisions to a typical NRI primary sampling unit (PSU) and sample points.

Figure 1 provides a graphical representation of the political subdivisions of a typical PLS county. Under the PLS system, a standard-sized county is a square tract, 24 miles on a side, divided into 16 square townships, each six miles on a side. Each township is further divided into 36 square sections, each one mile on a side. NRI strata are defined by dividing townships into three two-tiered groups of 12 sections, which are 2 miles $\times$ 6 miles in size. The most common PSU is a 160-acre (64.75 ha) square quarter-section, 0.5 miles (0.8 km) on each side (Fig. 1). Under this arrangement, there are 48 PSUs per stratum, and an approximate 4% sampling rate is obtained by selecting two PSUs within each stratum of 12 square miles (Fig. 2). In the second

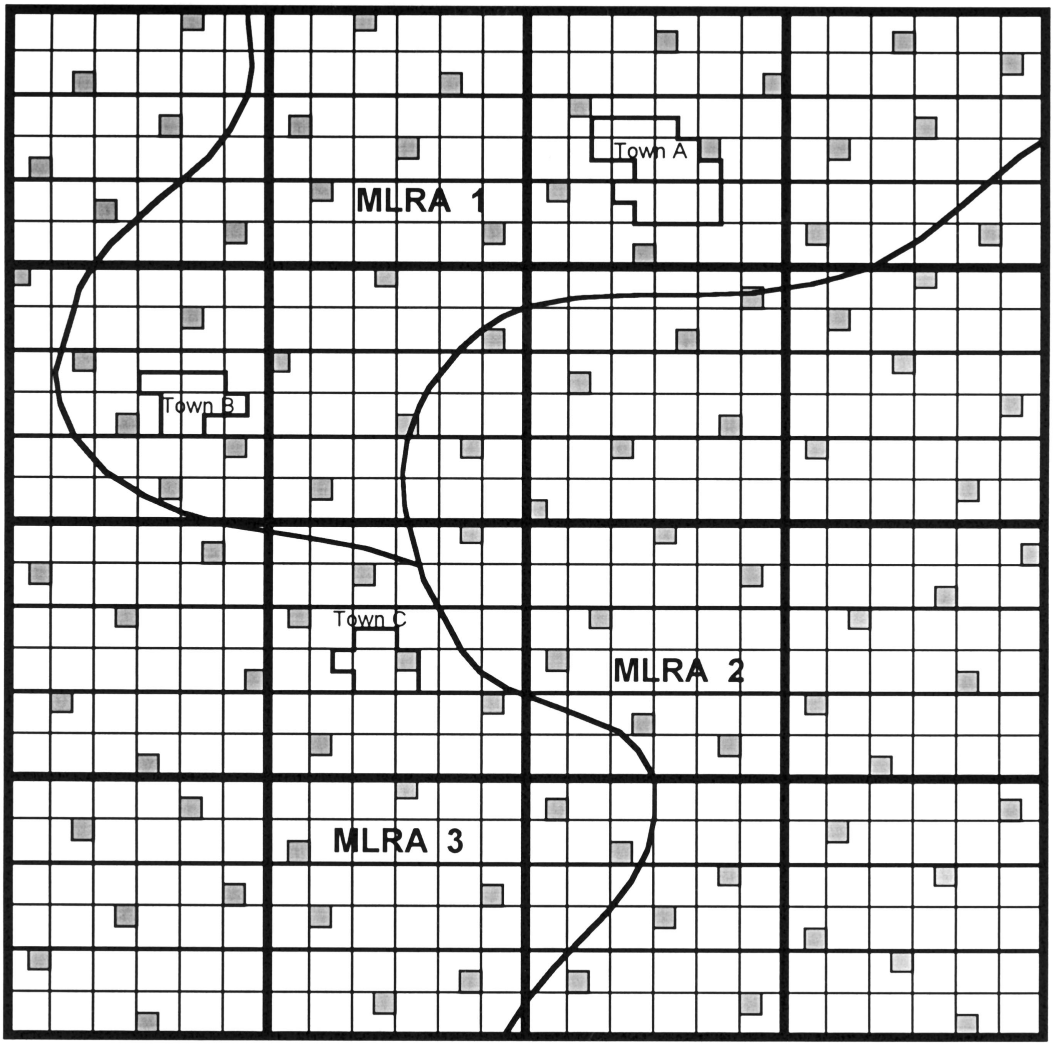

Figure 2. An example of a 4% NRI sample of PSUs in a county, with township boundaries (thick line) and Major Land Resource Area (MLRA) boundaries (curved line) displayed. The polygons defined by the intersection of county boundaries and MLRAs are called MLRACs.

stage of sampling, three sample points are selected within each PSU according to a restricted randomization procedure (see pp. 10–11 in Goebel and Baker, 1987).

Similar designs have been used for areas of the country not organized using the PLS system. For counties in the south-central and southeastern US, strata and PSUs were formed by imposing a grid system analogous to townships and sections. For counties in the 13 northeastern states, strata were formed using latitude and longitude, and PSUs were defined to be 20 seconds of latitude by 30 seconds of longitude. For Louisiana and the northwestern portion of Maine, sampling was based upon the Universal Transverse Mercator grid system, resulting in PSUs that are 0.5 kilometre squares.

There are instances throughout the US where sample design deviates from these standard procedures. For example, some PSU sizes are 40 or 640 acres, and fewer than three points are selected in some PSUs. Larger PSUs tend to be in areas whose natural resources are relatively homogeneous. Smaller PSUs tend to have fewer points. Also, stratification is altered in a number of counties based upon ownership patterns and specific resource conditions such as irrigation and urbanization.

For the comprehensive NRI surveys, sampling rates across the US generally range from 2 to 6% of the land area, though rates occasionally fall outside this range. The sampling rate within a county is increased when larger sample sizes are needed for special studies, or when heterogeneous patterns exist for irrigation soil types, land uses, major land resource areas, or hydrologic regions. For example, potential urban growth areas are sampled more intensively to study conversion of prime farmland to urban use. Sample size allocation is also frequently influenced by the availability of staff and funding.

The 1992 NRI sample contained 307 468 PSUs and over 800 000 sample points. The 1992 NRI sample was primarily selected from the 1982 and 1987 NRI samples. Over 6000 new PSUs were added for the 1992 survey to increase sample sizes in specific areas. The main 1982 sample consisted of 320 823 PSUs and nearly one million points, with an additional 43 966 PSUs selected for special studies conducted as part of the 1982 NRI. Staffing and funding constraints led to a 1987 sample of 107 757 PSUs, all but approximately 4000 of which were part of the 1982 sample. Most of these new PSUs were added in 116 counties where analyses of the 1982 data showed augmentation was needed to estimate changes in urban areas. All 1987 PSUs were included in the 1992 sample.

The objective in determining sample sizes for the 1992 NRI was to have a coefficient of variation of less than 10% for any estimate of surface area with a particular resource condition (or for other variables such as erosion rates) on areas that constitute at least 10% of the surface area in the multi-county region defined by a Major Land Resource Area (MLRA). MLRAs are distinguished by geography, soil, climate, water resource, and land use considerations (USDA, 1981). Other regions of interest include geographical areas associated with a common water drainage pattern as defined by the US Geological Survey's Hydrologic Unit system (USGS, 1982) and NRCS administrative areas.

3.3 *Data collection*

During the 1992 NRI, data were collected for three different reporting units: sample points, PSUs, and polygons defined by the intersection of MLRAs and counties, called MLRACs (see Fig. 2).

The majority of the variables were collected at the sample point, including land use and vegetative cover, soil information, agricultural and conservation practices, wildlife habitat variables, wetland classifications, and a number of condition-related factors for rangeland (Table 2). Variables collected for each PSU include surface area data for several land use categories, and information on federal ownership, regional natural resource classifications, and climate factors (Table 2). Definitions for the variables and procedures for data collection are explained in detail in the 1992 NRI instruction manual (USDA, 1991). Data gatherers were instructed to use photography and remote sensing materials to collect PSU and point data for the 1992 year, as well as county office records, field visits, and whatever additional materials were necessary to determine 1992 conditions. About one-fourth of the sample sites were field visited for

Table 2. Overview of variables collected for sample points, PSUs, and MLRACs.

Point	*PSU*	*MLRAC*
Land cover/use	Acres inside county	Rural roads and railroads
Soil classification	Ownership (all federal)	Federal land
Soil properties	MLRA	Large streams
Erosion factors	HUC	Large water bodies
Irrigation practices	Climate factors	
Conservation practices	Windbreaks (kind and size)	
Cropping history	Farmsteads	
Highly erodible land	Small urban (< 10 acres)	
Prime farmland	Large urban	
Conversion potential	Small streams (< 1/8 miles wide)	
Vegetative cover	Large streams	
Habitat distances	Small water bodies (< 40 acres)	
Wetland classification	Large water bodies	
Rangeland conditions		

range and tillage subsamples in 1992. Nearly all sample sites had been field visited in 1982. Land use data for PSUs not included in the 1987 sample were gathered retrospectively. NRCS staff were also required to review and edit existing 1982 and 1987 data during the 1992 NRI.

The data collected for MLRACs are called county base data, and are used as control totals in weighting procedures. These data represent a census of US land for surface area, including acres of rural roads and railroads, federal land, and Census water (Table 2). As part of an inter-agency agreement, the sum of the acres for various land uses in the county must match 1980 Census surface area acres for the county, adjusted for changes in political boundaries since 1980.

PSU and point data were collected using a computer-assisted survey instrument (computer software with screens and edit checks designed to support keyentry of survey data) on desktop computers, with edits designed to check ranges and consistency among variables. Data gatherers were required to resolve all errors detected by the edit program before submitting the data to the Iowa State University Statistical Laboratory. Submitted data were merged with soils information and additional PSU and point edit checks were performed.

3.4 *Imputing missing historical data*

Procedures were developed to impute missing data for sample units not observed in 1987 and for PSUs and points that were new to the sample in 1987 and 1992. Imputation methods were designed to provide a complete final data set. We describe methods used to impute missing 1987 data, and outline modifications to this algorithm used to impute historical information for new sample units.

In the 1987 NRI, complete information was collected on about one-third of the sample points. During the 1992 NRI, many variables such as 1987 land cover/use were collected retrospectively through photo-interpretation and ancillary materials. However, 1987 values for cropping history, conservation practices, kinds of conservation treatment needed, water and wind erosion equation factors, and rangeland condition required imputation.

Missing data were estimated using a hierarchical hot-deck imputation procedure (Tollefson, 1992 and 1994). For each point containing missing data, a nested set of imputation classes was constructed that contained potential donor points with similar values for spatial location, land use, land capability, irrigation status, and wind erosion. The imputation procedure began with the smallest set of potential donors, consisting of points with the same land use, same irrigation status, same MLRA within the county, same land capability class group, and same wind erosion equation status. If this set was nonempty, a donor point was randomly selected, and data from the donor point provided the missing information. If no suitable donor was found, the imputation class was enlarged by relaxing conditions for class membership until donors were found. For example, the relaxation of the spatial location condition went from the same MLRA within the county (MLRAC), to the same county, to the four nearest-neighbour counties, to the state. Limits were set on the number of times a donor point could be used to prevent overuse of selected points.

Occasionally, photography was not available to provide 1987 values for variables used to control donor point selection. When this occurred, a probability scheme was used to complete the imputation procedure described above. A set of control variables were selected with equal probability from the 1982 and 1992 data, and the imputation procedure described above was used to complete the data for the remaining variables.

The same basic algorithm was used for other missing values in the 1992 NRI, for example, to provide 1982 and 1987 data for points that were added to the sample in 1992. All imputed values were required to pass the same edit criteria as observed data.

3.5 *Imputing points to represent county base and PSU data*

1992 NRI statistical procedures were designed to combine information contained in the county base control (MLRAC) data, PSU data, and point data to produce a complete point data set with one weight for each point. Completeness was achieved by imputing missing data as described in Section 3.4. Additional imputation of 'pseudo points' was used to transfer MLRAC and PSU data to point-level data. These procedures were required because county base (MLRAC) and PSU data may contain patterns of change over time that are not observed at a point. For example, the county base data might indicate a reduction in federal land from 1987 to 1992 for a particular MLRAC. However, there may be no points in the MLRAC showing a change from federal ownership in 1987 to non-federal ownership in 1992. The approach used in the 1992 NRI was to impute point data for land use patterns observed in county base data or PSU data, but which were not found in the point data corresponding to the MLRAC or PSU, respectively. Pseudo points were created using imputation procedures similar to those described in Section 3.4 to reflect the properties of land associated with the observed land use pattern.

The general procedure for imputing point data to reflect changes observed in PSUs is described in McVey *et al.* (1994) and Breidt *et al.* (submitted). A similar procedure was used for county base pseudo points (Fuller, 1996). Table 3 contains example PSU data in which the surface area for the large urban land cover/use category (≥ 10 acres of contiguous urban land) increases from 1982 to 1987, but shows no change between 1987 and 1992. The corresponding real (observed) point land cover/use data are also presented in Table 3. Note that the only point in the PSU that is associated with large urban land is classified as large urban for all three sample years (point 3). To reflect the change in large urban acres observed in the PSU, a pseudo (imputed) point is generated that is classified as a land use other than large urban in 1982 and as large urban in 1987 and 1992.

Table 3. Example of a 160-acre PSU and associated PSU and point data. A land use pattern is observed in the PSU data, but not in the point data. One realization of an imputed pseudo point for the PSU is presented.

PSU data

Land cover/use	*Surface area in PSU (acres)* 1982	1987	1992
Large urban ($\geq$ 10 acres)	15	20	20
Other land cover/use	145	140	140
TOTAL	160	160	160

Point data for the PSU

Point type	*Point ID*	*Land cover/use* 1982	1987	1992	*Initial acre weight*[a]
Observed	1	Small water	Small water	Small water	70 p^{-1}
Observed	2	Pasture	Hayland	Hayland	70 p^{-1}
Observed	3	Large urban	Large urban	Large urban	15 p^{-1}
Imputed	4	Pasture[b]	Large urban[c]	Large urban[c]	5 p^{-1}

[a] p = PSU selection probability
[b] Donated from point 2
[c] Donated from point 3

The 1992 NRI imputation procedures for pseudo points required point data from two sources of real data. Donor points were selected to impute data for a land cover/use classification and associated variables for the year(s) with unknown land use and to impute data for the year(s) with known land use. In the example, the unknown land cover/use for the pseudo point is the 1982 land cover/use that is not large urban, and the known land cover/use for the pseudo point is large urban in 1987 and 1992. For each pseudo point associated with a PSU, hierarchical imputation classes of donor points were defined based on location and on appropriate land cover/uses for each survey year. For the years corresponding to unknown land cover/uses, a set of acceptable land cover uses was determined using guidelines outlined in Fuller (1996, Table 1). Sets of donor points containing acceptable known and unknown land/use patterns were divided into hierarchical subsets on the basis of location relative to the PSU. Donor points were always selected from within the PSU whenever possible. Subsequent imputation classes were defined by relaxing geographical constraints as described in Section 3.4. All pseudo points were required to pass the same edit checks as those applied to observed data. The final database contained 863 185 real points and 201 664 pseudo points.

Returning to the example, a donor point is needed that has a 1982 land cover/use that is an acceptable preceding cover/use to large urban land and a second donor point is needed that is classified as large urban in 1987 and 1992. The latter donor point is available within the PSU via point 3 (Table 3). Point 3's 1987 and 1992 data are used to impute the 1987 and 1992 pseudo point data. Acceptable land cover uses for the 1982 data are cropland, pasture, rangeland, forest land, and small built-up land (Fuller 1996, Table 1). Since two points (1 and 2) have 1982 land cover/uses that are acceptable, one is randomly selected from the pair with equal probability. In Table 3, point 2 is selected, and its 1982 data are used to impute the pseudo point's 1982 data.

3.6 *Initial weights*

After imputation was completed, initial acre weights were determined for real points and pseudo points that reflect acres for the observed PSU and county base patterns.

When real points were available in a PSU that exhibited land use patterns consistent with observed PSU dynamics (i.e. no pseudo point imputation was required), initial weights were constructed by allocating the acres associated with that land use pattern to the corresponding points. When there was more than one real sample point in the PSU that matched the observed land use patterns in the PSU, the initial weight was determined by dividing the acres associated with the PSU land use pattern by the number of eligible sample points. The initial weight for a point was equal to the acres allocated to that point divided by the PSU selection probability.

A modification of this procedure was used for PSUs that required pseudo points. First, pseudo points were assigned the acres of the associated PSU land use pattern. Then the remaining acres were allocated to the remaining real points. Consider the PSU pseudo point example in Table 3. The initial weight for the pseudo point showing the change in land use from 1982 to large urban in 1987 and 1992 would be equal to the observed 5-acre increase in large urban land in the PSU from 1982 to 1987 divided by the selection probability p for that PSU (Table 3). The initial weight for a real point with large urban cover/use in all three sample years would be set equal to the 1982 PSU large urban surface area (15 acres) divided by the PSU selection probability. This allocation insures that large urban acres in any of the sample years for that PSU are properly reflected in the point data. The remaining 140 acres would be allocated evenly to points 1 and 2. The initial weights for points 1 and 2 would be equal to 70 acres divided by the PSU selection probability p.

Weights for county base pseudo points were constructed to reproduce the county base acre data, just as PSU pseudo weights were constructed to reproduce PSU acres. The calculations for county base pseudo points were performed prior to the operations for PSU pseudo points and real points. Initial weights attached to pseudo points that represented items in the county base were not changed (i.e. frozen) in subsequent operations. The weights for PSU points could be modified in subsequent operations.

3.7 *Weight calculation*

Fourteen categories of land use were used in constructing a final set of area weights for the point data set. Three categories were derived from the MLRAC county base data: roads and railroads, federal land, and Census water (Table 2). Categories associated with PSU elements were farmsteads, small water, small streams, small built-up land, and large urban land (Table 2). The remaining six categories were constructed from point variables. They are prime cropland, non-prime cropland, pastureland, rangeland, forestland, and other land, which consists of all remaining land uses. These 14 weighting classes were used to guarantee that data for the 1992 study were consistent with previous NRIs. Weights were constructed such that estimates of 1982 acres for these classes as derived from the 1992 NRI database were consistent with corresponding estimates obtained from the 1982 survey. The 1982 NRI is believed to be the best source of surface area estimates for these categories during the 1982 growing season. Some exceptions in consistency were permitted, for example, when substantial revision of the data had occurred as a result of revised definitions or protocols.

Final weights were constructed in several phases so that (1) weights within a MLRAC sum to the acres for that MLRAC, (2) weights for the 1982 cover/use for each of 14 cover/use categories within a state sum to the previous 1982 estimated acres, and (3) weights for Conservation Reserve Program (CRP) points within a state sum to known CRP acres for the state. Recall that county base (MLRAC) acres were considered to be known population totals. Thus, final weights for the three county base categories were the frozen weight assigned in the pseudo point generation procedure.

Totals for urban land for MLRACs were estimated using a special procedure applied to each state (Fuller, 1996). After urban pseudo points were generated and initial weights were assigned, direct estimates for small built-up (≤ 10 acres of built-up land) and large urban acres in each survey year (1982, 1987, 1992) were calculated for the state and for MLRACs by summing the weights for points within the state or MLRAC with small built-up or large urban land cover/use in the designated year. Small area estimation techniques (Platek *et al.* 1987) were used to develop a second model-based estimate of urban changes in a MLRAC. The model included 1982 urban acres and changes in human population totals within the MLRAC for each 5-year interval. The final estimate of urban acres for the MLRAC in 1987 and 1992 was the average of the direct estimate and the model estimates. The sum of the MLRAC estimates is the direct state estimate. The convex combination of the model and direct estimates for a MLRAC was fixed in advance.

The final weights for the remaining nine land cover/uses were determined using a sequential ratio estimation procedure in which the MLRAC totals for a land cover/use were adjusted so that the sum of the MLRAC acres was equal to the 1982 control and all weights were non-negative. Weights were estimated for one cover/use category at a time, starting with the cover/use with the smallest acreage in the state. As the cover/use totals were modified, the modifications were constructed so that the sum of cover/use acres in a MLRAC is the total acres in the MLRAC. It was not always possible to simultaneously satisfy the state cover/use and MLRAC acre constraints. When the sequential procedure failed to satisfy both constraints, the state cover/use controls were relaxed.

3.8 *Analysis database and software*

The final product of the estimation process was a point database with one weight per point. The weight contains information from control totals and PSU data in a manner such that estimates constructed with the point data set equal the control totals. Stratum and cluster variables are included in the database, and all pseudo points and points containing imputed values are identified.

NRCS has developed NRI Data Analysis Software (USDA, 1994b) that provides a usable front-end for those not familiar with the 1992 NRI longitudinal database. An interface has been created that overlays existing software for tabulating data and for calculating variances from complex surveys, called PC CARP (Fuller *et al.* 1986).

3.9 *Quality control*

Procedures to control data quality were necessary at several steps in the survey process. Numerous staff with a variety of backgrounds collected the data, and observations at a sample point were made by different data gatherers for different sample years. The NRI programme invested considerable effort in constructing specific instructions and in training personnel so that protocols were applied as uniformly as possible throughout the country. The 1992 survey was used to review and check historical data when possible using these procedures.

As noted in Section 3.3, a computer-assisted survey instrument was used to enter the data. The software included edit checks for illegal or unlikely values and for inconsistencies among data from related variables. In addition to point-of-entry error checking, extensive checking and editing was conducted by Iowa State University Statistical Laboratory staff prior to and during the application of statistical procedures. Edit routines were run on all imputed historical data and pseudo points.

With the complexity of relationships between variables and the wide range of conditions observed in NRI, some problems cannot be detected until sampling weights are applied or until specialists are able to review summary results. Dozens of tables were constructed for each state with weighted summaries of the data. The tables were used to reveal patterns that are not detectable at the reporting unit level, such as the absence or overabundance of a particular condition. For example, tables included county-level estimates of acres in specific crops, irrigated land, prime farmland, wetlands, and other categories, as well as county-level estimates of water and wind erosion rates. The tables were reviewed by state and regional inventory specialists. When problems were detected, specific point data were analysed to determine causes, and data gatherers were asked to re-examine the sample data. When necessary, technical staff developed solutions to resolve apparent inconsistencies observed in the data or estimation products.

Computer programs for summarizing NRI data were programmed by two independent staff as part of the quality assurance process. Output from editing runs, statistical estimation results, and summary tables all provided information that was used to revise and correct editing routines.

The actual editing and estimation process was quite complex and iterative. Problems detected at various phases in the process were used to improve the edit routines and imputation models. Before the final data set was produced, data for each state were run through a common set of editing, imputation and weight calculation programs that had been fine-tuned during the initial editing and estimation runs.

4. Discussion

4.1 *Sample design*

The choice of design for any survey depends on the objectives of the survey, the nature of the population, and operational constraints. The NRI is a national multi-resource inventory and monitoring programme whose purpose is to produce a longitudinal data base containing numerous agro-environmental variables that will support scientific investigations and policy-related analyses. Most of NRI's specific objectives have been related to agriculture and natural resource conservation. These resources tend to exhibit moderately slow rates of change, although rapid alterations in land use, resource conditions, and practices may occur in response to discrete shifts in policy or environmental conditions. The five-year survey cycle has been historically appropriate due to the relatively slow rate of change observed in these systems and the use of NRI data to support Farm Bill development every five years. Annual data collection efforts to address specific assessment needs in interim years has provided flexibility when rapid changes need to be monitored.

Sample design for natural resource monitoring studies is challenging because of the diverse patterns and scales of spatial and temporal variation, as well as variation in the biotic aspects of ecosystems. In general, efficiency is increased if the sampling effort is allocated as a function of the variation in measurements across these dimensions. In the case of NRI, heterogeneity in land use and resource conditions are highly correlated with underlying physiographic factors and human population patterns. Because of this correlation, the national sample provides adequate sample sizes for studying many natural resource phenomena and for summarizing data for a wide range of domains.

The redefinition of target populations and subpopulations is an integral feature of multi-resource monitoring programmes. For NRI, an important component in providing the flexibility to define domains *post hoc* is the selection of a few PSUs within each of the numerous small strata covering the land area of the US. This approach was adopted early in the history of the NRI programme based on research conducted as part of the development of the Master Sample of Agriculture during the 1940s. The Master Sample of Agriculture is a frame that could be used to select an area probability sample of farm, open country, and urban area segments. The frame was created to support numerous agricultural and economic surveys with diverse purposes. The design specifications would typically require geographic spread throughout the entire nation and the ability to specify domains for special study.

The NRI design defines strata based primarily on fixed geographic units (e.g. sub-townships), rather than subject matter variables (e.g. specific land cover/uses such as cropland, rangeland, forest land). This approach has facilitated survey administration, which tends to be organized by geographic and/or political units, and has avoided problems that occur as resources shift spatially. The use of these units in the design has made it simple to provide analyses for policy development and implementation, which have historically focused on political units such as states and counties. A nationally-integrated design also simplifies inference strategies for producing timely summaries of the survey data. Design-based estimators can be quickly calculated to estimate parameters for various domains, avoiding the complications associated with modelling relationships among a collection of special regions to make inferences on a larger population.

When constructing sampling plans for monitoring studies in which the location of target populations shift, it is often advantageous to develop a sampling frame that extends beyond the target population. The location and extent of the target population can be identified at each time point and a foundation is established to support future changes in the target population. The samples from NRI's precursor surveys were selected from the entire land base (at that time, the 48 conterminous states) as a means of identifying the location of and estimating the size of non-federal rural lands. Today, national coverage is still used to detect changes in the boundaries of the target population, which has grown to include urban areas for investigations of urban growth and its consequences on the availability and productivity of agricultural land. Millions of acres have changed ownership status since NRI surveys were first conducted. The sample units located on federal lands are also available for conducting an inventory of federal lands in cooperation with other agencies.

Strategies to alter sampling rates are required to meet the changes in national objectives that regularly occur in long-term multi-resource monitoring programmes. In addition, special studies may require that additional sample units be selected in a small region or that a subsample be selected from the existing national sample across a broader study area.

The national NRI sample has frequently been used as a frame for special studies, for example, to estimate changes in wetland status, to investigate pesticide usage in relation to agricultural practices and water quality, to assess rangeland conditions and trends, and to study conservation tillage practices. These surveys may be conducted as part of the comprehensive national survey or as an independent study conducted during an interim year. The most recent examples of interim-year studies used to address immediate concerns include the 1995 NRI Erosion Update Study (Fuller *et al.* 1995) and a 1996 NRI special study to investigate the effect of the Federal Agricultural Improvement and Reform Act of 1996 and high commodity prices on land use and conservation practices. The 1995 study provided interim estimates of changes in erosion as a function of conservation practices and programmes.

When designing special studies, sampling strategies are modified to facilitate sample allocation and to address operational and resource constraints. For example, in the 1995 and 1996 NRI surveys, various types of cropland were of primary interest. Twenty-six states with large cropland acres were selected to be in the sample with certainty, and 13 additional states were selected from the remaining 22 conterminous (non-certainty) states. The non-certainty states were stratified by region and the sample states were selected randomly with probability proportional to a measure of the state's cropland acres. The availability of operational resources within states played a role in sample allocation across states. To facilitate efficient field operation, counties were selected within sample states. A subsample of segments (called PSUs in the national sample) was then selected within each county with higher selection probabilities assigned to segments containing highly erodible cropland. Since cropland classifications for points were based on information from previous studies, points classified as non-cropland in prior years were given a small positive probability of selection. Thus, points that moved from non-cropland to cropland were included in the sample, and the study population was the entire cropland domain for the study year.

Designs in which subsamples are selected from the national sample can be viewed as multiphase designs. The national sample plays the role of the phase I sample for which less intensive observations are made on numerous points. The subsample plays the role of the phase II sample, on which more expensive measurements are made. For NRI, the phase I measurements include land use classifications and other pertinent characteristics collected during a previous survey. Phase II measurements involve field observations or extensive photo-interpretation of the same characteristics and several new critical variables for the targeted study year.

The NRI sample framework also provides for intensive studies in small regions. In the NRI programme, augmentation has historically been accomplished using systems developed when the original frame and sample materials were constructed. The original sample materials were developed with multiple preselected samples of varying rates (e.g. 1%, 2%, 4%). The preselected PSUs not included in the basic NRI sample can be used to augment the original sample in specific areas to address new objectives. These preselected samples have been used in states such as Iowa and Kansas to create samples designed for county-level inference. In other cases, increased sample sizes have been used for special studies of a specific natural resource or geographic area. More recently, samples in areas not previously sampled have been selected using Geographic Information Systems (GIS), where latitude and longitude are used to define strata within regions (Breidt, 1995).

The temporal aspects of monitoring studies add numerous complications to study design. Objectives for such studies usually include estimating status at a particular time point as well as estimating trends over time. Longitudinal aspects of the NRI are based on repeated observations on sample PSUs and points. Repeated observations on sample points also make it possible to study the dynamics of land use as the point classification changes over time (Nusser *et al.* submitted). While area measurements from segments can be used to estimate net changes in acres associated with specific land uses, dynamics associated with changes from one type of land use to another are difficult to assess without point-specific data. When location links are available for sample points, spatial dynamics can also be studied.

As data collection methods change or sample units are discarded, designs can be implemented to measure the effects of the change in observation protocols or sample composition. For example, sample contamination may occur when land owners become aware that their land is included in a sample unit and subsequently alter their practices. In such cases, the sample unit may need to be replaced. It is useful to collect data on both the old and new units during an overlap period to estimate the effect of altering the sample. Such a procedure was implemented when new sample units were selected in 13 northeastern states during the 1982 NRI. Alternatively, new measurement procedures or definitions will arise that provide improved observations. New definitions and protocols should be incorporated into the study in a manner that allows the effect of such changes to be estimated. This usually takes the form of using the old and new protocols on the same sample units for a subset of the larger sample. These paired observations are then used to adjust historical estimates so that they are consistent with current measurements.

Finally, while much attention is given to the selection of sampling units in natural resource studies, sample design for measurements at a particular location also requires statistically-sound protocols to reduce the influence of subjectivity when collecting data. Well-defined protocols and sample design for measurements are necessary to insure repeatability across time, space and field observers. We must also recognize that in many cases variables are based on imprecise measurements or that a well-measured variable estimates a conceptual quantity with error. When measurement error is an important component, as can be the case with photo-interpretation or field measurement, sampling strategies can be devised to provide more precise estimates and to estimate the variance of the measurement error (see e.g. Francisco, 1986).

4.2 *Data collection*

While most of the procedures used to collect NRI data have remained stable over the past decade, some methods are being adapted to take advantage of new technologies. For example, the use of digital geospatial data layers in obtaining county base control totals is currently being investigated. The Census Bureau produces layers containing political boundaries and spatial information for county base categories such as water bodies. A new MLRA data layer is also being produced by the NRCS. These layers are being examined to see if they are reasonably consistent with previous figures and to determine whether complete coverage of sufficiently high quality exists for use in the NRI.

New computer-assisted survey information collection methods are also being explored. A system has recently been developed in which the computer-assisted survey instrument is executed on a hand-held, pen-based computer with the ability to exchange survey information with a central database server via modem (Nusser *et al.* 1996). This system is being used as part of the 1996 NRI data collection effort.

4.3 *Statistical methods for constructing databases*

While sampling considerations for national monitoring programmes have recently received attention from the statistical community, there is much to be learned about the development of large longitudinal data bases (see Kasprzyk *et al.* 1989). Such databases should be complete, simple to use, rapidly produced, and consistent with known information and historical estimates. In the early stages of a natural resource survey, estimation of changes is relatively simple due to the limited number of observations over time. As more time points are added, the complications multiply as the amount of information on a sample unit increases and as the effect of changes in objectives, sample composition, and measurements accrue. Sophisticated statistical procedures must be modified to address operational challenges associated with processing large volumes of data in a timely fashion.

The 1992 NRI estimation procedures were developed to accomplish a complex set of goals. The procedures involved many operations, including missing data imputation, pseudo point generation, smoothing of small area totals, and weight calculation. Many of the procedures developed during recent years have become necessary because of the longitudinal nature of the study and the increasing number and complexity of restrictions. Because these procedures were applied to a large volume of data with widely varying characteristics, it was desirable to construct relatively simple procedures that incorporated a complicated set of controls and restrictions in a manner that minimized decision-making external to the computer programs. Standard estimation techniques have been modified to facilitate extensive error checking in the estimation process.

Hot-deck procedures (Little and Rubin, 1987) have been widely used to impute missing data because the imputed values are drawn from a distribution that is presumed to match that of the missing data. The hierarchical hot-deck imputation methods used to impute historical data and pseudo points provided a practical solution to problems that arose during the 1992 NRI. A significant operational advantage was gained by developing a procedure that relied on a single imputation and that could be applied to several settings. In addition to imputing data containing variation similar to that of related observations, the method also preserved a number of the relationships among variables. Though the construction of the imputation classes and steps in the hierarchical collapse of the imputation classes required considerable care, it was simpler than constructing intricate models of the relationship among the missing and observed variables. A disadvantage of the imputation approach is that confidence interval coverage is overstated when imputed data are analysed as if the imputed data were observed.

The imputation procedure developed to generate pseudo points was used in lieu of two-phase estimation. In classical two-phase estimation, regression is used to incorporate PSU information into the weight for the secondary sampling unit. Two-phase estimates for small areas can be more variable due to the absence of points representing known land use changes in the PSU. For example, the lack of points following observed patterns results in many zero estimates of change for certain land use categories even though the actual PSU data may indicate that a change in land use exists. Breidt *et al.* (submitted) have shown that the imputation procedure is approximately unbiased and provides more precise estimates of change, for example, in acres in vegetative cover classes, than standard two-phase estimation.

4.4 *Analysis*

The form of the database and the associated software were developed in response to user needs for accessibility. While point level databases have been produced in association with inventory studies for some time, the NRCS Data Analysis Software (USDA, 1994b) is a new feature of the inventory, designed to allow 1992 NRI data to be used by a broader audience. It is expected that this software will continue to be developed in a manner that takes advantage of Internet access to a central site.

Variance estimates provided by the NRCS Data Analysis Software or other standard survey packages are approximate because they ignore the effect of controls and imputation. However, such variance calculations provide useful indications of the variability of the data. Research is being conducted to develop improved variance estimators (Breidt *et al.* submitted). Replication methods are being explored as a method of estimating variances that includes sources of variation due to imputation and recognizes the use of external controls.

4.5 *Integrated inventory programmes*

The NRI and other existing resource inventories do not currently provide all the information needed by scientists and policy makers to understand and to make informed decisions regarding the complex interactions between and within ecosystems. The Committee on the Environment and Natural Resources of the National Science and Technology Council has been assessing federal, state, and local efforts to monitor the environment. It is proposing a framework that provides for integration and coordination of a broad set of activities, including remote sensing, probability-based surveys, research into ecological modelling, long-term monitoring at purposively selected sites, and research station activities. Integration of probability-based monitoring surveys has also been proposed as a result of an inter-agency demonstration project conducted in Northern Oregon during 1995. This study investigated the feasibility of combining sampling frames and ecological measures from different disciplines and agencies. The project team included representatives from the National Agricultural Statistical Service (NASS), US Forest Service (USFS), Bureau of Land Management, National Biological Survey, Environmental Protection Agency, and NRCS. The sample was selected from a frame containing sample units from the NRI, the USFS's Forest Inventory and Analysis program, and the National Forest System's inventory system. Data were collected by inter-agency teams in two phases, using photo-interpretation and using intensive on-site data collection. Each field site was visited independently by two different field crews to evaluate the repeatability of measurement protocols (Goebel *et al.* 1996). Several facets of coordinated inventory activities for public lands are discussed by Max *et al.* (1996).

An early example of successful integration of programmes is the Alaska NRI sample, which was designed to include existing USFS plots (Fuller and Goebel, 1993). Collaborative efforts between NRCS, NASS, and the Economic Research Service have also occurred for the Area Studies in the early 1990s and more recently as part of the 1996 NRI. Farm operator data have been collected by NASS as part of the 1996 NRI special study.

Traditional modes of inter-agency cooperation for the NRI involved the use of statistical procedures to incorporate information and classification procedures developed by other government programmes. It is increasingly important to develop common definitions and measurement protocols, to collect data that enables *post hoc* classification according to individual agency definitions, and to link current data with historical definitions. These activities are most effective when inter-agency coordination and planning is accomplished as part of the preparation phase prior to data collection.

4.6 *Summary*

NRI plays an important role in supporting policy development for the nation's natural resources. Its longevity is partly a function of the pivotal role agricultural production plays in our society. In addition, the NRI programme has long benefited from the support of administrators who understand the importance of obtaining scientific information from well-designed studies as a basis for policy development.

NRI's success in fulfilling this role is also partly attributable to its strong statistical and scientific foundations. Theoretical procedures have been adapted to accommodate operational constraints in a statistically sound manner. Emphasis has been placed on simplicity whenever scientific and statistical validity permits. A stratified two-stage random sample and design-based analysis strategy have proved to be responsive to new political and scientific initiatives, and have facilitated rapid release of results. The resulting product is a powerful longitudinal multivariate database, suitable for investigating spatial and temporal dynamics of agricultural and ecological variables.

Acknowledgements

This work has been supported in part by cooperative agreement 68–3A75–43 between the USDA Natural Resources Conservation Service and Iowa State University. We gratefully acknowledge the research and editorial contributions of W.A. Fuller, F.J. Breidt, and J. Opsomer, and the support of our typist, J.A. Shafer. We also thank the reviewers for providing many useful comments on an earlier draft of this paper.

References

Baker, H.D., Fuller, W.A. and Goebel, J.J. (1992) *Sample design and estimation procedures for the 1987 National Resources Inventory*. Statistical Laboratory, Iowa State University.

Birdsey, R.A., Hahn, J.T., MacLean, C.D., Arner, S.L., Bechtold, W.A., Williams, M.S., Schreuder, H.T., Scott, C.T., Moisen, G.G., Stage, A.R. and Born, J.D. (1995) *Techniques for forest surveys when cluster plots straddle two or more conditions*. Forest Science Monograph 31.

Breidt, F.J. (1995) Markov chain designs for one-per-stratum sampling. *Survey Methodology* **21**, 63–70.

Breidt, F.J., McVey, A.M. and Fuller, W.A. (1996) Two-phase estimation by imputation. *Journal of the Indian Society of Agricultural Statistics* (Golden Jubilee number) **49**, 79–90.

Cassell, D.L. (1993) Inclusion probabilities for environmental monitoring in FHM and EMAP. *1992 Proceedings of the Section on Statistics and Environment*, American Statistical Association, 82–5.

Francisco, C.A. (1986) *A quality evaluation study of the 1982 National Resources Inventory*. Statistical Laboratory, Iowa State University.

Fuller, W.A. (1990) Analysis of repeated surveys. *Survey Methodology* **16**, 167–80.

Fuller, W.A. (1996) Estimation for the 1992 NRI. Unpublished manuscript.

Fuller, W.A. and Goebel, J.J. (1993) Sample design for the Alaska NRI. Unpublished manuscript.

Fuller, W.A., Breidt, F.J. and Huang, H.C. (1995) The 1995 Erosion Study sample. Unpublished manuscript.

Fuller, W.A., Kennedy, W.J., Schnell, D., Sullivan, G. and Park, H.J. (1986) PC CARP. Statistical Laboratory, Iowa State University.

Goebel, J.J. and Baker, H.D. (1987) The 1982 National Resources Inventory sample design and estimation procedures. Statistical Laboratory, Iowa State University.

Goebel, J.J. and George, T.A. (1996) Establishing a Consistent National Base for Assessing Natural Resource Issues. In *Evaluating Natural Resource Use in Agriculture: Analytical Tools Now and into the Future*, edited by T. Robertson, B.C. English and R.R. Alexander, University of Tennessee Institute of Agriculture, 319–40.

Goebel, J.J. and Schmude, K.O. (1982) Quality control and evaluation for the SCS National Resources Inventories. In *In-Place Resource Inventories: Principles and Practice*. Society of American Foresters, Publication No. 82–02, 871–6.

Goebel, J.J., Reiser, M. and Hickman, R.D. (1985) *Sampling and estimation in the 1982 National Resources Inventory*. Paper presented at the 145th annual meeting of the American Statistical Association, Las Vegas, Nevada.

Goebel, J.J., Schreuder, H.T., Geissler, P.H., House, C.C., Olsen, A.R. and Williams, W. (1996) *Reinventing the Nation's Environmental Efforts: A Proposal to Merge Federal Surveys*. Paper presented at American Statistical Association Workshop on Statistical Issues for Ecological and Natural Resource Monitoring Programs in the US, Washington, DC.

Helsel, D.R. (1995) Design of a relational water-quality assessment program. *1995 Proceedings of the Biometrics Section*, American Statistical Association, 60–7.

Kasprzyk, D., Duncan, D., Kalton, G. and Singh, M.P. (1989) *Panel Surveys*. Wiley, New York.

Kellogg, R.L., TeSelle, G.W. and Goebel, J.J. (1994) Highlights from the 1992 National Resources Inventory. *Journal of Soil and Water Conservation*, **49**, 521–7.

Little, R.J. and Rubin, D.B. (1987) *Statistical analysis with missing data*. Wiley, New York.

Max, T.A., Schreuder, H.T., Hazard, J.W., Oswald, D.D., Teply, J. and Alegria, J. (1996) *The Pacific Northwest Region Vegetation and Inventory Monitoring System*. Research Paper PNW-RP-493, US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Messer, J.J., Linthurst, R.A. and Overton, W.S. (1991) An EPA program for monitoring ecological status and trends. *Environmental Monitoring and Assessment* **17**, 67–78.

McCormack, R.M. (1988) Statistical challenges in the environmental sciences: a personal view. *Journal of the Royal Statistical Society, Series A* **151**, 201–10.

McVey, A.M., Breidt, F.J. and Fuller, W.A. (1994) Two-phase estimation through imputation. *1994 Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1053–8.

Nusser, S.M. (1995) Sampling issues in regional monitoring programs. *1995 Proceedings of the Biometrics Section*, American Statistical Association, 68–70.

Nusser, S.M., Breidt, F.J. and Fuller, W.A. Design and estimation for investigating the dynamics of natural resources. Submitted to the *Journal of Ecological Applications*.

Nusser, S.M., Thompson, D.M. and DeLozier, G.S. (1996) Using personal digital assistants to collect survey data. *1996 Proceedings of the Section on Survey Research Methods*, American Statistical Association (forthcoming).

Overton, W.S. and Stehman, S.V. (1993) Properties of designs for sampling continuous spatial resources from a triangular grid. *Communications in Statistics – Theory and Methods* **22**, 2641–60.

Overton, W.S. and Stehman, S.V. (1994) Variance estimation in the EMAP strategy for sampling discrete ecological resources. *Environmental and Ecological Statistics* **1**, 133–52.

Overton, W.S. and Stehman, S.V. (1995) Desirable design characteristics for long-term monitoring of ecological variables. *1995 Proceedings of the Biometrics Section*, American Statistical Association, 50–9.

Peck, S.L., Rawlings, J.O. and Finkner, A.L. (1992) A comparison of sampling design options for EMAP – Agro ecosystems. *1991 Proceedings of the Section on Survey Research Methods*, American Statistical Association, 191–5.

Platek, R., Rao, J.N.K., Särndal, C.E. and Singh, M.P. (1987) *Small Area Statistics*. Wiley, New York.
Powell, D.S., McWilliams, W.H. and Birdsey, R.A. (1994) History, change and the US forest inventory. *Journal of Forestry* **92**, 6–11.
Saeboe, H.V. (1983) Land use and environmental statistics obtained by point sampling. *Proceedings of the Session of the International Statistical Institute* **50**, 1317–42.
Schreuder, H.T. and Czaplewski, R.L. (1993) Long-term strategy for the statistical design of a forest health monitoring system. *Environmental Monitoring and Assessment* **27**, 82–94.
Schreuder, H.T., Gregoire, T.G. and Wood, G.B. (1993) *Sampling Methods for Multi-resource Forest Inventory*. Wiley, New York.
Stehman, S.V. and Overton, W.S. (1992) Properties of designs for sampling continuous spatial resources. *1991 Proceedings of the Section on Statistics and the Environment*, American Statistical Association, 182–7.
Stehman, S.V. and Overton, W.S. (1994a) Environmental Sampling and Monitoring. In *Handbook of Statistics*, Vol. 12, G.P. Patil and C.R. Rao (eds) Elsevier Science Publishers, Amsterdam, 263–306.
Stehman, S.V. and Overton, W.S. (1994b) Comparison of variance estimators of the Horvitz-Thompson estimator for randomized variable probability systematic sampling. *Journal of the American Statistical Association* **89**, 30–43.
Stevens, D.L. and Olsen, A.R. (1992) Statistical issues in environmental monitoring and assessment. *1991 Proceedings of the Section on Statistics and the Environment*, American Statistical Association, 76–85.
Tollefson, M.H. (1992) Variance estimation under random imputation. Ph.D. Thesis. Iowa State University.
Tollefson, M.H. (1994) Imputation programs for the 1992 NRI. Unpublished manuscript.
US Department of Agriculture (1945) *Soil and water conservation needs estimates for the United States by states*. Soil Conservation Service, Washington, DC.
US Department of Agriculture (1962) Basic statistics of the National Inventory of Soil and Water Conservation Needs. *Statistical Bulletin No. 317*, Washington, DC.
US Department of Agriculture (1981) Land resource regions and major land resource areas of the United States. *Agricultural Handbook 296*, Soil Conservation Service, Washington, DC.
US Department of Agriculture (1991) *Instructions for collecting 1992 National Resources Inventory sample data*. Soil Conservation Service, Washington, DC.
US Department of Agriculture (1994a) Summary Report: 1992 National Resources Inventory. Soil Conservation Service, Washington, DC., and Statistical Laboratory, Iowa State University.
US Department of Agriculture (1994b) *National Resources Inventory Data Analysis Software: User's Guide, Version 1*. Soil Conservation Service, Washington, DC.
US Geological Survey (1982) A US Geological Survey standard: Codes for the identification of hydrologic units in the United States and the Caribbean outlying areas. *Geological Survey Circular 878-A*.

Biographical sketches

S.M. Nusser is Associate Professor in the Department of Statistics, Professor-in-charge of the Statistical Laboratory's Survey Section at Iowa State University, and a member of the faculty of the Ecology and Evolutionary Biology programme. She holds a MS degree in Botany and a PhD in Statistics. She has collaborated with the US Department of Agriculture on the National Resources Inventory and on statistical methods for soil surveys. Her research interests include the use of statistics in biological and ecological studies, survey statistics, and statistical methods for dietary assessment.

J.J. Goebel is National Leader for the Resources Inventory Program within the Natural Resources Conservation Service, US Department of Agriculture. He served previously as National Statistician for the Soil Conservation Service, and was formerly on the faculty of the Department of Statistics at Iowa State University. He holds a PhD in Statistics, and his research interests include survey methodology, time series, and geospatial analysis.