United States
Environmental Protection
Agency

Office of Acid Deposition
Environmental Monitoring and
Quality Assurance
Washington DC 20460

EPA/600/8-90/055
July 1990

Research and Development

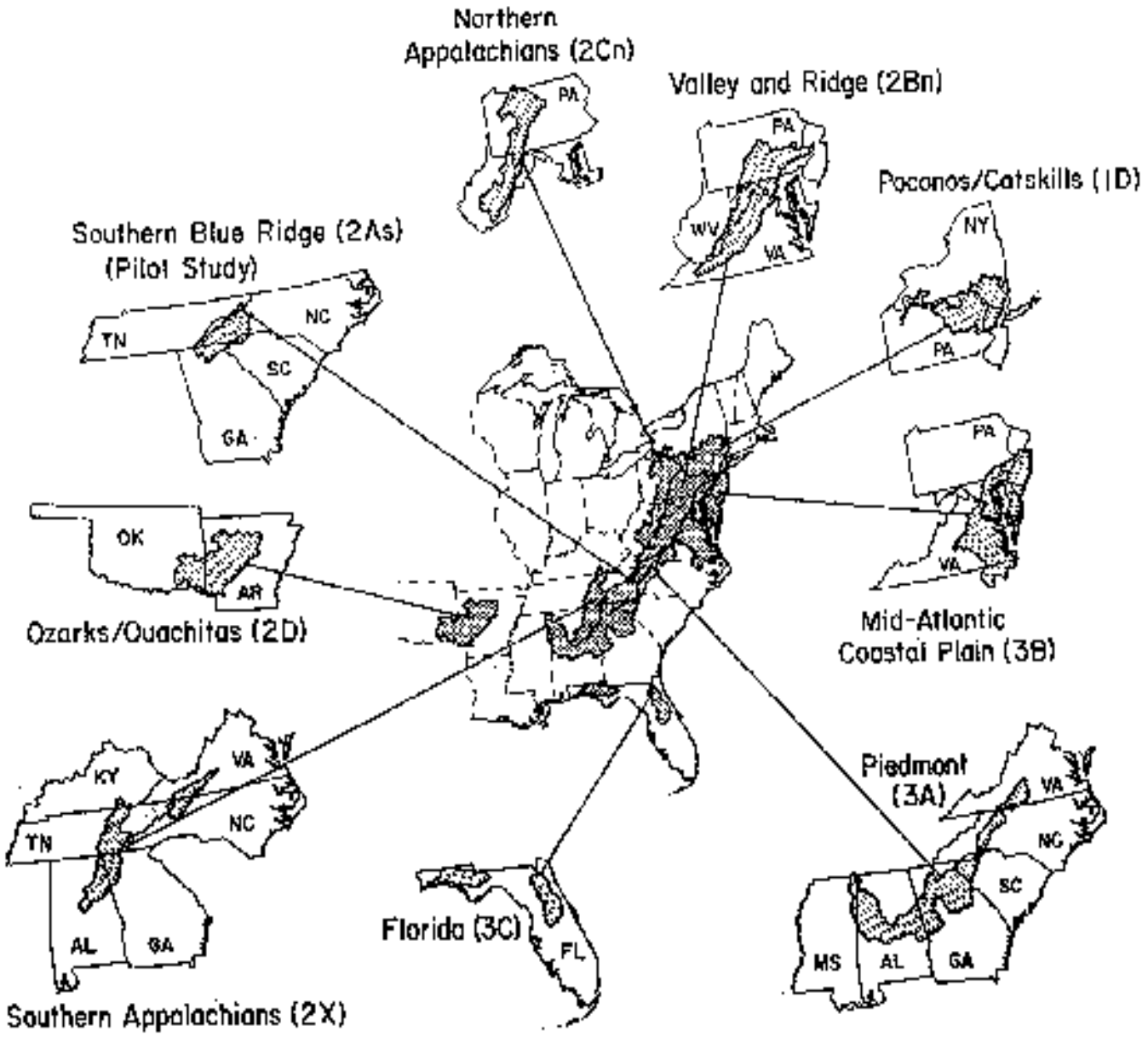
EPA

National Stream Survey Database Guide



Aquatic Effects Research Program

SUBREGIONS OF THE NATIONAL STREAM SURVEY - PHASE I



EPA/600/8-90/055

NATIONAL STREAM SURVEY DATABASE GUIDE

by

Mark E. Mitch

Philip R. Kaufmann

Alan T. Herlihy

W. Scott Overton

Michael J. Sale

A Contribution to the
National Acid Precipitation Assessment Program



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

Environmental Research Laboratory - Corvallis, OR 07333
Environmental Monitoring Systems Laboratory - Las Vegas, NV
80193

NOTICE

The research described in this document has been funded wholly or in part by the U.S. Environmental Protection Agency under Contract No.68-03-3249 to Lockheed Engineering and Management Services Company, Inc.; Contract Nos. 68-02-3889 and ~02-3994 to Radian Corporation; Contract Nos. 68-03-3246 and 68-C8-0006 to NSI Technology Services$_1$ Inc.; Interagency Agreement Nos. #0W89931368 and #1824-1557-A1 with the U.S. Department of Energy (Contract No. DE-ACO5-84OR21400 with Martin Marietta Energy Systems, Inc.); and Cooperative Agreements with Utah State University (CR815168) and Oregon State University (CR813061). It has been subject to the Agency's peer and administrative review, and has been approved for publication as an EPA document. Mention of corporation names, trade names, or commercial products does not constitute endorsement or recommendation for use.

The correct citation for this document is:

Mitch, M.E., P.R. Kaufmann, AT. Herlihy, W.S. Overton, and M.J. Sale. 1990. National Stream Survey Database Guide. EPN6OO/8-90/055. U.S. EPA Environmental Research Laboratory, Corvallis, Oregon. 92 pp.

Inquires regarding the National Stream Survey - Phase I Mid-Atlantic and Southeast should be directed in writing to:

Chief, Watershed Branch
U.S. Environmental Protection Agency
Environmental Research Laboratory
200 SW 35th Street
Corvallis, Oregon 97333

RELATED DOCUMENTS

Supplemental information on the National Stream Survey - Phase I (NSS-l) can be found in a series of ancillary manuals and reports. These publications include:

A Sampling Plan for Streams in the National Surface Water Survey. 1985. Technical Report 114 (July 1986). Overton, W.S. Department of Statistics, Oregon State University, Corvallis, Oregon 97331. 18 pp.

Draft Research Plan, National Surface Water Survey: National Stream Survey, Mid-Atlantic Phase I and Southeast Screening. 1985. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. 134 pp.

Eastern Lake Survey - Phase II, National Stream Survey - Phase I: Processing Laboratory Report - 1988. Arent, U., M.D. Morison, and C.S. Soong. EPA/6OO/4-88/025. U.S. Environmental Protection Agency, Washington, D.C. 86 pp.

National Surface Water Survey: National Stream Survey, Phase I - Pilot survey. 1986. Messer, J.J., C.W. Ariss, J.R. Baker, S.K. Drousé, K.N. Eshleman, P.R. Kaufmann, R.A. Linthurst, J.M. Omernik, W.S. Overton, M.J. Sale, R.D. Schonbrod, S.M. Stambaugh, and J.R. Tuschall, Jr. EPA/6OO/4-86/026. U.S. Environmental Protection Agency, Washington, D.C. 179 pp.

Quality Assurance Plan for the National Surface Water Survey, Stream Survey (Middle Atlantic Phase I, Southeast Screening, and Middle Atlantic Episodes Pilot). 1986. Drousé, S.K., D.C. Hillman, L.W. Creelman, and S.J. Simon. EPA/6OO/4-86/044. U.S. Environmental Protection Agency, Washington, D.C. 215 pp.

Field Operations Report, National Surface Water Survey, National Stream Survey, Pilot Survey. 1987. Knapp, C.H., C.L Mayer, D.V. Peck, J.R. Baker, and G.J. Filbin. EPA/6OO/8-87/01 9. U.S. Environmental Protection Agency, Washington, D.C. 110 pp.

Evaluation of Quality Assurance and Quality Control Sample Data for the National Stream Survey (Phase I - Pilot Survey). 1987. Drousé, S.K EPA/6OO/8-87/057. U.S. Environmental Protection Agency, Washington, D.C. 45 pp.

Analytical Methods Manual for the National Surface Water Survey, Stream Survey (Middle Atlantic Phase I, Southeast Screening, and Middle Atlantic Episodes Pilot). 1987.

Hillman, D.C., S.H. Pia, and S.J. Simon. EPA/6OO/8-87/005. U.S. Environmental Protection Agency, Washington, D.C. 265 pp.

A Sampling and Analysis Plan for Streams in the National Surface Water Survey. 1987. Technical Report 117. Overton, W.S. Department of Statistics, Oregon State University, Corvallis, Oregon 97331. 50 pp.

Data Management and Analysis Procedures for the National Stream Survey. 1990. Sale, M.J. (editor). ORNL/TM. Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831. (Draft).

National Surface Water Survey: National Stream Survey (Phase I, Southeast Screening, and Episodes Pilot). Field Operations Report. 1988. Hagley, C.A., C.L. Mayer, and R. Hoenicke. EPA/6OO/4-88/023. U.S. Environmental Protection Agency, Washington, D.C. 36 pp.

National Stream Survey - Phase I Quality Assurance Report. 1988. Cougan, K.A., D.W. Sutton, D.V. Peck, V. Miller, J.E. Pollard, and J. Teborg. EPA/6OO/4-88/01 8. U.S. Environmental Protection Agency, Washington, D.C. 205 pp.

Chemical Characteristics of Streams in the Mid-Atlantic and Southeastern United States. Volume I: Population Descriptions and Physico-Chemical Relationships. 1988. Kaufmann, P.R., A.T. Herlihy, J.W. Elwood, M.E. Mitch, W.S. Overton, M.J. Sale, J.J. Messer, K.A. Cougan, D.V. Peck, KH. Reckhow, A.J. Kinney, S.J. Christie, D.D. Brown, C.A. Hagley, and H.l. Jager. EPA/6OO/3-88/021a. U.S. Environmental Protection Agency, Washington, D.C. 397 pp.

Chemical Characteristics of Streams in the Mid-Atlantic and Southeastern United States. Volume II: Streams Sampled, Descriptive Statistics, and Compendium of Physical and Chemical Data. 1988. Sale, M.J., P.R. Kaufmann, H.I. Jager, J.M. Coe, K.A. Cougan, A.J. Kinney, M.E. Mitch, and W.S., Overton. EPA/6OO/3-88/021b. U.S. Environmental Protection Agency, Washington, D.C. 595 pp.

TABLE OF CONTENTS

Section Page

Noticeii
Related Documents . . . iii
Acknowledgements . . . viii
1.INTRODUCTION . . . 1
1.1 Purpose and Scope . . . 1
1.2 NSS-I Database Distribution Notes . . . 2
1.2.1 Description of NSS-I and Pilot Survey Data Sets . . . 2
1.2.1.1 NSS-I Report Data . . . 2
1.2.1.2 Data Set NSSIDS3 . . . 4
1.2.1.3 Data Set NSSIDS4 . . . 4
1.2.1.4 Data Set SBRYN . . . 5
1.2.1.5 Data Set NSSFSO . . . 5
1.2.1.6 NSS-I Pilot Report Data . . . 5
1.2.1.7 Data Set PILOTDS3 . . . 6
1.2.1.8 DataSetPILOTDS4 . . . 6
1.2.2 Transfer Media . . . 7
1.3 Notice of Caution and Issues of Interest to NSS Data Users . . . 7
1.4 Episode Pilot Survey . . . 8
2. SURVEY DESIGN . . . 9
2.1 Objectives . . . 9
2.2 Target Population . . . 9
2.3 Sample Reach Selection . . . 11
2.3.1 Stage I . . . 12
2.3.2 Stage II . . . 12
2.3.3 Effective Sample Size (ESS) . . . 13
2.3.4 Special Interest Sites . . . 13
2.4 Data Collection . . . 15
3. DATABASE DEVELOPMENT . . . 19
3.1 General . . . 19
3.2 Database Evolution and Review . . . 19
3.3 NSS-I Data Qualifiers: Flags . . . 21
3.4 Enhanced Data (Data Set 4) . . . 21
3.5 Selection and Use of Data Sets . . . 22
3.5.1 Data Set 3 Versus Data Set . . . 423
3.5.2 The NSS-I Pilot Survey: The Southern Blue Ridge Province . . . 23
4. DEFINING THE TARGET POPULATION . . . 25
4.1 Evaluation Process . . . 25
4.2 Identifying Noninterest Observations and Sites . . . 25
4.2.1 Episode Identification . . . 26
4.2.2 Acid Mine Drainage . . . 26
4.3 Classification of noninterest sites in the NSS-I Database . . . 26
4.4 Drop Code Values . . . 28
5. DATABASE APPLICATION . . . 29
5.1 Generating the Database of Index Values From the Reaches Sampled . . . 29
5.2 Extrapolation to the Target Population . . . 29

5.2.1 Sample Weightings for Population Estimates 29
5.2.2 Estimating the Target Population 32
5.2.3 Variance Estimates and Confidence Bounds 33
5.3 Describing the Target Population 34
5.3.1 The Cumulative Distribution Function (CDF) Curve 35
5.3.2 Length Estimates 38
5.3.2.1 First Approach: Length Estimates Based on Node Chemistry 38
5.3.2.2 Second Approach: Length Estimates Based on Interpolated Length 38

ILLUSTRATIONS

Figure

Pa

ge

2-1 NSS-I subregions 10
3-1 The NSS-I database development process 20
5-1 Calculation of an example cumulative distribution function curve for downstream sites 37
5-2 Interpolated length estimates for subregion 2As 39

TABLES

Table Page

1-1 NSS-I Databases Available 3
2-1 NSS-I Grid Point and Effective Sample Size Summary 14
2-2 Variable Measured in the NSS-I 6
5-1 Number of Visits, Total Observations, Target and Nontarget Samples, and Indexed Observations by Subregion 30
5-2 Total Resource Estimates of the NSS Refined Target Population 36

ACKNOWLEDGMENTS

We gratefully acknowledge the many people who helped create the NSS-I and Pilot Survey databases. A project as massive as the NSS could not have been successfully completed without the cooperation of literally hundreds of individuals, including geographers, statisticians, field and laboratory crews, QA staff, data base

managers, and project scientists.

We thank Allison Pollock (SAI), Joe Bernert (E&S Chemistry), Barry Rochelle (NSI), and Deb Chaloud (EMSLV) for their reviews of the manuscript.

We would like especially to thank Susan Christie (NSI) for editing and making suggestions for this document; Jan Coe, Yetta Jager, and Mary Alice Faulkner (SAI), whose data management expertise and programming skills were crucial in bringing the database into existence; Stasia Allen, Doug Brown, and Jeff Irish (NSI) for their geographic expertise and GIS skills; Karen Cougan, Valerie Sheppe, Chuck Monaco, and Sevda Drousé (LEMSCO) for their fastidious attention to detail and quality assurance; Jim Blick (NSI) and Charles Ariss (Utah State University) for their perseverance through the unending hours of tedium and for their reverence for statistics.

SECTION 1
INTRODUCTION

1.1 PURPOSE AND SCOPE

The National Stream Survey (NSS) is one component of the National Surface Water Survey (NSWS), a project implemented by the United States Environmental Protection Agency (EPA) as part of the Aquatic Effects Research Program (AERP). The AERP, which includes several integrated studies conducted in areas containing surface waters potentially sensitive to change as a result of acidic deposition, addresses four major policy issues relating to the effects of acidic deposition on aquatic ecosystems: (1) the present status and extent of acidic and low alkalinity surface waters in the United States, (2) the extent and magnitude of past change, (3) the change to be expected in the future under various rates of acidic deposition, (4) the maximum rates of deposition below which further change is not expected, and (5) the rate of change or recovery of aquatic ecosystems if deposition rates are decreased.

The data contained on the accompanying tape or floppy diskettes were collected during the first phase of the National Stream Survey (NSS-I), conducted in the mid-Atlantic and southeastern United States (Kaufmann et al., 1988), and include results of a pilot stream survey conducted in the Southern Blue Ridge Province (Messer et al., 1986,1988). Like other components of the U.S EPA'S National Surface Water Survey (NSWS), the NSS is based on a probability sample from an explicitly defined population of surface waters. Data were collected from these surface waters during what was considered to be a representative index period. Sample information was then extrapolated to represent a target stream population within surveyed geographic regions. In the NSS-I, these data were based on samples collected at upstream and downstream locations on stream reaches during the spring of 1986 (1985 for the Pilot Survey).

The flexibility of the NSS design permits the examination of any subpopulation of the total number of sampled streams, based on measured attributes. For example, a subpopulation of interest could be defined as that set of streams located in the state of Maryland or a set of streams with pH values less than or equal to a particular reference value (e.g., $pH \leq 5.5$).

This document serves as a database guide, providing an overview of various aspects of survey design, database structure, and statistical applications of the NSS-I and Pilot Survey databases. It focuses on specific issues that should be kept in mind during analysis and interpretation of the data, such as the criteria used to identify noninterest sites during the database development process.

An additional document, *Data Management and Analysis Procedures for the National Stream Survey* (Sale et al., 1990), details the statistical analysis procedures and computer code listings of programs used to

generate the population estimates presented in Volumes I and II of the NSS-I report (Kaufmann et al., 1988; Sale et al., 1988).

This description of the design, analysis, and application of the data has been extracted and summarized from Volumes I and II of the NSS-I report (Kaufmann et al., 1988; Sale et al., 1988). This document covers the statistical and conceptual design of the surveys (Section 2), the structure and components of the database (Section 3), and the method by which the sampled stream reaches were used to describe the NSS-I target population (Section 4). It also discusses estimation of variance and other statistical issues regarding the computation of regional population estimates (Section 5).

Appendix A discusses specific issues the NSS data user should be aware of. Appendix B contains a data dictionary that lists variable names and their definitions. Appendix C summarizes differences in the estimates reported in the NSS-I Pilot Survey report (Messer et al., 1986) and in the NSS-I final report (Kaufmann et al., 1988). Appendix D lists supplemental variables on which data were collected in the NSS-I.

Table 1-1 lists the NSS databases available for distribution. The enclosed data files are referred to as “Data Set 3" and “Data Set 4". Data Set 3 is a verified and validated data set containing the original data along with observation-specific flags and tags. Data Set 4, a subset of Data Set 3, incorporates the averaging of field duplicate (quality control (QC)J samples with corresponding routine water samples and the replacement of values for missing and/or erroneous data identified in a series of intensive quality assurance (QA) reviews. Data Set 4 (NSSIDS4 and SBRSYN) includes data for both probability sample and special interest reaches and contains unique identifiers that can be used to subset the data and calculate population estimates presented in Kaufmann et al. (1988). Data set 4 retains data for multiple observations (e.g., multiple sampling visits).

1.2 NSS-I DATABASE DISTRIBUTION NOTES

1.2.1 Description of NSS-I and Pilot Survey Data Sets

Six data sets are available for distribution Table 1-1). Four data sets, NSSIDS3, NSSIDS4, NSSIFSO, and SBRSYN, were used to make the NSS-I estimates presented in Kaufmann et a.,

(1988). The two remaining data sets, PILOTDS3 and PILOTDS4, are associated with estimates presented for the NSS-I Pilot Survey report (Messer et al., 1986).

1.2.1.1 NSS-I Report Data--

Data sets NSSIDS3, NSSIDS4, and SBRSYN were used for data analyses and presentations made by Kaufmann et a. (1988) and Sale et a. (1988). These data sets contain data collected during field activities conducted in the spring of 1986 in portions of the mid-Atlantic and southeastern United States, in addition to data from the Southern Blue Ridge Province, which was revised after the original Survey report was published (see Section A.1 1).

TABLE 1-1. NSS-I DATABASES AVAILABLE

	NSS-I Data Set 3 (Validated) NSSIDS3	NSS-I Data Set 4[a] (Enhanced) NSSIDS4	Pilot Survey Data Set 3 (Validated) PILOTDS3	Pilot Survey Data Set 4 (Enhanced) PILOTDS4	Pilot Survey Synthezided Data[b] SBRSYN	Field Site Observations NSSFSO
Number of Observations	1487	1765	397	339	34	1068
Number of Variables	117	94	107	86	61	73
Flags/Tags Present	Yes	No	Yes	No	NO	N/A
Missing/Erroneous Data	Present	Substituted	Presented	Substituted	N/A	N/A
Duplicate Samples	Retained	Averaged	Retained	Averaged	N/A	N/A
Approximate Size	2.0 Mb	1.8 Mb	0.5 Mb	0.3 Mb	0.03 Mb	1.3 Mb
Unique Identifiers	BAT_ID with SAM_ID	STRM_ID with SAMRN	BAT_ID with SAM_ID	STRM_ID with SAMRN	STRM_ID with SAMRN	STRM_ID with DATSMP
NSS Subregion	All	All	2As	2As	2As	All

[a] Data for the NSS Pilot Survey reaches have been appended to NSSIDS4. Data comparable to the full-scale survey's springindex baseflow period are encoded as SAMRN = 1-3 in the Pilot Survey.

[b] All chemistry data contained in this data set were calculated. These values were used to provide estimates of chemistry for 34 upstream sampling sites not sampled during the spring baseflow period in the Pilot Survey. In order to uniquely identify these observations and still incorporate them in to the spring index calculations all 34 of the enhanced observations have a sample visit number (SAMRN) value of 1.5 (see Section 1.2.2.4).

Note: NSS-I population estimates were made using Data Set r, after averaging multiple site visits (see Section 5.1 for detailed instruction).

The NSS Phase I survey was a broader regional application of methods developed during the Pilot survey. In the Mid- Atlantic, 250 steam reaches were visited twice upstream and downstream sampling locations. In the Southeast, 200 steam reaches were visited once upstream and downstream sampling locations. These sample visits are encoded in the database as SAMRN = 1 or 2. Each sample site is identified with a 9-character stream identification number comprised of an 8-digit reach ID and a sample site position (U = upstream; L = downstream). The data from field sampling crews, the processing laboratory, and analytical laboratories have been merged into one file. In contrast, Pilot survey sampling sites are designated by 8-character stream identification codes.

1.2.1.2 Data Set NSSIDS3–

Data set NSSIDS3 contains NSS-l data without any type of enhancement or substitution. The data have undergone an intensive verification and validation process. Data from field duplicate (OC) water samples are included in the data set as separate observations. Values that have been identified as suspect, erroneous, or missing are flagged in the data set but are not replaced. Data qualifiers, flags (see Section 3.3), are included for each chemical variable to denote suspect values, as well as specific conditions or circumstances pertaining to individual water samples (e.g., holding time violations). The parameters are presented in the units and precision in which they were originally measured (see Appendix B, Database Dictionary).

1.2.1.3 Data Set NSSIDS4–

Data set NSSIDS4 is considered the final data set and is the end product of intensive quality review. This data set is used to generate population estimates. Based on chemical relationships within the data, erroneous and missing values have been replaced with estimated values (see Section 3.4). Chemical values from field duplicate (OC) water samples have been averaged with corresponding routine samples. After the original Pilot Stream Survey Report was published, a small portion of the data was revised. The Pilot Survey data, with all revisions, are included in NSSIDS4. The resulting data set contains observations for 450 NSS-l probability sample reaches, *M* Pilot Survey reaches, and 44 special interest streams. Data are <u>not</u> averaged between

sample visits. Note that the reaches sampled during the Pilot Survey have 8-character stream identification codes, whereas reaches sampled during the NSS-I have 9-character codes. Sites considered noninterest in generating estimates of the NSS-I target population of streams are identified by a variable (DRPCDE) that contains a sample exclusion “drop” code: 0-5 (see Section 4.4).

1.2.1.4 Data Set SBRSYN-

In the Pilot Survey, only 20 of the 54 probability reaches were sampled at both upstream and downstream sampling locations during the spring index base flow period. A supplemental data set (SBRSYN) was synthesized to provide spring upstream estimates for the Southern Blue Ridge Province compatible with data from other NSS-I subregions. Data for 22 chemical variables for the 34 upstream sites that were not sampled were synthesized, based on empirical relationships with data from the sampled streams and on data collected at all sites during the summer flow period (see Kaufmann et al., 1988, Appendix B). These observations have been assigned a sample visit number (SAM RN) of 1.5 to allow the user to identify these sites and merge their data with other Southern Blue Ridge data that have sample visit numbers 0 - 4. The observations that represent the spring index period in the Southern Blue Ridge have sample visit numbers 1 - 3. These observations are identified by a similar 8-character variable, STRM_ID, used for Pilot Survey Southern Blue Ridge sites. In order to calculate population estimates for the Southern Blue Ridge Province that duplicate those in the NSS-I final report, these synthesized values should be appended to the NSSIDS4 data set prior to subsetting and averaging data to make population estimates.

1.2.1.5 Data Set NSSIFSO-

Originally titled “Watershed Disturbance Characteristics”, this data set, NSSIFSO, contains data that describe the watershed in the immediate vicinity of field sampling sites. This information includes details on immediate watershed disturbances, bank vegetation cover, stream substrate, and additional field comments about the sample site. Assessment of substrate type and bank vegetation are made as percentage estimates as: absent (0%), sparse (< 25%), moderate (25-75%), and heavy (> 75%). Data for both the Pilot Survey and the Phase I Survey are combined in one data set. This information is based on unvalidated observations of field crews and is therefore subjective. Vegetation coverage and substrate compositi9n estimates were based on the crew's judgment and are only rough estimates based on visual assessments. This information has not been subjected to a stringent QA review similar to that received by the chemical data. Although specific

observations, coverage estimates, and substrate assessments may be difficult to validate, these data can be a useful tool in examining specific conditions that may have existed at the time of sampling (Appendix D).

1.2.1.6 NSS-l Pilot Report Data--

The Pilot Survey data (PILOTDS3, PILOTDS4) were collected during field activities conducted in the spring and early summer of 1985 in the Southern Blue Ridge Province, Subregion ~s (Messer et al., 1986,1988). Five field sampling visits were made from mid-March to mid-June. The Pilot Survey was conducted to test the logistical
and analytical protocols planned for the full-scale NSS-l in the Mid-Atlantic and Southeast. Data from the Pilot Survey were used to evaluate the statistical sampling design, logistics plan, quality assurance plan, data management program, and data analysis plan.

Up to 5 sampling visits were made to 54 probability sites and an additional 7 special interest sites. The 5 Pilot Survey sampling visits are numbered OA (using the variable SAMRN) in the database. This includes an initial reconnaissance and methods development visit (coded as '0' in the database) that is not used in making population estimates but may be of interest for examining temporal variability. The spring index base flow period is represented by sample visits numbered 1 - 3. The data associated with sampling visit 4 were collected during a summer base flow period, but may be useful for additional analyses. Each Pilot Survey sample site is identified by an 8-character stream reach identification code made up of a 7-character reach ID and a sample site location code (U = upstream; L = downstream).

After the Pilot Survey Report was published, a small portion of the data was revised (see Appendix C). The updated and revised data set has been appended to NSSIDS4. PILOTDS3 and PILOTDS4 are associated with the Pilot Survey Report and do not contain the revised Pilot Survey data included in NSSIDS4.

1.2.1.7 Data Set PILOTDS3--

Data set PILOTDS3 contains NSS-l Pilot Survey data without any type of enhancement or substitution. The data have undergone a rigorous verification and validation review. Chemical values from duplicate water samples are included in the data set as separate observations. Values identified as suspect, erroneous, or missing are flagged in the data set but are not replaced. Data qualifiers,

flags, are included for each chemical variable to denote suspect values, as well as notes about specific conditions or circumstances pertaining to individual water samples (e.g., holding time violations). Parameters are presented in the units in which they were originally measured (Appendix B).

1.2.1.8 Data Set PILOTDS4--

Data set PILOTDS4 is the final data set for the Pilot Survey, and is the end product of intensive quality review. This data set is used to generate population estimates. Based on chemical relationships within the data, erroneous and missing values have been replaced with estimated values. Field duplicate (QC) chemical values have been averaged with corresponding routine samples. Note that all five of the Pilot Survey sample visits are Included In the data set. The spring Index base flow period Is represented by sample visits (SAMRN): 1 - 3.

1.2.2 Transfer Media

The transfer media on which the NSS-l databases are available include either 94rack magnetic tape or 5 1/4 inch high-density floppy disks in a card4mage ASCII format or in a SAS (SAS, 1985) format as a SAS export data set. Missing values are replaced in Data Set 4 for both the Pilot and Phase I databases. In Data Set 3, missing values are represented with the number -999.000 for card image formats. Standard SAS notation for missing values is used in all SAS files (i.e., a “.” for numeric variables and or blank for character variables). The card-image format definitions are discussed in Appendix E.

1.3 NOTICE OF CAUTION AND ISSUES OF INTEREST TO NSS DATA USERS

Use of the NSS data can be very complex. Estimates can be generated for a variety of populations. The user should understand the following specific issues before performing any analyses of NSS data. Details and context of these issues are summarized in Appendix A.

- °Use of Drop Codes: DRPCDE
- °Generating a Working Data Set
- °Data Set of Field Site Observations
- °Use of Sample Weights: W
- °Chemical Variables With Similar Names
- °Population Estimates for Geographic Subsets of Streams
- °Reach Length Estimates: RCH_LN versus L2
- °Topographic Drainage Area Measurements: a_1, a_2, a_3, a_4, a_5, a_{total} (A1, A2, A3, A4, A5 and A_WS)
- °Synthesized Data for the Southern Blue Ridge (Subregion 2AS)
- °Differences Between the NSS-I and the NSS-I Pilot Survey
- °Revisions to NSS-I Pilot Survey Data
- °Comparison of Parameter Units in the NSS-I and the NLS
- °Revisions to a_1
- °NSS Database Variable Formats
- °Subregion Identification Codes
- °Using DIC and pH in Calculated Variables

1.4 EPISODE PILOT SURVEY

Concurrent with the NSS-I field sampling activity, an Episode Pilot Survey was conducted in the Mid-Atlantic Region. The primary objective of the Episode Pilot Survey was to assess the feasibility of a synoptic assessment of precipitation episodes using a probability-based sampling design. Although based on the NSS-I, the Episodes Pilot Survey was not able to collect an adequate number of Episodes samples to make regional estimates of extent, duration, or frequency of episodic conditions in streams. The primary conclusion of investigators was that this type of synoptic survey of episodes would not be cost effective (Messer and Eshleman, 1987). The data collected as part of the Episodes Pilot Survey has not been subjected to the same extensive QA review as the NSS-I samples and is not distributed with the NSS-I data sets.

SECTION 2
SURVEY DESIGN

2.1 OBJECTIVES

The NSS-I was designed to chemically and physically characterize a target population of streams, based on a probability sample. The primary objectives of the NSS-I were:

- To determine the percentage, extent (number, length and drainage area), location, and chemical characteristics of streams in the United States that are presently acidic, or that have a low acid neutralizing capacity (ANC) and thus might become acidic.
- To identify streams representative of important classes in each region that might be selected for more intensive study or long4erm monitoring.

The NSS sampling design employed a randomized, systematic technique for selecting a probability sample of stream reaches within areas of the United States expected to contain waters of low acid neutralizing capacity (ANC) (Messer et al., 1986; Overton, 1986,1987; Kaufmann et al., 1988). Statistically, the stream survey is a double sample, stratified on subregion and expected ANC class (< 50 μeq L^{-1} and > 50 μeq $L^{-1)}$, as represented on maps of expected alkalinity prepared by Omernik and Powers (1983) and Omernik and Kinney (1985).

2.2 TARGET POPULATION

The NSS-I focused on a population of stream reaches defined in terms of size, general water quality, and location. This target population of reaches, in a broad sense, can be thought of as all reaches appearing as blue-lines on 1 :250,000-scale United States Geological Survey (USGS) topographic maps in areas expected to contain surface water ANC $\leq$ 400 μeq L^{-1}. This population was restricted to reaches that drain watersheds < 155 km^2 and that are not grossly polluted. The primary intention of surveying reaches in this size range was to examine those streams large enough to be recreationally and economically important for fish habitat, yet still small enough to be susceptible to change as a result of acidic deposition. These criteria identified a different set of streams in each geographic area (e.g., forested upland sites are noticeably different from those located in the Florida

lowlands). The attributes of large bodies of flowing water (i.e., those with Strahler orders > 5 on 1:250,000-scale USGS maps) were not assessed or described in the NSS-l.

SUBREGIONS OF THE NATIONAL STREAM SURVEY - PHASE I

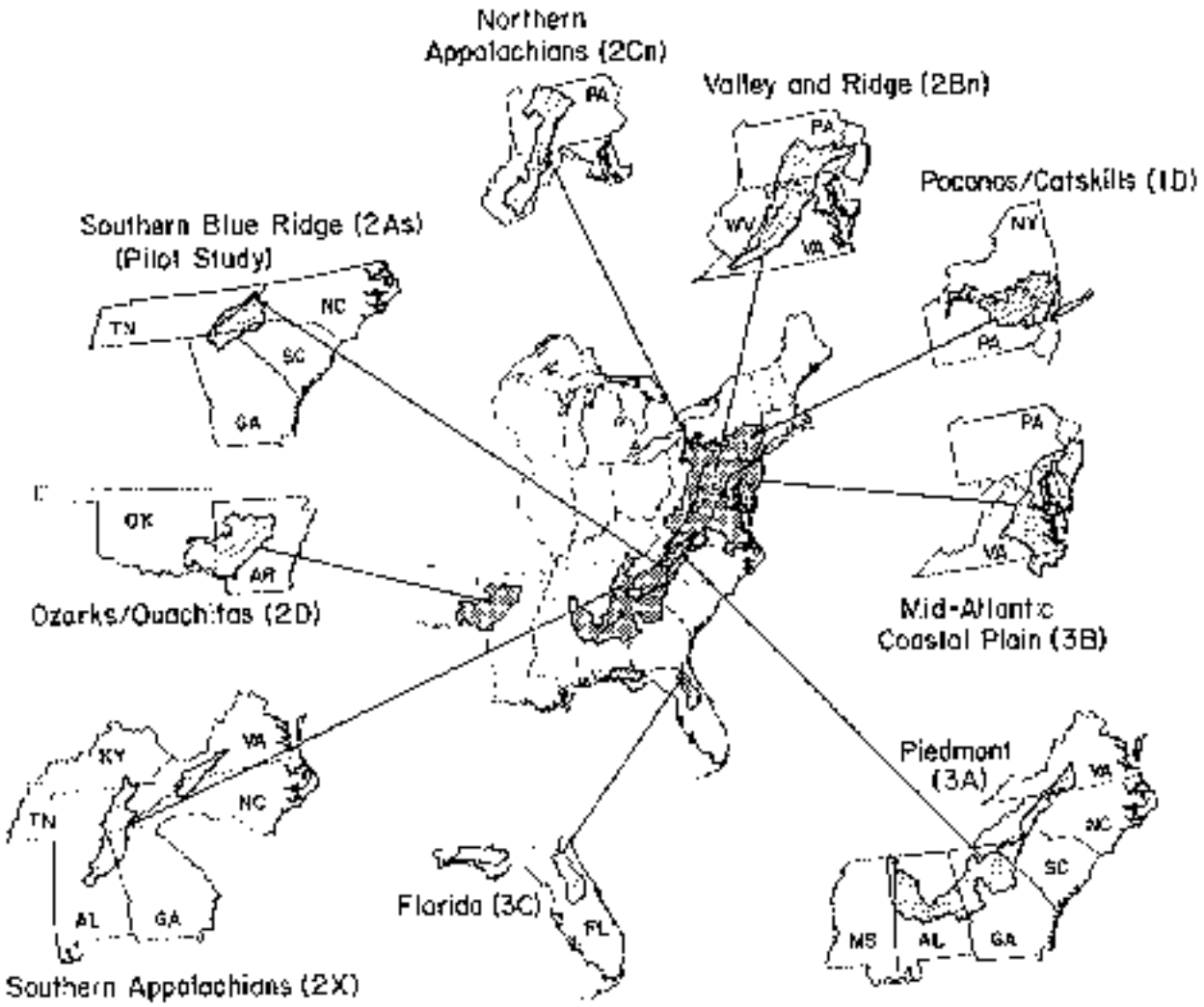


Figura 2-1. NSS-I subregions.

The NSS-I was conducted in two regions of the United States, the Mid-Atlantic and the Southeast, focusing on areas with high sulfate deposition that were considered most likely to contain clear water streams (low in organic acids) with low alkalinity. Nine subregions were identified within these areas (Figure 2-1):

1. Poconos and Catskills (Subregion 1 D).
2. Mid-Atlantic Coastal Plain (3B), including the New Jersey Pine Barrens and the Chesapeake Bay area.
3. Northern portion of the Valley and Ridge Province (2Bn).
4. Northern Appalachians (2Cn).
5. Piedmont (3A).
6. Ozarks and Ouachitas (2D).
7. Southern Blue Ridge (~s), NSS-I Pilot Survey study area sampled in 1985.
8. Southern Appalachians (2X), a combined area of portions of the Southern Appalachian Plateau, the southern area of the Valley and Ridge Province and a northern portion of the Blue Ridge Mountains.
9. Florida (3C).

Section A. 15 details the identification and identification codes of NSS-I subregions in the NSS database.

Initially, the NSS-I was divided into two studies: the full-scale Phase I Survey in the Mid-Atlantic and a screening survey in the Southeast. The principal survey, designed to meet the NSWS primary and secondary objectives, was conducted in the Mid-Atlantic. In contrast, the Southeast Screening Survey was designed to assess the need for additional sampling efforts outside the Mid-Atlantic Region. The difference between these two studies was in the number of water samples collected at each reach end. In the Mid-Atlantic, two water samples were collected from each stream reach end; a single water sample was collected for each reach end in the Southeast.

Results from the NSS-I Pilot Survey in the Southern Blue Ridge indicated that a single spring baseflow sample could adequately represent baseflow conditions and could be subjected to the same assessment and analysis as the multi-sample Mid-Atlantic Survey (Messer et al., 1986). Based on these results, data collected from the Mid-Atlantic and the Southeast are considered to make up the regions sampled in the NSS-I.

2.3 SAMPLE REACH SELECTION

Within the NSS-I, the sampling unit of interest (or description) is the stream reach, defined as

a blue-line segment appearing on 1:250,000-scale maps delimited at each end by a confluence of blue lines or by the end of a blue line segment (I.e., a headwater reach), conforming to criteria that excluded noninterest systems (e.g., watershed area > 155 km^2, urbanized watersheds, reservoirs or lakes). The NSS-I sample reaches were selected using a stratified two-stage variable probability process that selectively identified reaches representing the target population of streams in each subregion. The complete selection process is detailed in Overton (1987).

2.3.1 Stage I

In contrast to the list sampling frame employed by the National Lake Survey, the NSS-I employed an area/point frame. A rectangular dot-grid overlay, made of transparent acetate, that projected 64 mi^2 per point at a scale of 1:250,000 (i.e., the mapped distance between points horizontally and vertically was 8 miles) was overlaid on USGS maps. This process identified the Stage I sample, 3,082 blue-line reaches and 222 nonreaches (lakes, reservoirs, swamps). Of the 3,082 potential sample reaches, 781 were excluded because of drainage area size (> 155 km^2) and proximity to urbanized areas (i.e., more than 20% of the watershed was located within an urban area as indicated on 1 :250,000-scale USGS maps). A reach's probability of inclusion at this stage of sampling was directly related to its direct drainage (a_1) and the density of points portrayed on the acetate grid (i.e., 64 mi^2lpoint) (Overton, 1987). In order to increase resolution in the high-interest portion of the target population, a separate stratum of Stage 110w ANC sites was identified (those areas where ANC was expected to be very low, ANC < 50 μeq L^{-1}).

2.3.2 Stage II

The Stage II sample, a subset of the Stage I sites, was composed of those sites identified for field visits (selected as a variable probability sample). All low ANC Stage I reaches were included in the Stage II sample (except in the Pilot Survey). Thus the Stage II sample included the low ANC sites and a probability subsample from the higher ANC stratum. Because all of the Stage I reaches in the low ANC strata were included in the Stage II sample, these reaches were selected with a Stage II inclusion probability = 1. In addition, some Stage I reaches had such small a_1 values that they also entered into the Stage II sample with an inclusion probability = 1.

Except for reaches in the low ANC strata and reaches with small a_1 values, a reach was selected in Stage II from the Stage I sample with an inclusion probability inversely proportional to its direct drainage, a_1. The final sample weights, used in making population estimates, are inversely proportional to the overall inclusion probabilities. These weights are equal to the number of reaches

each stream reach represents in the target population. The Stage II sampling process selected 504 reaches for field sampling and chemical measurement.

2.3.3 Effective Sample Size (ESS)

A reach's stratum (ANC and subregion) and the effective sample size (ESS) for that stratum are important values used in calculating variance estimates. A reach's ANC stratum can be identified in the database by the variable STRATUM. STRATUM is equal to 1, indicating the higher ANC stratum, to 2, indicating the low ANC stratum, or to 3, indicating a ~small a_1[1] stratum. This breakdown of the number of Stage I and Stage II sites in each stratum is shown in Table 2-1.

Overton (1987) discusses the rationale for calculating the NSS-I ESS. Conceptually, the effective sample size can be thought of as the number of grid points (interest and noninterest combined) that would have to be examined in order to obtain a desired number of sample sites. For example, an effective size of 75 for a subregion indicates that 75 grid points would have to be examined to identify the desired number of Stage II sites (50 reaches per subregion). For Stage I, this value is the total number of grid points falling within the boundaries of the study area. Because the Stage II sample is a subset of the Stage I target sample, the Stage II effective sample size is not explicitly defined. However, an indirect estimate of this value can be made based on the assumption that the ratio of target to nontarget sites in the Stage I sample is applicable to the Stage II sample. For the NSS-I main survey (in contrast to the Pilot Survey[1]), this value was calculated as:

$$n'' = n_2 n'/n$$

where:

- n'' is Stage II effective sample size,
- n^2 is the Stage II sample size (# of reaches selected for field sampling),
- n' is Stage I effective sample size,
- n is Stage I target reaches (# of Stage I reaches considered targets for Stage II sampling).

2.3.4 Special Interest Sites

In addition to the reaches selected for the Stage II sample, the NSS-I sampled 36 special interest reaches. These reaches are similar to those in the NSS-I target population (in terms of size), but are sites where intensive

[1]The Pilot Survey Stage II selection process chose every other Stage I reach from the list of reaches ordered by grid-point locations. This is in contrast to the variable probability selection process used in the main survey. In the Pilot Survey, the ESS was defined as half the number of grid points (84) or the number of points examined to obtain the number of reaches to sample in the field. Because the main survey used a probability sample to obtain the Stage II sites, the ESS was estimated.

TABLE 2-1. NSS-I GRIDPOINT AND EFFECTIVE SAMPLE SIZE SUMMARY

	Stage I Sample						
Subregion	Stratum (1 = ANC > 50 μeq L^{-1}) 2 = low ANC) 3 = small a_1, ANC > 50 μeq L^{-1})	# of Nonreach or Nontarget Points	# of Target Points (n)	Total # of Points (n')	# of Selected Reaches (n^2)	# of Sampled Reaches	Effective Sample size[a] (n")
1D	1 or 3	108	180	288	48	48	76.8
	2	1	13	14	13	13	14.0
2Bn	1 or 3	116	380	496	50	50	65.3
	2	2	3	5	3	3	5.0
2Cn	1or 3	93	206	279	55	55	74.5
	2	7	19	26	19	19	26.0
3B	1 or 3	209	336	545	50	50	81.1
	2	7	12	19	12	12	19.0
2As	1 or 3	32	136	168	54	54	84.04[b]
2D	1 or 3	94	266	360	50	50	67.7
2X	1 or 3	91	243	333	50	48	65.8
3A	1 or 3	144	423	607	50	50	71.7
3C	1 or 3	81	84	265	50	48	94.3

[a] Effective sample size (ESS) is the numer of grid points examined to obtain the needed number of State II sample reaches (to be visited in the field). ESS is stratum specific. Exact vaules for each ANC stratum are necessary when calculating variance.

[b] The effective sample size is calculated differently in the Pilot Survey than in the full NSS-I survey (see Section 2.3.3).

process-oriented studies or long-term monitoring data are available. Although field sampling and chemical analyses were undertaken at these sites, they were not used in making estimates of the target population and their weights are set to zero in the database. These sites, however, are used to examine watershed processes and Iong-term trends. The observations for special interest sites in the NSS database are identified by the following variables:

- SUB_ID = “SI”
- DRPCDE = “5 “
- W = “0”
- The sixth character in RCH_ID and STRM_ID = “9"

2.4 DATA COLLECTION

The collection of water samples and attributes for the Stage II sample constitutes a third stage of sampling in the statistical design of the NSS-l. Sampling at this stage characterizes or indexes the chemical and physical properties of a sample reach. The NSS-l relied on samples taken during an appropriate season from a representative sample of water bodies to provide an index of the chemical characteristics of a target population of surface waters (Messer et al., 1986,1987,1988). In the NSS, this index value depicts stream chemistry during spring baseflow between snowmelt and leafout (approximately March 15 to May 15), when sensitive life stages of important fish species are present and chemical conditions potentially limiting to aquatic organisms are likely to exist. The rationale of the spring index sampling period is detailed in Volume I, Section 2.5 of the NSS-I report (Kaufmann et al., 1988) and in Messer et al. (1988).

Field visits were made during the spring index period to NSS-I reaches in a geographic area of more than 530,000 km^2. The NSS-l water sampling and analysis methodology is detailed in Section 3 of Kaufmann et al. (1988) and in Hillman et al. (1987). On each sample visit, field crews collected a 3.8-L water sample and four 60-mL syringe samples, in addition to recording watershed and hydro--logic descriptive characteristics and making in situ chemical measurements (Hagley et al., 1988). Water samples were then transported to a centralized processing laboratory where they were stabilized. Chemical measurements were made within 36 hours of sample collection (Arent et al., 1988). The processed samples, aliquoted and preserved, were then shipped to contract analytical laboratories for chemical analyses. Table 2-2 lists the physical and chemical data collected in the field, at the processing laboratory, and at the analytical laboratory, as well as the geographic data recorded for the sample sites.

In addition to routine water samples, OC samples were collected and used to ensure that

sampling and analytical methods were performed according to specifications and to evaluate overall data quality for the survey. The types and uses of these samples are detailed in Cougan et al. (1988), Drousé et al. (1986), and Kaufmann et al. (1988).

TABLE 2-2. VARIABLES MEASURED IN THE NSS-I

CHEMICAL AND PHYSICAL VARIABLES (see Appendix B)			
Parameter	Variable Name	Parameter	Variable Name
Field Site/In Situ Measurements			
pH in situ	**PH_R**	Specific conductance	**CONIS**
Temperature	**TMPSTR**	Dissolved oxygen	**DO_IS**
processing Laboratory Measurements			
Monomeric Al	**ALDSVL, ALDS16**	Specific conductance	**CONVAL**
Organic monomeric Al	**ALORVL, ALOR11, ALOR16**	pH, closed system	**PHSTVL**
Dissolved inorganic carbon (closed system)	**DICVAL**	True Color	**COLVAL**
		Turbidity	**TURVAL**
Contract analytical Laboratory [Measurements]			
Acid neutralizing capacity (ANC)	**ALKA11**	Iron	**FE11, FE16**
		Magnesium	**MG16, MG16**
Extractable Al	**ALEX11, ALEX16**	Manganese	**MN11, MN16**
Total Al	**ALTL11, ALTL16**	Nitrate	**NO311, NO316**
Ammonium	**NH411, NH416**	pH; initial acidity	
Base neutralizing capacity (BNC)	**ACCO11**	Titration pH; initial alkalinity	**PHAC11**
Calcium	**CA11, CA16**	Titration	**PHAL11**
Chloride	**CL11, CL16**	pH; air equilibrated	**PHEQ11**
Dissolved inorganic carbon (initial)	**DICI11**	Phosphorus	**PTL11, PTL16 PTD11, PTD16**
Dissolved inorganic carbon (air equil.)	**DICE11**	Potassium	**K11, NA16**
		Silica	**SIO211, SIO2115**
Dissolved organic carbon	**DOC11**	Sodium	**NA11, NA16**
		Specific conductance	**COND11**
Fluoride	**FTL11, FTL16**	Sulfate	**SO11, SO416**

a Multiple variable names exist for those parameters with measurements reported in multiple units (e.g, mg L^{-1} and μeq L^{-1}) or

that have been measured using different methods (ace Appendix B).

TABLE 2-2. VARIABLES MEASURED IN NSS-l (Continued)

Parameter	Variable Name	Parameter	Variable Name
Calculated Varlablea			
Anion deficit	**ANDEF**	Hydroxide conc.	**OH16**
Total anions	**ANSUM**	Organic anion conc.	**ORGION**
Total cations	**CATSUM**	Sum of base cations	**SOBC**
Carbonate concentration	**C0316**	Sum of strong mineral acid anions	**SOS MA**
Bicarbonate	**HC0316**	Inorganic monomeric aluminum	**ALINOR**
Hydronium concentration	**H16**		

GEOGRAPHIC VARIABLES

Geographic Attribute	USGS Map Scale	Variable Name
Watershed area contributing to mapped sampling point (km^2)	1:24,000	**A_WS**
Direct drainage area (mi^2) between upstream and downstream reach ends	1:24,000	**A1**
Drainage area between upstream/downstream sample sites	1:24,000	**A4**
Drainage area above upstream sample site	1:24,000	**A5**
Site elevation (m)	1:24,000	**ELEV**
Stream gradient (%)	1:24,000	**GRADE**
Site latitude (decimal degrees)	1:24,000	**LAT_STD**
Site longitude (decimal degrees)	1:24,000	**LON_STD**
Length of reach between upper and lower sampling sites (km)	1:24,000	**L2**
Length of reach between stream confluences (km)	1:24,000	**RCH_LN**
Name of map(s) showing watershed location	1:24,000	**MAP1..6**
Name of county for reach and watershed	1:24,000	**COUNTY1..4**
Name of 1:250,000-scale map showing watershed	1:250,000	**QUAD**
Number of headwater reaches (Shreve order)	1:250,000	**RCH_HW**
Shreve order	1:24,000	**SHREV75**
ANC stratum (1 or 2) in statistical design	(section 2.4)	**STRATUM**
Strahler order	1:24,000	**STRA75**
Strahler order	1:250,000	**STR_5O**
Reach identification code	N/A	**RCH_ID**

Stream identification code	N/A	**STRM_ID**
State (2-character code)	1:250,000	**STATE_ID**
Stream name	1:24,000	**STRMNAM**
Subregion identification code	N/A	**SUB_ID**

It should be noted that a reach was considered sampled if a visit was made by a field crew to a mapped sampling point. Though specific stream characteristics and/or watershed conditions might have eliminated a reach from the NSS target population, all collected information (even for noninterest conditions) can be used to describe the characteristics of some populations of reaches represented in the Stage II sample.

SECTION 3

DATABASE DEVELOPMENT

3.1 GENERAL

This section provides background information on the database development process used the NSS-l and Pilot Surveys. An important component of all NSWS projects is the approach used in database development to ensure that the collected and recorded data are representative of the physical and chemical characteristics of the water body at the time of sampling. The NSS-l database contains over 56,000 individual values, including physical and chemical parameters as well as quality assurance data, all of which were reviewed individually and in the context of subregion chemistry. Additional details of the database development process and data analysis procedures are presented in Kaufmann et al. (1988) and Sale et al. (1989). As with all NSWS databases, the NSS-l data has undergone an external audit (Pollack and Grosser, 1988) to verify and validate data quality assessments and data set evolution.

3.2 DATABASE EVOLUTION AND REVIEW

The final data sets used in making NSS-I population estimates have been subjected to four levels of QA evaluation. The completion of each level of QA review produced a new working data set of greater refinement (Figure 3-1). These working data sets are defined as: raw (Data Set 1), verified (Data Set 2), validated (Data Set 3), and enhanced (Data Set 4). The final product of this refinement process, the enhanced data set (Data Set 4), incorporates data substitution and replacement of missing values. Data Set 4 was used in calculating the NSS-l population estimates presented in Volume I of the NSS-l report (Kaufmann et al., 1988).

The verification process focused on the internal consistency of chemical measurements within a water sample. This process identified individual chemical values that appeared as exceptions to results predicted from calculations based on chemical relationships (e.g., anion/cation balances, conductance estimates, and protolyte analysis). Concurrent with this process, OC samples (including audit samples, field blank data, and instrument detection limit values) were assessed for potential analytical bias in the laboratory or the field. The final product from the verification process is Data Set 2.

The validation process examined stream chemistry in the context of the total group of sample streams within a subregion Observations that were identified as “atypical' during review of data in a subregional context were considered outliers from the rest of the data. Sites showing a number of unusual chemical values generally were not suspected of having serious analytical errors, but rather were associated with site observations that suggested probable impacts from watershed disturbances, such as acid mine drainage, tidal influence, or urban runoff. Individual chemical outliers were examined for errors that might have occurred during transcription or analysis.

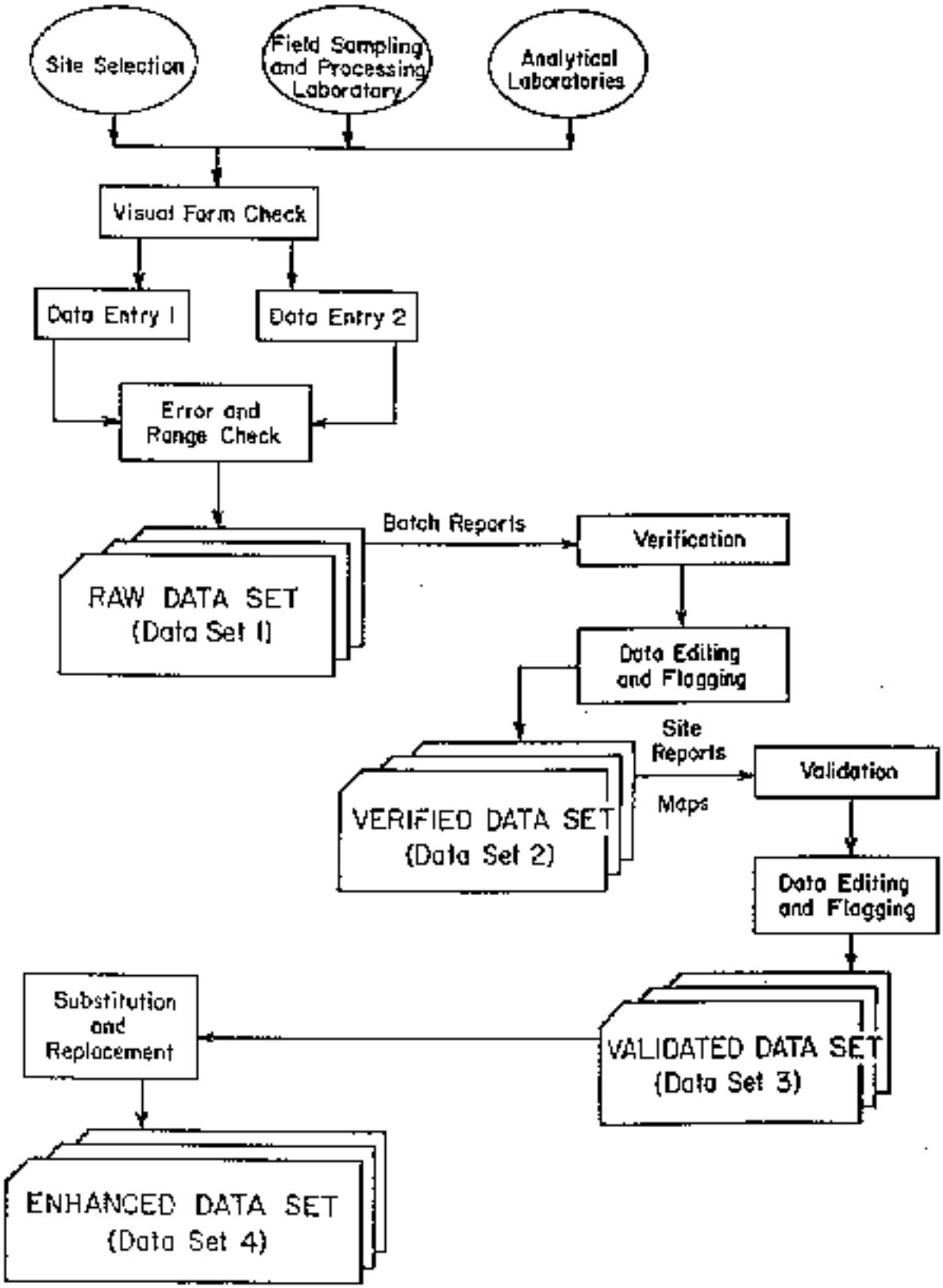


Figure 3-1. The NSS-I database development process.

Erroneous or missing values were considered for possible replacement for calculating population estimates. The final product of the validation process is Data Set 3.

3.3 NSS-I DATA QUALIFIERS: FLAGS

In the development of the NSS-I databases, data qualifiers were used as a tool to mark individual values or even an entire water sample as having particular features (e.g., sample holding time issues, analytical instrument errors, QA discrepancies) that could be pertinent to data interpretation. These qualifiers, described as flag variables, identify observations or notes made during the QA process (verification and validation) employed in creating the enhanced database. Flag variables help identify nonrepresentative or questionable data. The NSS-I flags are included in Data Set 3 but not in Data Set 4 because of the averaging of duplicate samples used in generating Data Set 4 and for ease of use of the data set.

Each chemical variable has a complementary flag variable with a prefix of the chemical variable name and an “F” to indicate a flag. For example, the flag variable for ALKA1 is listed under the variable name ALKA11F. A complete listing of all flag definitions is presented in the NSS-I Data Dictionary (Appendix B, Table B-1). A flag code is composed of two characters, an alpha and a numeric. The first, the alpha character, identifies a problem or concern category (e.g., anion/cation balance discrepancies). The second, the numeric, indicates a specific problem or note within the category. For example, if the flag variable MG11F contained the code “A3", this indicates an anion/cation balance discrepancy with possible cation contamination. Flag variables may also contain multiple codes. If a variable contained the code “A3B0H0", this would indicate three flagged conditions: (1) a problem with the anion/cation balance, (2) a problem with a field blank water sample for the batch of samples, and (3) holding time criteria were not met.

3.4 ENHANCED DATA (DATA SET 4)

Data Set 4, a subset of Data Set 3, is used to generate a data set for calculating population estimates and associated statistics. Such estimates are difficult to generate if there are inconsistencies in the data (e.g., missing values). Data Set 4 was prepared to resolve problems of missing and erroneous data. When necessary, substitutions were perf6rmed according to the following protocol:

1. Whenever possible, values from duplicate (QC) water samples were used.
2. If a duplicate measurement was not available, a value from an alternate visit to the site was used.

I

3. If a duplicate measurement or a measurement from an alternate visit was not available, a substitution value was calculated by means of a linear regression model or ion balance estimate. This predicted value was calculated based on observed chemical relationships and spatial relationships between upstream and downstream sites.

All values generated for substitution were examined for consistency with other data before placing them in the final data set. Only data from routine and duplicate samples were assessed for substitution or enhancement. Of the 44, 975 values, 83 (< 0.5% of the database) were flagged. Only 34 values were replaced (enhanced) in Data Set 4. Of these 34 values, 20 were replaced with data from alternate visits. The remaining 14 were replaced using multiple linear regression and ion balances. In addition to substitutions for erroneous and missing values, negative values for parameters other than ANC and base neutralizing capacity (BN9) were set equal to zero[2]. All modified values in the final data set are flagged.

In roughly 10% of the samples collected, duplicate samples were collected as QC checks to estimate sample precision. The data from these duplicate samples were averaged with data from the routine samples in generating the enhanced data set.

3.5 SELECTION AND USE OF DATA SETS

The distributed NSS-I and Pilot Survey databases contain chemistry data, geographic attributes, and watershed descriptive information. For both the NSS-I and the Pilot Survey, three data sets are available: Data Set 3 (NSSIDS3 or PILOTDS3), Data Set 4 (NSSIDS4 or PILOTDS4), and a data set of field observations recorded by field crews about the immediate watershed area and sample site substrate (NSSIFSO). The decision as to which data set to use depends on the intended use. To examine population attributes (e.g., means, percentiles, or variances), Data Set 4 should be used. It is important that the statistical weighting factor (W) be used when examining population characteristics. Estimating population parameters without using sample weights can lead to biased estimates and inaccurate interpretation. However, if the intended use of the data is to examine the chemistry of individual water samples, Data Set 3 may be more useful, as it contains flags and tags that identify anomalies in the data or in the methodology used (e.g., holding time violations or analytical equipment discrepancies).

[2]A total of 225 values (0.5% of the database) across 9 variables (extractable aluminum, Doc, iron, manganese, ammonium, nitrate, total dissolved phosphorus, and silica) had values scored to zero, with total dissolved phosphorus having the most scoring changes, at 99 values. The bias due to this adjustment did not affect the population estimates presented. The range of negative observations scored to zero were low in absolute value and below instrument detection limits in all cases (cougan St al., 1988).

3.5.1 Data Set 3 Versus Data Set 4

The information contained in Data Set 3 is verified and validated but has not undergone any type of missing or erroneous value replacement or enhancement. Duplicate observations in Data Set 3 are not averaged. Each observation in Data Set 3 contains data associated with a particular container of water collected, including QC samples (e.g, blanks, duplicates, and audit samples); thus this data set is sample or visit specific. Quality control samples, such as duplicates, can be identified using the variable SAMCOD (“R” - routine, “D”- duplicate, “E” - suspected precipitation events). Unique observations can be identified in Data Set 3 by using the variables for batch and sample identification codes: BAT_ID and SAM_ID.

Data Set 4, in contrast, is visit specific, in that observations are intended to portray the chemistry that existed at the time of a visit to a sample reach. Quality control samples, except for field duplicate samples, have been removed in Data Set 4. In addition, field duplicate (QC) samples are averaged with the corresponding routine sample, erroneous and missing values are replaced, and impossible negative values are scored to zero. Because duplicate samples were averaged with routine samples, batch and sample identification codes are not included in Data Set 4. As in Data Set 3, data for multiple visits are retained (not averaged). Observations are uniquely identified using the stream identification code (STRM_ID) and the sample visit number (SAMRN).

3.5.2 The NSS-l Pilot Survev: The Southern Blue Ridge Province

In contrast to the NSS-l, which sampled both the upstream and downstream ends of a reach, the Pilot Survey sampled the upstream ends of only 20 of the 54 reaches during a spring index baseflow period in 1985 (Messer et al., 1986,1988). All 54 reaches, however, were sampled at both the upstream and downstream points during the summer 1985 (versus spring) index period. To provide population estimates compatible with the spring upstream estimates for other subregions, empirical relationships were used to synthesize missing values for 22 chemical variables for the 34 upstream sites not sampled during the spring index period. These synthesized values are not included in the enhanced database, but are provided as a separate small data set. The calculation of this synthesized data set is detailed in Appendix B of Kaufmann et al. (1988). The results presented for the Southern Blue Ridge in the NSS final report were generated using a data set made by appending SBRSYN to NSSIDS4.

It is important to note that after the Pilot Survey report was published, a number of data values were updated, resulting in a change in the estimated total number of reaches in the Southern Blue Ridge Province from 2,021 in the Pilot Survey report to 2,031 in the NSS-l report. These updated values were incorporated into data used in the NSS-I report to describe the Southern Blue Ridge Province (subregion 2As). The principal updates included the revising of one direct drainage value, a_1 (Al in the database), and a refinement in the method used to calculate ANC and BNC from Gran data.

The differences between the data sets used for the NSS-l report and the Pilot Survey re discussed in Appendix C. Because a_1 is used in calculating the final sample weight of each reach (the number of reaches the observation is estimated to represent in the target population), the number of estimated reaches in the Southern Blue Ridge Province changed by 10 reaches. The Pilot survey data sets, PILOTDS3 and PILOTDS4, were used in generating results presented in the F Survey report (Messer et a., 1986). The revised data for the Southern Blue Ridge are included NSSIDS4. This is the data set used in making estimates presented in the NSS-l report (Kaufmann, 1988).

SECTION 4

DEFINING THE TARGET POPULATION

4.1 EVALUATION PROCESS

This section discusses how the Stage II probability sample of reaches was evaluated and refined to a subset that represents the NSS-I target population of stream reaches. Although many nontarget reaches were screened out (using map information) during the Stage I selection process, it was necessary to further refine the set of "selected" reaches, based on field and chemical data, to best represent the target population of interest. In addition to identifying erroneous data, the validation process helped to identify unusual sites in the context of subregion populations. For example, reaches impacted from acid mine drainage comprise a portion of the total population for which the additional impact from acidic deposition is extremely difficult to ascertain. Therefore, these reaches were not included in estimates of the target population. Chemical data for such sites were not deleted from the NSS-I database, but were marked as noninterest to allow their inclusion or exclusion in statistical analyses with variables DRPCDE and SIT_CLS.

4.2 IDENTIFYING NONINTEREST OBSERVATIONS AND SITES

Noninterest observations were identified based on chemistry, field observations, and the investigation of watershed characteristics. The following basic criteria were used to identify sites and observations not included in estimates of the NSS-I target population:

- Intermittent: reaches that were at least 90% dry or stagnant.
- High specific conductance: reaches having an in situ specific conductivity > 500 μS cm^{-1} (e.g., reaches contaminated by oil well brine, industrial pollution, severe acid mine drainage).
- Episode (other than spring baseflow conditions): reach chemistry influenced by a precipitation event at or near the time of sampling (e.g., high turbidity and high flow) (Section 4.2.1).
- Low pH: reaches having a field pH value of 3.3 or less [e.g., severe acid mine drainage impact (Section 4.2.2)].
- Tidal influence: coastal reaches with water chemistry influenced by seawater (e.g., specific conductivity greater than 250 μS cm^{-1}).

  Reservoir:reach inundated by water project.

  ◦ No channel, dry: no sample could be collected because of a lack of water (dry) or a lack of explicit stream channel with flowing water (swamp).

Because the NSS-I target population estimates were intended to represent spring baseflow chemistry, field sampling was explicitly avoided during precipitation episodes. Although great care was taken not to collect samples during a precipitation episode, such conditions might not have b apparent to a sampling team. Samples inadvertently collected during a precipitation or snowmelt episode, or influenced by one, were identified as those meeting all of the following criteria:

- Site identified as a validation outlier (see Section 3.2).
- Change in stream gauge height of 7.5 cm or more between site visits, supporting evidence of flood stage, except for the Southeast sites where only one visit to a site was made.
- Field comments indicating a precipitation or snow melt event within 24 hours of sampling.
- Unusually high concentrations of turbidity, total aluminum, manganese, or iron relative to other visits at the same site or its corresponding upstream or downstream node.

4.2.2 Acid Mine Drainage

Streams impacted by acid mine drainage (AMD) (i.e., those distinguished by low alkalinity and I and markedly high sulfate) comprise a category that had to be identified in order to distinguish then from streams impacted by acidic deposition. Sites impacted by AMD were excluded in the calculation of NSS-I population estimates. These streams were identified as those meeting all of the following criteria, the rationale for which is discussed in subsection 9.3.1 of Kaufmann et al. (1988):

- ANC ≤ 0 $\mu eq L^{-1}$
- Sum of base cations > 400 $\mu eq\ L^{-1}$
- Sulfate/sum of anions > 75%
- DOC < 5 mgL^{-1}
- $[SO_4^{2-}] > 300\ \mu eq\ L^{-1}$ in the Mid-Atlantic; $[SO_4^{2-}] > 200\ \mu eq\ L^{-1}$ in the Southeast Mining activity confirmed by means of maps, field visits, or aerial photographs

4.3 CLASSIFICATION OF NONINTEREST OBSERVATIONS IN THE NSS-I DATABASE

Once specific site exclusion criteria were evaluated, observations classified as noninterest were assigned a site class code (SIT_CLS) to identify specific noninterest conditions. The first part of this code identifies the noninterest features and the second, the reach end or node (upstream or downstream) at which the condition was noted. These features are coded in the NSS-I database as:

Code	Condition

A	Sample point impacted by acid mine drainage
C	High conductivity at sample point (> 500 μS cm^{-1})
	Intermittent flow
T	Site chemically impacted from tidal activity
R	Random miss (see discussion below)
S	Special cases requiring the removal of specific observation (e.g., visit made to the wrong reach)

It is important to note that a site was considered sampled if a field crew visited a mapped sampling point and obtained information about the site, regardless of the presence of water. There were, however, four cases in which field crews were not able to visit the mapped sampling point (e.g., reach not found by field crews) and thus, no information about these missed sites was obtained. Such occurrences were considered to occur at random and each one is identified as a random miss marked with an 'R' in the SIT_CLS variable.[3]

The second character of the site class code designates the point on the reach at which the noninterest condition existed: "1 for the upstream site, “2" for the downstream site, or “3" for both sites. For example, a reach assessed as having high conductivity might have a SIT_CLS code of Cl to note the condition at the upstream point, C2 to note the condition at the downstream point, or C3 to note the condition at both points. When multiple noninterest conditions occurred on a reach, these codes were combined, (e.g., high conductivities found at both sample sites, in addition to acid mine drainage evidence at a downstream end would be classified as "A2C3").

The second character in SIT_CLS provides exclusion criteria for the entire reach regardless of the observation (or node) being examined. In the above example, the SIT_CLS code “A2C3" would appear in all observations for that reach.

[3] The occurrence of a random miss within a stratum required an adjustment in the final weighting factor for all reaches In that stratum (Overton, 1987).

4.4 DROP CODE VALUES

The drop code variable, DRPCDE, was created to allow the selective exclusion of noninterest observations in statistical analyses. All observations were assigned a code to indicate whether they are to be considered part of the target population or the noninterest group of reaches.

Population	Drop Code	Exclusion Criteria Description
Target	0	NSS-l target observation for both the upper and lower node.
Noninterest	1	Alternate node of noninterest site (see discussion that follows).
Noninterest	2	Noninterest sites (Kaufmann et al., 1988, Section 3.2).
Noninterest	3	Sites acidic due to acid mine drainage (Kaufmann et al., 1988, Subsection 9.3.1).
Noninterest	4	Pilot Survey nonspring index data (water samples not collected between March 15 and May 1) (in Kaufmann et al., 1988, Subsection 1.3.2.1).
Noninterest	5	NSS-I special interest sites (Kaufmann et al., 1988, Section 2.6).
Noninterest	13	Combination of DRPCDE 1 and DRPCDE 3 (i.e., an alternate node of a noninterest site is acidic due to acid mine drainage).

The set of observations considered to represent the target population of streams are those with a DRPCDE value of “1" or less. Population estimates were made for upstream and downstream locations as separate populations. However, when conducting analyses that examine those reaches at which both the upstream and the downstream ends fit the target criteria, the data can be subsetted by including only those observations with a DRPCDE value of 0.

Because specific noninterest conditions (e.g., intermittent flow) may eliminate either reach end, upstream or downstream, as a target, it is possible to have one end of a reach included in the target population, but not the other. For example, estimates were made in the NSS-l for different numbers of upstream and downstream ends.

SECTION 5
DATABASE APPLICATION

5.1 GENERATING THE DATA SET OF INDEX VALUES FROM THE REACHES SAMPLED

This section discusses the method used to calculate NSS population estimates. An important aspect of the NSS data design is that population estimates are based on a single observation for each stream reach. In the NSS-I report, separate estimates are made based on observations representing upstream and downstream sample sites. Using Data Set 4 (NSSIDS4, SBRSYN, PILOTDS4), this is accomplished by averaging multiple observations for each sample site. The resulting data set will contain information for the upstream and downstream ends of reaches, which are treated as representing separate populations of interest. We chose not to average upstream observations with downstream observations, but doing so does not violate the NSS design.

The process of generating an “indexed” data set includes:

- Identifying appropriate data sets (NSSIDS4 and SBRSYN for NSS-I estimates and PILOTIDS4 for NSS-I Pilot Survey estimates).
- Subsetting observations with a DRPCDE $\leq$ 1 (or deleting observations with DRPCDE $\geq$ 2).
- Calculating a mean value of the parameter(s) of interest (e.g., ANC, pH) for each STRMJD. Each resulting observation should contain a value for the parameter(s) of interest (e.g., ANC, pH), as well as data for SUB_ID, STRM_ID, NODE, W, STRATUM, and any other classification unit desired (e.g., STATE1, RCH_HW). Note: It is recommended that the mean of multiple pH values be calculated as the mean of hydrogen ion concentrations and converted into pH.

Table 5-1 lists the number of visits, total observations, high-interest samples, and nontarget samples, and the number of indexed observations by subregion.

5.2 EXTRAPOLATION TO THE TARGET POPULATION

The following section briefly discusses the methods used for making NSS-I population estimates. The computer programs used to generate the statistical estimates and graphics (cumulative distribution function curves, trilinear plots, etc.) are included in Sale et al., 1990, *Data Management and Analysis Procedures.*

5.2.1 Samp!e Weightings For Population Estimates

The NSS-I statistical design uses the attributes (e.g., chemistry, geographic features) of the 500 sampled reaches to describe the characteristics of an estimated 64,260 reaches in the target population. In the NSS-I database, each sample reach is assigned a calculated weight that indicates how many other reaches in the population are represented by that reach.

TABLE 5-1. NUMBER OF VISITS, TOTAL OBSERVATIONS, TARGET AND NONTARGET SAMPLES, AND INDEXED OBSERVATIONS BY SUBREGION

Region	Subregion	# of Sample Visits per Site[a]	# of Observations		# of Noninterest Observations		# of Target Observations		# of Spring Indexed Observations	
			Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower
Mid-Atlantic	1D	2	122	122	6	10	116	112	58	56
"	2Bn	2	103	102	16	11	87	91	44	47
"	2Cn	2	145	144	14	25	131	119	67	61
"	3B	2	119	114	7	6	112	108	57	58
Southeast	2D	1	50	50	1	2	49	48	49	48
"	2X	1	50	50	11	10	39	40	39	40
"	3A	1	50	50	3	3	47	47	47	47
"	3C	1	50	50	19	16	31	34	31	34
"	2Aa	1-5[c]	109d	233	55	78	54[d]	155	54[d]	54
Special Interest Sites	SI	1-4	2	84	–	–	–	–	–	–

[a] The number of sample visits correspond to the variable SAMRN in the NSS-I and Pilot Survey data base.

[b] The number of observations in the "indexed" (averaged) data set.

[c] Observations for 2As, the Southern Blue Ridge Province, are identified in the data base as 0-4. Only sample visits 1-3 are within the spring index period.

[d] 34 observations for the Southern Blue Ridge were synthesized for the NSS-I data report (see Section A.1.

This weighting factor is inversely proportional to a reach's probability of being selected and is the product of sample weights for two stages of sampling. Recall that in the first stage, the probability of selecting a reach was directly proportional to the direct drainage area of the reach, a_1[4]. In the second stage, sites were selected with probabilities inversely related to their first stage inclusion probability, largely equalizing the final sample weightings. After completion of the Stage I selection process, a few a_1 values were updated. As in the Pilot Survey, the final sample weight of a reach is directly related to its a_1. To adjust for these changes, an update was made to the Stage II conditional weights. Details of estimation and the statistical foundation are provided in Kaufmann et al. (1988), Section 2, Overton (1986, 1987), Blick et al. (1987), Overton and Stehman (1987), and Stehman and Overton (1987a, 1987b). Two weighting variables are present in the database, ,"WC", the conditional Stage II subsampling weight, and "W1". The final NSS-I sample weight, W, is used in extrapolating from the sample to the population:

$$W = WC\ (64/a_1)$$

where:

64 = Area (mi^2)/grid point

Note: If $a_1 < 0.2\ mi^2$, a_1 was scored to $0.2\ mi^2$ in the calculation of W. Although not present in the NSS database, an unscored weight can be generated by multiplying WC by $64/a_1$.

Whenever statistical descriptions of NSS-I reach populations are made, the sample weight, W, should be used. It is recommended that weights be used when calculating frequency distributions and other univariate analyses of chemical data, including the calculation of means. However, NSS-I data users are cautioned to assess the use of weights in bivariate or muitivariate explorations of the data. Scatter plots, correlations, and regressions are affected by the use of weights and careful thought should be given, especially if data are combined over strata (e.g., across subregions). Keep in mind that unweighted parameter estimates do not necessarily represent those of the population.

An important issue in the use of weights to generate population statistics is how to calculate standard errors associated with estimates. The NSS-I bases standard error estimates on a variation of the Horvitz-Thompson variance estimator (Overton, 1987). Users are cautioned in the use of sample weights in statistical analyses (see Section A.4). Any analyses that incorporate a standard error estimate must use the standard error algorithm used in the NSS.

[4] In a small number of reaches, a_1 was less than O.2 mi^2. In these cases the value $0.2\ mi^2$ was substituted as the divisor to reduce the variance of the population estimates (Overton, 1987, discusses the rationale and application of this scoring process)

5.2.2

Estimating the Target Population

In the NSS-I, the basic parameter estimate (and variance) of the population of interest is based on the total number of reaches, as portrayed by the upstream sites or downstream sites separately, or the total of any attribute of interest, using the Horvitz-Thompson (1952) estimator:

$$\hat{T}_y = \sum_{i \in S} w_i y_i$$

where:

- $\hat{T}_y$ is the estimate total of any attribute, y, over the population.
- y is any attribute of interest over the sample$_1$ S. When making estimates of the number of reaches, y = 1; for estimates of reach length, y RCH_LN; for total watershed area estimates, y = A_WS; for direct drainage area, Y = AI.
- w is the sample weight assigned to each reach.
- $\sum_{i \in S}$ indicates summation over the entire sample (or subset) of target reaches.

A primary product of the NSS-I, these estimates of the distribution of target population attributes were made by assigning different definitions to y (e.g., total length), and summing over different sets of sample units, S (e.g., Subregion). The NSS-I examines several attributes of the target population:

- Number of target reaches $\hat{N}(x), y = 1$: $\hat{N}(x) = \sum_{X \le x} w_i$
- Total watershed area $\hat{A}(x), y = A_i$: $\hat{A}(x) = \sum_{X \le x} w_i A_i$
- Length of target reach $\hat{L}(x), y = l_i$: $\hat{L}(x) = \sum_{X \le x} w_i / i$

where:

X is the attribute of interest (e.g., ANC, pH).

x is an individual value or observation.

- $X \preceq x$ refers to those conditions in which the estimate is incrementally made over a range of an attribute (i.e., $X(x)$ is the estimated number of reaches in the target population with a value $\preceq$ x).

- A_i is the topographic watershed drainage above a sample site ("A_WS" in the NSS database).

- I is the length of reach segment between mapped confluences ("RCH_LN" in the NSS-I database).

The distribution of attributes across a range of chemistry is calculated by computing estimates for each value of x, and then, in a cumulative manner, dividing by the total number of estimated reaches (or sum of weights). The following shows a distribution for example data.

ANC	Sample Weight	Cumulative Weight	Percent of Total	Cumulative Percent
-50.0	15	15	0.021	0.021
-25.5	150	165	0.209	0.231
0.0	50	215	0.069	0.301
25.7	250	465	0.348	0.650
50.3	35	500	0.049	0.699
200.1	215	715	0.301	1.000

For each distribution, it is possible to estimate quartiles, the median, and quintiles referred to as Qi, 02, Q3, 04. The median is the value of x such that it is ½ or 0.5 on a cumulative proportion F(X) curve (Section 5.3.1). These statistics are defined and presented for all distributions in Sale et al. (1988). Additionally, the mean and standard deviation of the variable x in the population can be estimated.

$$Mean(x) = \sum w_i x_i / \sum w_i$$

$$SD(x) = \sqrt{\sum w_i x_i^2 / \sum w_i - \left[\sum w_i x_i / \sum w_i\right]^2}$$

5.2.3 Variance Estimates and Confidence Bounds

The variance estimates, leading to estimated standard errors (SE) were obtained by application of an original variation of the Horvitz-Thompson variance estimator (Overton, 1986; Stehman and Overton, 1 987a). The formula for the estimated variance is:

$$\hat{V}\left(\hat{T}y\right) = \sum_{i \in S} y_i^2 w_i \left(w_i - 1\right) + \sum_{i \in S} \sum_{\substack{j \in S \\ j \neq i}} y_i y_j v_{ij}$$

where:

- if observations i and j are from the same stratum (Subregion and ANC), then $V_{ij} = \left[\left(w_i + w_i\right)/2 - w_i\, w_j\right]/(n-1)$ else, if i and j are from different strata then $v_{ij} = o$
- I s the Stage II effective sample size for each stratum (from Table 2-1).

Simply put, the first portion of this equation is a variance component while the second is a covariance component, specific to the observations of each stratum.

Once variance estimates are made, the associated standard errors and one-sided 95% upper confidence bounds can be generated. The following statistics are associated with N, the estimated total number of target reaches, estimated as [$\hat{T}_y$, where y = 1 (see Section 5.2.2)]:

- The weighted cumulative proportion estimates calculated as . $\hat{F}(x) = \hat{N}(x) / \hat{N}..$
- The standard error for this estimate (A (x)n), calculated as SE $\left(\hat{N}(x)\right) = \sqrt{\hat{V}\left(\hat{N}(x)\right)}.$
- one-sided 95% upper confidence bound[5], calculated (assuming a normal distribution) as $\hat{N}_u(x) = \hat{N}(x) + 1.645\left[SE\left(\hat{N}(x)\right)\right]..$
- Weighted percentile estimates of one-sided 95% upper confidence bound, calculated as $F_u(x) = \hat{N}_u(x) / \hat{N}..$

Although not presented in the NSS-I report, a two-sided 95% confidence bound can be generated as $\hat{N}_L(X) = \hat{N}(x) \pm 1.96\left[SE\left(\hat{N}(x)\right)\right]..$

5.3 DESCRIBING THE TARGET POPULATION

This section outlines ways the NSS-I data can be used to describe the target population of streams. The database structure facilitates the examination of the target population in a number of ways. Whereas the range of target streams is explicitly defined in terms of size, watershed area, and general water quality conditions, the Survey design allows the end user to examine the entire population of sampled streams, and not just those used to calculate refined target estimates. For example, if one is interested in examining that portion of the NSS-I population impacted by acid mine drainage,

[5]For the majority of the NSS-I estimates, only a one-sided upper 95% confidence bound was generated as depicted In distribution plots in Sale et a. (1988). A lower one-sided 95% confidence bound can be generated, if needed, by assuming that a lower confidence bound is approximately equal to $\hat{N} - 1.645\left[SE\left(\hat{N}\right)\right].$

the data can be subsetted using the appropriate drop codes and weights. Examples of other NSS-I populations might include:

- The initial NSS-I target population (i.e., the population represented by the complete set of sampled Stage II reaches, including noninterest and interest sites).

  [Subsetting observations based on a DRPCDE < 4 and then indexing] The set of headwater

  reaches.

  [Subsetting observations based on RCH_HW = 1 and then indexing]

- The set of reaches located within a specific geographic area (e.g., NSS-I reaches whose upstream or downstream ends are located in Maryland).

  [Subsetting observations based on STATE1 = 'MD']

  The set of reaches defined by some chemical attribute (i.e., reaches with DOC < 10 mg L^{-1}).

  [Subsetting observations based on DOC11 $\leq$ 10]

Again, it is important to remember that during examination of the population represented by the sampled reaches, the sample weighting factor, W (contained in Data Set 4), must be used.

Table 5-2 presents estimates of the target stream population by subregion and node for the total number of sites and total reach length. Estimates were made using the equations presented in Section 5.2. The set of streams represented in Table 5-2 is identified in the NSS-I database as those sites with a DRPCDE value < 1. Special interest sites are not used in making statistical estimates of the target population. This data subset includes only those sites that represent spring baseflow conditions and are not grossly impacted from influences (e.g., acid mine drainage, tidal influence) that mask the effects of acidic deposition.

5.3.1 The Cumulative Distribution Function (CDF) Curve

In the NSS-I, distributions are generated for measured attributes using the general estimators (e.g., numbers, length, watershed area) and presented using the cumulative distribution function curve, Fix), sometimes called a cumulative proportion (Overton, 1989). The distribution, Fix), shown in Figure 5-1, is interpreted as the proportion of target reaches in the population having the attribute X $\leq$ x. To read this figure, pick a value of x of the attribute X, along the horizontal axis (ANC in this example) and read the y-axis value of the two curves, F(x) at this value. The F(x) is the estimated proportion of reaches in the population with a value of the attribute equal to or less than X. In this example, the median or 5Oth percentile (i.e., $\hat{F}(x)$ = 0.5) is read as 257 μeq $^{-1}$ The estimated number of downstream reach ends, which is less than or equal to this value, or $\hat{N}(x)$, can be

TABLE 5-2.

TOTAL RESOURCE ESTIMATES OF THE NSS REFINED TARGET POPULATION

Subregion	Node[a]	n	$\hat{N}$	SE(N)	$\hat{L}$	SE $\left(\hat{L}\right)$
ID	U	58	3244	347	15270	1911
	L	56	3235	347	15144	1912
2Bn	U	44	13038	1249	32687	4492
	L	47	13992	1213	36405	4678
2Cn	U	67	8663	807	22373	2732
	L	61	8488	814	21738	2746
3B	U	57	11284	1078	40296	5799
	L	58	11287	1078	40344	5799
2As	U	54	2031	326	9036	960
	L	54	2031	326	9036	960
2D	U	49	4204	406	22753	2485
	L	48	4116	410	22480	2507
2X	U	39	4936	529	21892	2807
	L	40	5057	526	23015	2895
3A	U	47	7515	650	33531	4402
	L	47	7515	650	33531	4402
3C	U	34	1727	437	4312	690
	L	31	1555	306	4820	731

[a]U = upstream or “upper node” sample sites; L = downstream or”lower node” sample sites.

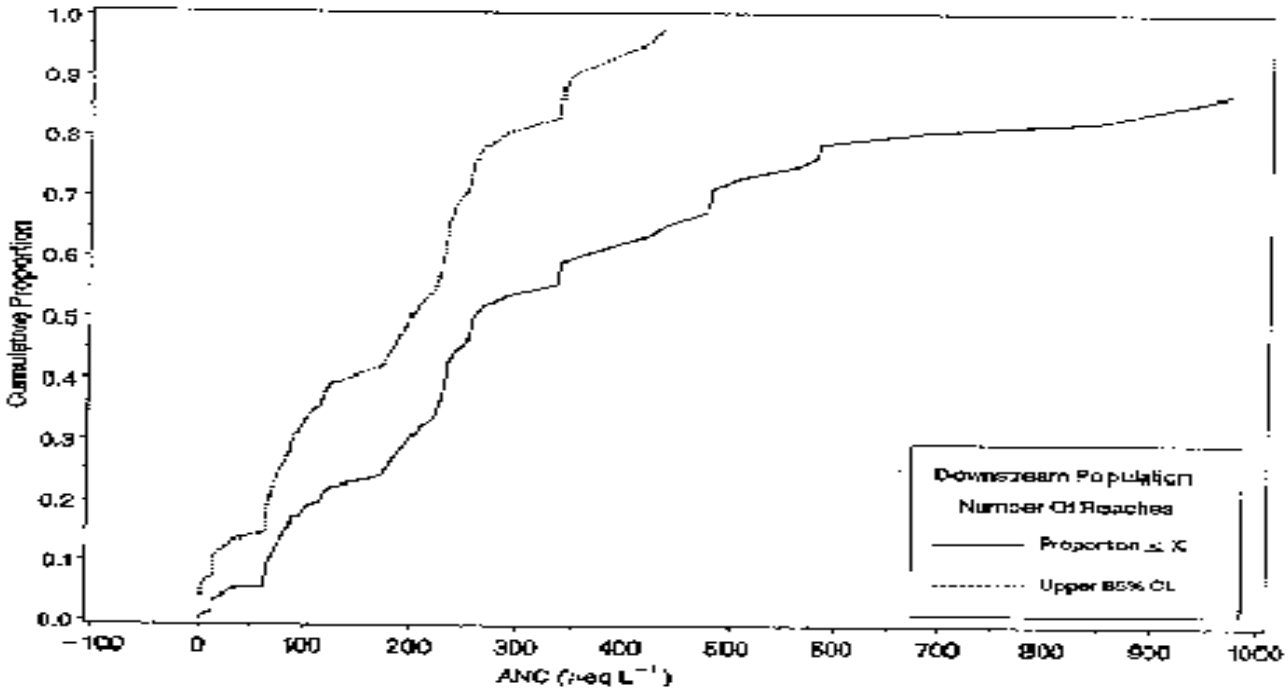
n = number of sampled stream reaches.

$\hat{N}$ = estimated target population size.

SE $\left(\hat{N}\right)$ = standard error of $\hat{N}$.

$\hat{L}$ = estimated total stream reach length of the target population, based on the variable RCH_LN (reach length).

SE $\left(\hat{L}\right)$ = standard error of $\hat{L}$.



A cumulative distribution function curve can be calculated for the NSS target population of downstream sites based on the following.

1. Subset the data such that:
   *Only those observations that have a DRPCDE < 1 are included, and
   *Only those observations for the downstream points (NODE=L) of interest.

   2.Generate a data set of mean chemistry values of Interest for each reach (by the variable STRM_ID in the database). Note: be sure to maintain weights in the new data set At this point, only one observation should exist for each downstream sample site.
   3.Identify the population subset to analyze (e.g., streams in Maryland).
   4.Sort the observations based on the parameter of interest (i.e., ANC).
   5.In a cumulative manner by each observation of x, divide the weight by the estimated population total (i.e., a cumulative percentage is calculated by summing up the weights over the distribution of interest and dividing by the population total) such that for each value of x, $F(x) = \hat{N}(x)\hat{N}.$

   6.The resulting data can be plotted as an XV plot with the cumulative percentage (from 0 to 1) on the Y-axis and the parameter of interest on the x-axis. Assuming that values lying between the calculated percentages are located along a straight line between plotted points, these points are connected by a straight line.
   7.The upper one-sided confidence bound is estimated for each value of x (Section 5.2.3) and a cumulative proportion is estimated [on each value of N(x)], based on the estimate of the total number of sites, N.

Figure 5-1. Calculation of an example cumulative distribution function curve for downstream sites.

calculated by multiplying $\hat{F}(x) by \hat{N}_{total}$. The dotted line in this figure can be used to estimate a one sided 95% confidence bound on $\hat{N}(x)$ representing $\hat{N}_u(x)$. In this example, $\hat{N}_u(x)$ for the median is calculated as 0.76 * $\hat{N}_{total}$.

In addition to distributions presented as F(x), some are presented as an inverse, or descending, cumulative proportion, or 1 - F(x). For these, read the distribution as the estimated proportion of reaches having values equal to or greater than $x(X \leq x)$.

5.3.2 Length Estimates

In addition to distributions in terms of the numbers and percentages of stream reaches, estimates were made of the combined length of reaches in the target population (Overton, 1989). These length estimates were calculated using two different approaches that yielded different estimates.

5.3.2.1 First Approach: Length Estimates Based on Node Chemistry--

The first approach used to calculate the length estimates presented in Sale et al. (1988), assigns a measured chemical parameter value (e.g., ANC) at the downstream node to the entire length of sample reach (using the variable RCH_LN).

1. Subset over unit of sample interest (e.g., subregion 2As) for observations with DRPCDE $\leq$ 1
2. Generate data set of mean values (Section 5.1) with a single value for each upstream and downstream end.
3. From data set of index values, generate data set that merges the upstream and downstream data into a single observation, keeping the chemistry associated with each node identified separately. For each observation, there will be data from upstream and downstream ends.
4. Calculate weighted length for each reach by multiplying each value of RCH_LN by reach sample weight, W.
5. Sort data generated in step 4 by the chemistry value of interest and generate a CDF (using method described in Section 5.3.1).

5.3.2.2 Second Approach: Length Estimates Based on Interpolated Length--

The second method (Illustrated in Kaufmann et al., 1988) assumes the chemical concentrations change in a linear fashion from the upstream sampling point to the lower one. This method estimates a series of linearly interpolated chemical values along a reach length between the upstream and downstream ends. Each segment is then assigned the reach's respective sample weight, thereby extrapolating to the target population of reaches. Figure 5-2 presents cumulative length estimates for chemistry. As expected, the interpolated upper node chemistry, and interpolated length subregion 2As,

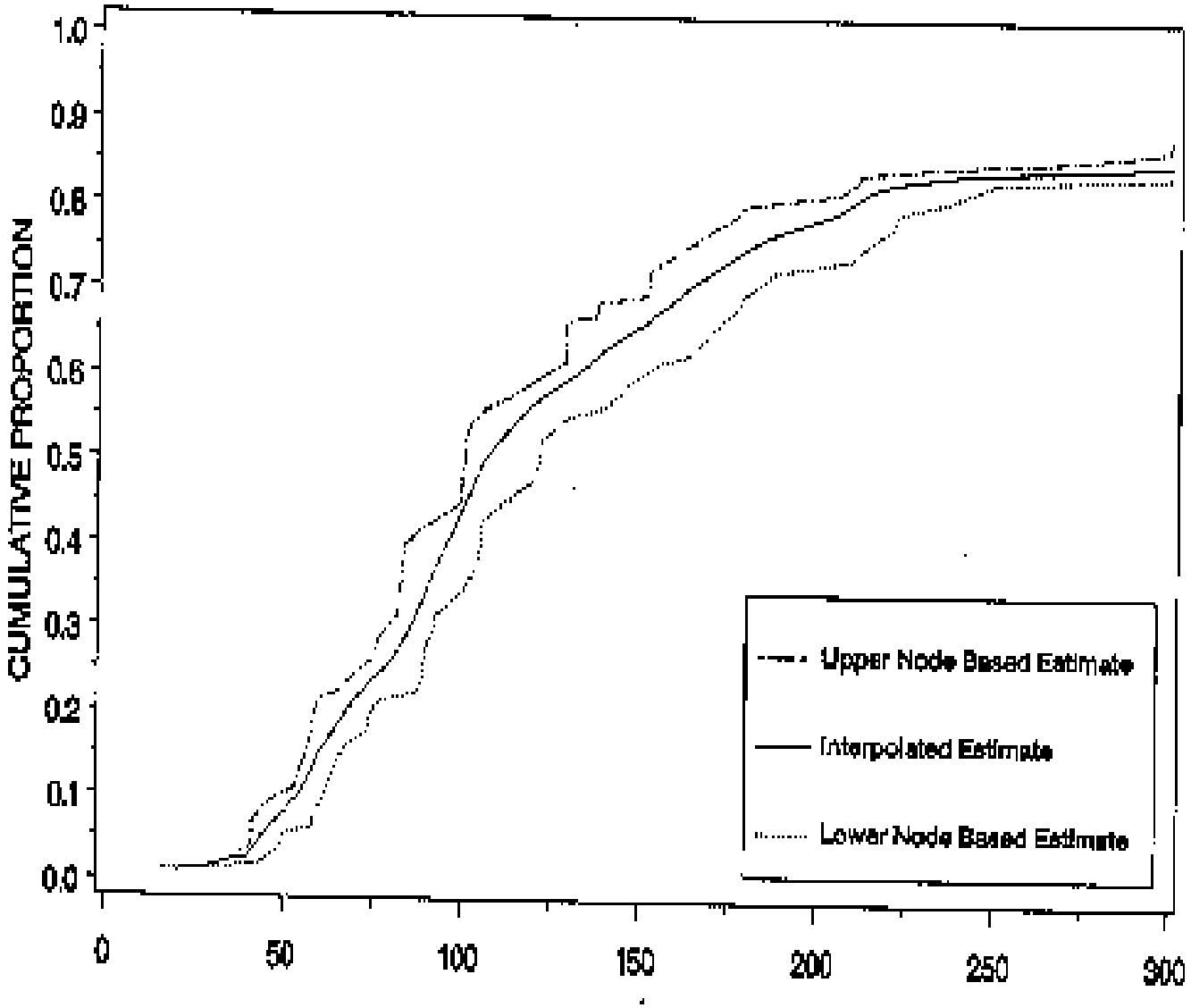


Figure 5-2. Interpolated length estimates for subregion 2As. Overlay of three CDFs of ANC concentration versus cumulative reach length based on (a) chemistry from upstream sites or nodes, (b) chemistry from downstream sites or nodes, and (c) an interpolated chemistry and length value between the upstream and downstream site or node values.

length distributions generally are bounded by the upstream and downstream distributions. Note: when distribution estimates are made for pH, the interpolation is actually done on [H+] and converted to pH values (based on PHSTVL). The following two methods can be used to estimate interpolated NSS-I reach lengths.

Estimated Interpolated Length: Method 1 - Interpolation of Reach Segments

1. Subset over unit of sample interest (e.g., Subregion 2As) for observations with DRPCDE $\leq$ 1. (Sites with DRPCDE = 0 have both upstream and downstream reach ends included in the target population.)

2. Generate data set of mean values (Section 5.1) with a single value for each upstream and downstream end.

3. From data set of index values, generate data set that merges the upstream and downstream data into a single observation, keeping the chemistry associated with each node identified separately. For each observation, there will be data from upstream and downstream ends.

4. Divide each reach length (using variable RCH_LN) into 0.1 km segments. Determine the difference for the chemistry value of interest between upstream and downstream ends for each reach and divide by number of 0.1 km segments estimated for reach. Resulting value represents estimated incremental change that occurs over each 0.1 km segment. Starting at one end, assign a chemistry for each segment, incrementally estimating each segment chemistry with each increment in length.

5. Sort data set generated in step 4 by the chemistry value and generate a CDF (using the method described in Section 5.3.1).

Estimated Interpolated Length: Method 2 - Simple Linear Interpolation

1. Subset over unit of sample interest (e.g., Subregion ~s) for observations with DRPCDE $\leq$ 1.

2. Generate data set of mean values (Section 5.1) with a single value for each upstream and downstream end.

3. From data set of index values, generate data set that merges upstream and downstream data into a single observation, keeping the chemistry associated with each node identified separately. For each observation, there will be data from upstream and downstream ends.

4. Using reach length (RCH_LN) and difference in chemistry between ends of each reach, it is possible to interpolate a length of each stream that is < a particular reference value. By doing this over an incremented range of chemistry, interpolated length estimates can be made for an entire subset of data. (In Figure 5-2, estimates were made for each value of ANC from 0 to 1,000, in 1 μeq L^{-1}increments.) **Note:** this can only be done if the range of reference values encompasses the range of observed values in the data. Also, the resolution of this method is limited by the increment size of the reference values (in this example, 1 μeq L^{-1}).

5. Sort data set generated in step 4 by the chemistry value and generate a CDF (using the method described in Section 5.3.1). Note: a standard error can be calculated for interpolated length estimates by assigning interpolated length as the attribute of interest (see Section 5.2.3). **Note:** a standard error can be calculated for interpolated length estimates by assigning interpolated length as the attribute of interest (see Section 5.2.3).

American Society for Testing and Materials. 1984. Standard specification of reagent water. Dl193-77 (reapproved 1983). In: Annual Book of ASTM Standards. Vol.11.01. Philadelphia, Pennsylvania

Arent, L.J., M.D. Morison, and C.S. Soong. 1988. Eastern Lake Survey - Phase II, National Stream Survey - Phase I: Processing Laboratory Report. EPA/600/4-88/025. U.S. Environmental Protection Agency, Washington, D.C. 86 pp.

Binder, D., J. Kovar, S. Kuar, D. Paton, and A. van Baaren. 1987. Analytic uses of survey data: a review. Pages 243-284. In: 1.8. MacNeill and G.J. Umphrey, eds. Applied Probability, Stochastic Processes, and Sampling Theory. D. Reidel Publishing Co.

Blick, D.J., J.J. Messer, D.H. Landers, and W.S. Overton. 1987. Statistical basis for the design and interpretation of the National Surface Water Survey, Phase I: Lakes and Streams. Lake Reserv. Manage. 3:470-475.

Cougan, KA., D.W. Sutton, D.V. Peck, V.J. Miller, and J.E. Pollard. 1988. National Stream Survey - Phase I: Quality Assurance Report. EPA/600/4-88/018. U.S. Environmental Protection Agency, Washington, D.C. 205 pp.

Cummins, K.W. 1962. An evaluation of some techniques for the collection and analysis of benthic samples with special emphasis on lotic waters. Am. MidI. Nat. 67:477-504.

Drousé, S.K, D.C. Hiliman, L.W. Creelman, and S.J. Simon. 1986. National Surface Water Survey: National Stream Survey (Pilot, Middle-Atlantic Phase I, Southeast Screening, and Episodes Pilot). Quality Assurance Plan. EPN6OO/4-86/044. U.S. Environmental Protection Agency, Washington, D.C. 215 pp.

Drousé, S.K 1987. Evaluation of Quality Assurance and Quality Control Sample Data for the National Stream Survey (Phase I - Pilot Survey). EPN6OOI8-87/057. U.S. Environmental Protection Agency, Washington, D.C. 45 pp.

Hagley, C.A., C.L Mayer, and R. Hoenicke. 1988. National Surface Water Survey: National Stream Survey (Phase I, Southeast Screening, and Episodes Pilot). Field Operations Report. EPN6OO/4- 88/023. U.S. Environmental Protection Agency, Washington, D.C. 38 pp.

Hillman, D.C., S.L Pia, and S.J. Simon. 1987. National Surface Water Survey: National Stream Survey (Pilot, Middle-Atlantic Phase I, Southeast Screening, and Middle-Atlantic Episodes Pilot). Analytical Methods Manual. EPN6OO/8-871005. U.S. Environmental Protection Agency, Washington, D.C. 265 pp.

Kanciruk, P., R.J. Olson, and R.A. McCord. 1986. Quality control in research databases: the U.S. Environmental Protectional Agency National Surface Water survey experience. In: W.K. Michener (ed). Research Data Management in the Ecological Sciences, University of South Carolina Press, Columbia, South Carolina. pp.193-207.

Kaufmann, P.R., A.T. Herlihy, J.W. Elwood, M.E. Mitch, W.S. Overton, M.J. Sale, J.J. Messer, KA. Cougan, D.V. Peck, KH. Reckhow, A.J. Kinney, S.J. Christie, D.D. Brown, C.A. Hagley, and H.l. Jager. 1988. Chemical Characteristics of Streams in the Mid-Atlantic and Southeastern United States. Volume I: Population Descriptions and Physico-Chemical Relationships. EPN6OO/3-88/021 a. U.S. Environmental Protection Agency, Washington, D.C. 397 pp.

Knapp, C.M., C.L Mayer, D.V. Peck, J.R. Baker, and G.J. Filbin. 1987. National Surface Water Survey: National Stream Survey (Pilot Survey). Field Operations Report. EPN60OI8-87/01 9. U.S. Environmental Protection Agency, Washington, D.C. 66 pp.

Kramer, J.R. 1984. Modified Gran analysis for acid and base titrations. Environmental Geochemistry Report No.1984-2.
McMaster University, Hamilton, Ontario, Canada.

Landers, D.H., J.M. Eilers, D.F. Brakke, W.S. Overton, P.E. Kellar, M.E. Silverstein, R.D. Schonbrod, R.E. Crowe, R.A. Linthurst, J.M. Omernik, S.A. Teague, and E.P. Meier. 1987. Characteristics of Lakes in the Western United States. Volume I. Population Descriptions and Physico-Chemical Relationships. EPN600/3-86/OMa. U.S. Environmental Protection Agency, Washington, D.C.176 pp.

Linthurst, R.A., D.H. Landers, J.M. Eilers, D.F. Brakke, W.S. Overton, E.P. Meier, and R.E. Crowe. 1986. Characteristics of Lakes in the Eastern United States. Volume I. Population Descriptions and Physico-Chemical Relationships. EPN6OO/4-86/007a. U.S. Environmental Protection Agency, Washington, D.C. 136 pp.

Messer, J.J., and KN. Eshleman. 1987. The Feasibility of Quantifying the Regional Extent, Magnitude, Duration, and Frequency of Episodes: Results of the National Surface Water Survey Pilot Studies. Presentation at NAPAP Aquatic Effects Task Group VI Peer Review, May 17-23, New Orleans, Louisiana.

Messer, J.J., C.W. Ariss, J.R. Baker, S.K Drous~, KN. Eshleman, P.R. Kaufmann, R.A. Linthurst, J.M. Omernik, W.S. Overton, M.J. Sale, R.D. Schonbrod, S.M. Stambaugh, and J.R. Tuschall, Jr. 1986. National Stream Survey Phase I - Pilot Survey. EPN6OO/4-861026. U.S. Environmental Protection Agency, Washington, D.C. 179 pp.

Messer, J.J., D.H. Landers, R.A. Linthurst, and W.S. Overton. 1987. Critical design and interpretive aspects of the National
Surface Water Survey. Lake Reserv. Manage. 3:463469.

Messer, J.J., C.W. Ariss, J.R. Baker, S.K Drousé, KN. Eshleman, A.J. Kinney, W.S. Overton, M.J Sale, R.D. Schonbrod. 1988. Stream chemistry in the Southern Blue Ridge: feasibility of a regional synoptic sampling approach. Water Resour. Bull. 24:821-829.

Nathan, G. 1988. Inference based on data from complex sample. Pages 247-266. In: P.R. Krishnaiah and C.R. Rao, eds. Designs Hand of Statistics. Volume 6. Elseveier Science Publishers B.V.

O'Dell, J.W., J.D. Pfaff, M.E. Gales, and G.D. McKee. 1984. Technical Addition to Methods for the Chemical Analysis of Water and Wastes. Method 300.0. The Determination of Inorganic Anions in Water by Ion Chromatography. EPN6OO/4-85101 7. U.S. Environmental Protection Agency, Cincinnati, Ohio.

Oliver, B.G., E.M. Thurman, and R.K Malcolm. 1983. The contribution of humic substances to the acidity of colored natural waters. Geochim. Cosmochim. Acta 47:2031-2035.

Omernik, J.M., and A.J. Kinney. 1985. Total alkalinity of surface waters: a map of the New England and New York Region. EPN6OO/D-84121 6. U.S. EPA Environmental Research Laboratory, Corvallis, Oregon.

Omernik, J.M., and C.F. Powers. 1983. Total alkalinity of surface waters--a national map. Ann. Assoc. Am. Geog. 73:133-136.

Overton, W.S. 1986. A Sampling Plan for Streams in the National Stream Survey. Technical Report 114. Department of Statistics, Oregon State University, Corvallis, Oregon. 18 pp.

Overton, W.S. 1987. A Sampling and Analysis Plan for Streams in the National Surface Water Survey. Technical Report 117. Department of Statistics, Oregon State University, Corvallis, Oregon. 50 pp.

Overton, W.S., and S.V. Stehman. 1987. An Empirical Investigation of Sampling and Other Errors in the National Stream Survey; Analysis of a Replicated Sample of Streams. Technical Report 119. Department of Statistics, Oregon State University, Corvallis, Oregon. 21 pp.

Overton, W.S. 1989. Effects of Measurement and Other Extraneous Errors on Estimated Distribution Functions in the National Surface Water Surveys. Technical Report 129. Department of Statistics, Oregon State University, Corvallis, Oregon. 21 pp.

Pollack, A.K, and S.C. Grosser. 1988. National Stream Survey Data Base Audit. SYS APPB8/088. Systems Applications, Inc., San Rafael, California. 49 pp.

Sale, J.J. (ed.). 1990. Data Management and Analysis Procedures for the National Stream Survey. Oak Ridge National laboratory, Oak Ridge, TN. (Draft)

Sale, M.J., P.R. Kaufmann, H.l. Jager, J.M. Coe, KA. Cougan, A.J. Kinney, M.E. Mitch, and W.S. Overton. 1988. Chemical Characteristics of Streams in the Mid-Atlantic and Southeastern United States. Volume II: Streams Sampled, Descriptive Statistics, and Compendium of Physical and Chemical Data EPN600/3~8I021 b. U.S. Environmental Protection Agency, Washington, D.C. 595 pp.

SAS Institute, Inc. 1985. SAS User's Guide: Statistics, Version 5 Edition. SAS Institute, Inc., Cary North Carolina. 956 pp.

Shreve, R.L 1966. Statistical law of stream numbers. J. Geol. 74:17-37.

Skougstad, M.W., J. Fishman, LC. Friedman, D.E. Erdmary, and 8.8. Duncan, eds. 1979. Methods for Determination of Inorganic Substances in Water and Fluvial Sediments: Techniques of Water Resources Investigations of the United States Geological Survey. Book 5, Chapter Al. U.S. Government Printing Office, Washington, D.C.

Steh man, S.V. and W.S. Overton. 1 987a. Estimating the variance of the Horvitz-Thompson estimator for variable probability, systematic samples. Proceedings of the Survey Research Section. 1987 ASA Annual Meetings, San Francisco, CA.

Stehman, S.V. and W.S. Overton. 1 987b. An Empirical Investigation of the Variance Estimation Methodology Prescribed for the National Stream Survey: Simulated Sampling from Stream Data Sets. Technical Report 118. Department of Statistics, Oregon State University, Corvallis, OR. 19 pp.

Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. Trans. Am. Geophys. Union 38:913-920.

U.S. Environmental Protection Agency. 1983. Methods for Chemical Analysis of Water and Wastes. EPN6OO/4-79/020. U.S. Environmental Protection Agency, Cincinnati, Ohio.

U.S. Environmental Protection Agency. 1987. Handbook of Methods for Acid Deposition Studies: Laboratory Analyses for Surface Water Chemistry. EPN6OOI4-871026. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. 342 pp.

APPENDIX A
NOTES OF CAUTION AND ISSUES OF INTEREST TO NSS DATA USERS

Working with the NSS data can be very complex. This appendix is provided to assist the data user with issues and cautions pertaining to data use. Although some topics are discussed in other sections of this guide or in other documents, this section details all known issues of concern and sources of potential problems for NSS data users. These include:

Section	Description
A.1	Use of Drop Codes: DRPCDE
A.2	Generating a Working Data Set
A.3	Data Set of Field Site Observations
A.4	Use of Sample Weights: W
A.5	Chemical Variables with Similar Names
A.6	Population Estimates for Geographic Subsets of Streams
A.7	Reach Length Estimates: RCH_LN versus L2
A.8	Topographic Drainage Area Measurements: a_1, a_2, a_3, a_4, a_5, a_{total} (Al, A2, A3, A4, A5, and A_WS)
A.9	Synthesized Data for the Southern Blue Ridge (Subregion 2As)
A10	Differences Between the NSS-I and the NSS Pilot Survey
A11	All Revisions to NSS-I Pilot Survey Data
A.12	Comparison of Parameter Units in the NSS-I and the NLS
A.13	Revisions to a_1
A.14	NSS Database Variable Formats
A15	Subregion Identification Codes
A.16	Using DIC and pH in Calculated Variables

A.1 USE OF DROP CODE VARIABLE: DRPCDE

(detailed In Section 4.4)

The distributed NSS-I data sets contain information for all sampled streams, including those having sampling points or conditions considered to be “noninterest” in the context of the NSS-I target population. These noninterest conditions indicate influence by factors that make it difficult to discern the impact from acid deposition (e.g, acid mine drainage, tidal influences, or sample collection during a baseflow period other than spring). Such observations were not used in generating statistical estimates of the NSS-I target population (in terms of status and extent) but are included in the distributed data to allow for the examination of a variety of stream subsets.

Noninterest sites can be identified by using the variable, DRPCDE, created to allow the inclusion or exclusion of noninterest sites from specific analyses. By subsetting the data for only those sites with DRPCDE < 1, the resulting data set includes only those observations considered to represent spring baseflow chemistry in the target population. This data subset will exclude data from sites found to be grossly polluted, sampled during a summer (non-spring) base flow period, or influenced by precipitation episode conditions. In addition, data from special interest sites were not used when generating NSS-I target population estimates and are excluded when the drop code variable is applied.

A.2 GENERATING A WORKING DATA SET

(Detailed In Section 5.1)

All analyses and estimates made for the NSS-I target population were generated using Data Set 4 (NSSIDS4). The NSS-I data comprises information collected from one or more visits to upstream and downstream ends of reaches. An important aspect of using the NSS-I data to generate population estimates is that upstream ends of reaches are considered to represent a population of stream locations different and separate from a population of downstream ends. Therefore, a working data set is generated in which there is only one observation for each upstream reach end and one observation for each downstream reach end. This is accomplished by first subsetting the data using the drop code variable and then averaging multiple visits to the same sampling site, but not upstream sites with downstream sites. Again, it is important to remember that population estimates for upstream reach ends were generated separately from estimates of NSS-I downstream reach ends in the final NSS report (Kaufmann et a., 1988; Sale et a., 1988).

A.3 DATA SET OF FIELD SITE OBSERVATIONS

(Detailed In appendix D)

Originally categorized as "Watershed Characteristics", this information, contained in NSSIFSO, pertains only to the area in the immediate vicinity of the sampling site. Field crews recorded observations of watershed characteristics near the site but did not perform an extensive field reconnaissance of the entire watershed area to identify disturbances that may influence reach chemistry. These data were collected to document potential watershed influences on sample chemistry and do not identify all potential disturbances in the entire watershed. These observations were recorded in a standardized format for disturbances (e.g., nearby roads, housing, agriculture, industry, and logging) and substrate composition and vegetation coverage of the immediate sampling location. Although none of this information has undergone the level of quality assurance that was applied to the chemistry data, such information is helpful when interpreting individual sample chemistries.

A.4 THE USE OF SAMPLE WEIGHTS: W

(Detailed In Section 5.2.1)

Generating statistical estimates of NSS-l populations can be very complex. Because streams were sampled with varying probabilities, the weighting factor, W, should be used to obtain unbiased estimates. In this manner, population means, totals, and proportions can be calculated in a relatively straightforward manner. However, standard errors of these estimates are difficult to compute. Since most statistical packages do not compute a standard error in a manner comparable to the Horvitz-Thompson theorem, the algorithm used in the NSS (see Overton, 1988), care must be taken when generating such estimates. The algorithm used to estimate standard error in the NSS is described in Section 5.2.1. The use of weights in computing regressions and other more complicated statistics is also potentially very complex and should be undertaken with care (Binder et al., 1987; Nathan, 1988).

A.5 CHEMICAL VARIABLES WITH SIMILAR NAMES

Care should be taken in identifying the most appropriate variable for each analysis. There are as many as five different types of variables for aluminum, pH, DIC, specific conductance, and phosphorus, as well as the flag variables. In addition, certain variables are listed with suffixes of both “11" and "16" (e.g., CA11 and CA16). The "11" suffix refers to values presented in the units measured at analytical laboratories, and the "16" suffix identifies variables that have been converted or transposed into a unit considered more appropriate for reporting results. The suffix "VAL" is used to

refer to certain parameters measured at the processing laboratory (e.g., CONVAL, DICVAL, etc.). The data dictionary (Appendix B) should be consulted for review of specific parameters and their units.

A.6 POPULATION ESTIMATES FOR GEOGRAPHIC SUBSETS OF STREAMS

When making estimates of the target population of reaches within a geographic subunit of the NSS (e.g.$_1$ state of West Virginia or New York), note that such estimates apply only to streams that are located within the areas actually surveyed and that fit the target reach criteria. In other words, the NSS-I may have sampled only part of a larger geographic area of interest (e.g., two areas in the state of Florida) and it would not be appropriate to assume that the NSS-I estimates apply to areas outside the NSS-I subregion boundaries (e.g., the entire state of Florida).

A.7 REACH LENGTH ESTIMATES: RCH_LN VERSUS L2

(Detailed in Section 5.3.2)

Two different variables designate measurements of reach length in the NSS-I database: RCH_LN and 1-2. The variable RCH_LN is considered to be a 'map'' attribute, measured during the site selection process on 1:24,000-scale United States Geological Survey (USGS) topographic maps. This variable is a measure of the length of reach between the intended sampling points (originally mapped reach ends). In contrast, the variable 1-2 is a measurement of the distance between the exact sampling points actually visited by field crews. In general, the 1-2 measurement is shorter than RCH_LN, because field crews very seldom were able to visit a reach at the exact spot originally identified. RCH_LN is the variable used to make NSS-I reach length population estimates. RCH_LN is a measure of the length of each reach in the NSS sample. The total reach length represented in the NSS target population can be estimated by summing, for all reaches, the product of RCH_LN and the sample weight, W.

NSS-I population estimates of stream length can be generated in several ways. One method assigns the chemistry of one reach end (upstream or downstream) to the entire reach length. A second method interpolates chemistry along a reach based on the chemistry of both reach ends. Note: interpolated length estimates can be made only by using the observations of reaches from which water samples were collected at <u>both</u> upstream and downstream ends. It is important to pay attention to drop code values when subsetting data for making population estimates of reach length. Drop code (DRPCDE) < 1 identifies the set of observations for which both the upstream and downstream reach ends were sampled, in contrast to DRPCDE < 1, which identifies estimates for all target observations, regardless of whether both reach ends were sampled.

A.8 TOPOGRAPHIC DRAINAGE AREA MEASUREMENTS: a_1, a_2, a_3, a_4, a_5, a_{total} (A1, A2, A3, A4, A5, AND A_WS)

(Detailed In Kaulmann et al., 1988; Section 2.4.1)

There are several different measurements of topographic drainage area in the NSS4 database. Each reflects a different component of reach topographic drainage area. The first, Al, is a measurement of the direct drainage area contributing to the length of reach designated by the mapped location of reach ends (upstream and downstream ends) identified during the site selection process. This measurement is based on the location of reach ends, as identified on 1 :250,000-scale maps during the site selection process. These locations were then transferred to 1:24,000-scale maps on which the drainage areas were measured.

The topographic drainage areas of NSS upstream ends are contained in the variables A2 and A3. For nonheadwater reaches, A2 is a measure of the drainage area contributing flow to the upstream reach end (measured on 1 :24,000-scale maps). This value will be zero for headwater reaches (those observations with RCH_HW=l). For headwater reaches, A3 is a measure of the drainage area contributing flow to the upstream reach end. This value will be zero for nonheadwater reaches (i.e., those with RCH_HW > 1).

The variable A_WS is the total topographic drainage area contributing to stream flow at the mapped sampling point (reach end) and is the drainage area variable used in making NSS-l population estimates of watershed area A_WS is the total drainage area contributing to a reach end. As with Al, A_WS is based on the 'map" location of reach ends identified during the site selection process. Whereas the variable A1 is a measure of the direct drainage area to the intended downstream point on a reach, the variable A_WS is equal to the sum of A1 and A2 for nonheadwater reaches. A3 is incorporated into A1 for headwater reaches. Thus, for headwaters, A_WS = A1. For downstream observations (where NODE=L), A_WS is equal to the drainage of the entire reach, calculated as:

Note: In contrast to Al, a measure of the direct drainage area of a mapped reach (as identified on 1:250,000-scale maps), A4 is a measure of the reach drainage area that drains the area between the exact upstream and downstream field sampling locations. In turn, the variable AS is a measure of the topographic drainage area that contributes to the exact upstream field sampling location (i.e., the drainage area above the upstream sampling point).

Total drainage areas, based on the exact locations where water samples were collected (analogous to A~WS), can be calculated as the sum of the variables A4 and AS, and are usually slightly different than the A_WS measurements. This type of drainage area measurement may be useful for data analyses that require a more precise association between water chemistry and drainage area (e.g., examination of conductivity versus drainage area).

The units for the variables A1, A1 PRIME, A2, and A3 are mi^2, whereas the units for the variables A4, A5, and A_WS are km^2. The conversion factor of 2.59 was used to convert mi^2 to km^2. NSS population estimates of watershed area were calculated in km^2.

A.9 SYNTHESIZED DATA FOR THE SOUTHERN BLUE RIDGE

(Detailed In Section 3.5.2)

Each observation in NSSIDS4 represents an actual water sample collected during the NSS-l or Pilot Survey. In the Southern Blue Ridge Province (identified as subregion ~S in NSSID84), only 20 of the 54 probability reaches were sampled at both upstream and downstream sampling locations during the spring base flow period. In order to generate population estimates compatible with other NSS-I subregions (in which equal numbers of upstream and downstream sites were sampled), a supplemental data set was synthesized for 22 chemical parameters for the 34 upstream sites that were not sampled. This information was generated from regression relationships among sampled streams and from data collected during a summer sampling period. The rationale and equations used to generate the data are presented in Kaufmann et al. (1988), Appendix B. This synthesized information is essential for replicating reported NSS-l population estimates of upstream nodes in the Southern Blue Ridge Province. To use these data, data set SBRSYN should be appended to NSSIDS4. Caution should be used. however. in applving this information to trend analvsis or multivariate examination of data. since the data were generated from repression relationships of data.

A.l0 DIFFERENCES BETWEEN THE NSS-l AND THE NSS-l PILOT SURVEY

(Detailed In Kaufmann et 81., 1988; Section 3.9)

Although most of the sampling and analysis methods used in the NSS-l main survey were developed during the Pilot Survey, conducted a year earlier, some specific differences in the methods used in the NSS-l should be noted, for example, phosphorus analysis, sample holding times, methods of fractionation, and determination of aluminum species.

A.11 REVISIONS TO NSS-I PILOT SURVEY DATA

(Detailed In Appendix C)

After the release of the NSS-I Pilot Survey report (Messer et al., 1988), revisions were made to a small number of variables in the database. These revisions are discussed in detail in Appendix C. The principal updates include the revision of one direct drainage measurements (Al) and revised ANC and BNC measurements based on the refinement in the calculation algorithm. The one Al change is important because Al is used in calculating the final sample weight for each stream. The estimated total number of target reaches in the Southern Blue Ridge Province changed by 10 reaches. Revisions to ANC and BNC were made on all observations in the NSS-I Pilot Survey database. The original Pilot Survey data, as presented in the Pilot Survey report, are contained in PILOTDS3 and PILOTDS4. The revised data are included in NSSIDS4 as data for subregion 2As. the Southern Blue Ridge Province.

A.12 COMPARISON OF PARAMETER UNITS IN THE NSS-I AND THE NLS

In the development of the databases associated with the National Surface Water Survey, there was a conscious effort to maintain consistency in database content and variable format. There are, however, some differences in the units maintained in the ELS and WLS databases and the NSS-I. The following is a brief comparison of principal parameters measured in the NSWS. Units of calculated parameters are maintained in the units of their component variables (e.g., the unit for sum of anions is μeq L^{-1}).

	NSS-I		NLS	
Parameter(s)	Units Measured	Units Converted[a]	Units Measured	Units Converted[a]
Anions/Cations	mg L^{-1}	μeq L^{-1}	mg L^{-1}	μeq L^{-1}
ANC/BNC	μeq L^{-1}	-------	μeq L^{1}	-----
Conductivity	μS cm-1	------	μS cm^{-1}	-----
$A1^{2}$+, Fe, Mn, P	mg L^{-1}	μmol L-1	μg L^{-1}	-----
SiO_2 mg L1	mg L^{-1}	μmol L^{1}	mg L^{1}	-----
Watershed Area	km^2 or mi^2	----	ha	-----
DOC, DIC	mg L^{-1}	-----	mg L^{-1}	-----

a ----- indicates parameters maintained in originally measured units.

In the NSS-I, parameters calculated from pH and DIC (e.g., HCO; CO_3^{2-}, H^+, OH^-, A^- are based on the variables PHSTVL and DICVAL (measured at the analytical laboratory on closed-headspace samples). In contrast, in the NLS, these parameters are calculated from the variables PHAC11 and DICI11 (measured at the analytical laboratory). Section A.16 details the reasoning for using these specific pH and DIC measurements.

A.13 REVISIONS TO a_1

(Detailed In Appendix 0)

After the release of the NSS-I final report (Kaufmann et al., 1988), several direct drainage measurements (a_1) were revised. Any changes in a_1 are important because they are used in generating each reach's final sample weight. If a substantial weight change occurs, a subsequent change will occur in the estimated total number of sites, which in turn may alter the reported estimated number of reaches within a specific chemical range. Examination of the impact of revising the a_1 and sample weights indicates that a significant change in population distributions does not occur and therefore the conclusions presented in the NSS-I final report (see Appendix E) do not change. The distributed NSS database includes the data necessary to replicate any estimates presented in the NSS-I report, as well as the revised a_1 measurements. The variable Al contains the a_1 measurement used to generate population estimates, whereas the variable Al PRIME contains the revised a_1 values. Note: if a revision to a_1 did not occur, then Al = Al PRIME.

A.14 NSS DATABASE VARIABLE FORMATS

All NSS database values were recorded in the format in which they were reported by the analytical laboratory, processing laboratory, or field sampling crew. Table B-I in Appendix B shows the SAS variable formats used in printouts of data listings. For both calculated and reported chemical values, printed decimal places do not necessarily represent actual precision. Results of NSS-I chemistry precision assessments are reported in Cougan et al. (1988) and in Section 4 of Kaufmann et a. (1988). In general, these reporting formats maintain a decimal place beyond the intended level of interest for the data. No data values were rounded off before or during data entry. Variations in reporting precision are most likely to result from the format in which values were originally recorded by the analytical laboratory, processing laboratory, or field sampling crews. Values are stored in SAS data sets using floating point precision. Calculated variables, including those with a “I6" suffix (e.g., 80416), are not rounded off in the database, but maintain the same reporting format as the variables that were used to calculate them (e.g., SO411).

A.15 SUBREGION IDENTIFICATION CODES

The variables RCH_ID and STRM_ID uniquely identify each sampling location. Both variables contain concatenated information for reach subregion and a 1 :250,000-scale map identification code (from maps used in the site selection process). Note: The first two characters of these variables, which indicate general mapped subregion areas, may not accurately identify the appropriate subregion for a given reach.

The variable SUB_ID refers to specific subregions (e.g., 1 D, 2BN, 2)(, etc.) for which population estimates are made. For example, reaches that have SUB_ID = °2X" include those that may have the first two characters of RCH_ID or STRM_ID equal to "2A", °2B", or "2C". This group of reaches is from a combined area of the Southern Appalachians, the southern area of the Valley and Ridge Province, and a northern portion of the Blue Ridge Mountains. Refer to Figure 2-1 for the specific geographic areas within each subregion. It is also important to note that the NSS is not based on the same region and subregion numbering scheme as the NLS. For example, subregion "2D" in the NSS is in the Ozarks and Quachitas of Arkansas and Oklahoma, while subregion '"2D" in the NLS is in the Upper Peninsula of Michigan, Wisconsin, and Minnesota.

Streams in the following subregions may have different values for the SUBJD and STRM_ID variable prefix:

SUB_ID Value	STRM_ID/RCH_ID Prefix	Geographic Area
2X	2A	Northern portion of the Blue Ridge
2X	2B	Southern Appalachians
2X	2C	Southern portion of Valley and Ridge
2As	2A	NSS Pilot Survey area sampled in 1985. Observations for these sites also have a shorter STRMJD and RCH_ID than do those from the full-scale survey
2BN	2B	Northern portion of the Valley and Ridge Province
2CN	2C	Northern Appalachians

A.16 USING DlC AND pH IN CALCULATED VARIABLES

Several calculated variables in the NSS database(s) incorporate measurements of dissolved inorganic carbon (DIC) and pH (e.g., HC0316, C0316, H16, 0H16, ANSUM, and CATSUM). In the NSS, as with other NSWS surveys, these parameters were originally calculated from measurements made at the analytical laboratory (i.e., database variables DICI11 and PHAC11). These specific variables were

used because they were considered to represent the chemistry of the

cubitainer water sample (from which other parameters are measured such as anions and cations). Because NSS water samples were often over-saturated with respect to $C0_2$, it is important that both pH and DIC were measured at the same time so that any changes in DIC are reflected in measured pH.

After the release of the NSS-l report, re-examination of calculated variables (e.g., HCO31 8, C0316, etc.) indicated that DC (DICI11) may not always have been measured close enough in time to pH (PHAC11) to avoid $C0_2$ concentration changes from degassing between the two measurements. Examination of verification flags identified approximately 25% of NSS-l routine water samples as not having initial DIC measurements made within the recommended 14-day holding time, while only about 1% of the corresponding pH measurements were not measured within the recommended 14-day holding time. Based on this information, it is recommended that calculated variables which incorporate DIC or pH measurements use those made at the processing laboratory (i.e., database variables DICVAL and PHSTVL). These (closed~system) syringe sample measurements minimize any $C0_2$ degassing prior to pH or DIC measurement. In addition, these samples were ~processed' within 24-36 hours after the sample was collected from a stream. Because DICVAL and PHSTVL are considered to provide a “matched” set of DIC and pH measurements, they were used to calculate components of the carbonate buffering system (HCO316, C0316, OH1 6, and H16). Please note, however, that these measurements were <u>not</u> made on the same cubitainer subsample as the anions and cations.

The variables HC0316, C0316, 0H16, and H16 in the NSS-l database have been revised accordingly (using DICVAL and PHSTVL) and will therefore be slightly different than those presented in the NSS-l report volumes.

APPENDIX B

NSS-I DATA DICTIONARY

The following data dictionary describes the contents of the U.S. Environmental Protection Agency's (EPA) National Stream Survey (NSS) Pilot and Phase-I database. This dictionary is provided to aid data managers, programmers, and users of the NSS database in the accurate transfer and use of the NSS data on their own computer systems. EPA methods referenced in this appendix are taken from U.S. EPA (1983); USGS methods are from Skougstad et al. (1979). When appropriate, the method or equation used to generate each variable is discussed or referenced. Additional method descriptions are discussed in Sections 2 and 3 of Kaufmann et al. (1988).

Detailed protocol descriptions are presented in the following documents:

Field Sampling Protocols
- Pilot Survey - Knapp et al. (1987)
- NSS-l - Hag ley et al. (1988)

Processing Laboratory Protocols
- Pilot Survey - Knapp et al. (1987)
- NSS-l - Arent et al. (1988)

Analytical Laboratory Protocols
- Pilot Survey - Drousé et al. (1986,1987)
- NSS-l - Hillman et al. (1987); Cougan et al. (1988)

TABLE B-1. NSS-I DATA DICTIONARY

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
ACCO11	μeq L^{-1}	Num	9.3	Acidity, or base neutralizing capacity, measured at the analytical laboratory; a measure of the amount of base needed to neutralize carbonate species, hydronium, and other acids in the sample. Determined by Gran analysis of base titration data (Hillman et al., 1987; Kramer, 1984).
ALDSVL	mg L^{-1}	Num	9.3	An estimate of total monomeric aluminum measured by pyrocatechol violet (PCV) colorimetry and automated flow injection analyzer at the processing laboratory, not measured in the Pilot Survey (Hillman et a., 1987).
ALDS16	μmol L^{1}	Num	9.3	ALDS16 = ALDSVL * 1000 /26.982
ALEX11	mg L^{-1}	Num	9.3	Extractable aluminum, an estimate of monomeric aluminumcomplexes (A~+). A filtered unacidified sample wascomplexed with 8-hydroxyquiniline, extracted with methyl-isobutyl ketone (MIBK), and analyzed by Graphite furnace (GF) atomic absorption spectroscopy (MS) at the analytical laboratory (Hillman et al., 1987; EPA method 202.2).
ALEX16	μmol L^{-1}	Num	9.3	ALEX16 = ALEX11 * 1000/26.982
ALINOR	μmol L^{-1}	Num	8.3	Calculated inorganic monomeric aluminum. ALINR = ALEX16 - ALOR16: in NSS-I Pilot Survey (based on MIBK aluminum methods).. ALINOR = ALDSI6 - ALORI6: in NSS-I (based on PCV methods). *Note:* Negative values have been scored to 0.

ALKA11	μeq L^{-1}	Num	9.3	Alkalinity, or acid neutralizing capacity (ANG); a measure of the amount of acid necessary to neutralize the bicarbonate, carbonate, hydroxyl, and other bases in the sample; measured at the analytical laboratory by Gran analysis of acid titration data (Hillman et al., 1987; Kramer, 1984).

TABLE B-1. NSS-I DATA DICTIONARY (continued)

SAS Variable Name	Units/ Format	Suggested Format Type	Width	Variable Definition
ALORVL	mg L^{-1}	Num	9.3	An estimate of nonextractable monomeric (organic) aluminum measured at the processing laboratory using the MIBK method after passing the sample through a strong cation exchange column. Measured only in Phase I samples (not in Pilot Survey) Hillman et al., 1987).
ALOR11	mgL^{-1}	Num	9.3	Nonextractable organic monomeric aluminum measured at the processing laboratory using the PVC method after passing the sample through a strong cation exchange column. Measured only in Pilot Survey (EPA method 202.2).
ALOR16	μmol L^{-1}	Num	9.3	ALOR16 = ALORVL * 1000/26.982: in Phase I. ALOR16 = ALOR11* 1000/26.982: in Pilot Survey.
ALTL11	mg L^{-1}	Num	9.3	Total aluminum measured on an unfiltered, acidified (HNO_3) aliquot at the analytical laboratory after digestion; analyzed by graphite furnace MS (Hillman et al., 1987; EPA method 202.2).
ALTLI 6	μmol L^{-1}	Num	9.3	ALTLI 6 = ALTL11 * 10001/26.982.
ANDEF	μeq L^{1}	Num	9.3	Anion deficit, or total cations (CATSUM) minus total anions (ANSUM).
ANSUM	μeq L^{1}	Num	9.3	Total anions, defined as: ANSUM = HCO316 + CO316 + CL16 + NO316 + SO416 + FTL16 + 0H16
A1	mi^2	Num	7.3	Direct drainage area. The portion of the watershed that drains directly into a reach between the upstream and downstream confluences interpreted from topographic maps. The variable Al is used in calculating samDle weights that are used in making statistical estimates of the NSS-l target population.

Measurements of Al are maintained in their measured units and not converted to metric values for reasons of statistical protocol. Measured on 1:24,000-scale USGS topographic maps.

SAS Variable Name	Units/ Format		Suggested Format Type	Width	Variable Definition
A4	km^2	Num	7.3		The watershed area draining directly to the segment of stream between the upstream and downstream points where water samples were collected. *Note:* while conceptually similar, the variables Al and A4 will not necessarily be equal as field crews were seldom able to collect water samples at the exact locations of the upstream and downstream ends of the mapped reach. Measured on 1:24,000-scale USGS topographic maps.
AS	km^2	Num		7.3	The watershed area draining to the upstream site where water samples were collected.
A_WS	km^2	Num		7.3	The total watershed area contributing to stream flow at a mapped reach end (upstream or downstream). This is the drainage area measurement used in making NSS population estimates. This measurement is based on the location of reach ends, as identified on 1:250,000-scale maps prior to sampling. These locations were then transferred and drainage areas were measured on 1:24,000-scale USGS topographic maps. *Note:* The variable A_WS is not sum of A4 and A5. Topographic drainage area based on the actual locations where water samples were collected, calculated using the variables A4 and A5, are usually slightly different than the A_WS measurements (see Appendix A, Section A.8).
BAT_ID		Char		6	Identification code for a batch (group) of samples assigned at the processing laboratory. The combination of BAT_ID and SAM_ID form a unique sample identifier.

CATSUM	μeq L^{-1}	Num	9.3	Total cations, defined as: CATSUM = CA16 + MG16 + K16 + NA16 + NH416 + H16.

TABLE B-1. NSS-I DATA DICTIONARY (continued)

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
CA11	mgL^{-1}	Num	9.3	Calcium measured at the analytical laboratory on an acidified,filtered aliquot using flame MS or ICPES (EPA method 215.1).
CAI6	μeqL^{-1}	Num	9.3	CA16 = CA11 * 49.90.
CL11	$mg\ L^{-1}$	Num	9.3	Chloride ion measured at the analytical laboratory on a filtered unacidified aliquot using ion chromatography (ASTM, 1984; O'Dell et a., 1984).
CL16	$\mu eq\ L^{-1}$	Num	9.3	CL16 = CL11 * 28.21.
COLVAL	Platinum Cobalt Units	Num	8.0	True color measured after centrifuging at the processing laboratory using a Hach Model CO-i Color Test Kit (EPA method 110.2 modified).
COND11	$\mu S\ cm^{-1}$	Num	9.3	Specific conductance, temperature corrected, measured at the analytical laboratory (EPA method 120.1).
CONIS	$\mu S\ cm^{-1}$	Num	9.3	Specific conductance measured in situ by field crews. Temperature corrected in Phase I, but not corrected in the Pilot Survey (Hagley et a., 1988).
CONVAL	$\mu S\ cm^{-1}$	Num	9.3	Specific conductance, temperature corrected, measured at the processing laboratory. Measured only in Phase I survey.
COUNTY1		Char	20	Name of the county where the sample site and watershed are located. Identified on 1:24,000-scale USGS topographic maps.
COUNTY2		Char	20	Name of second county, if necessary, where the sample watershed is located. Identified on 1:24,000-scale USGS topographic maps.
COUNTY3		Char	20	Name of third county, if necessary, where the sample watershed is located. Identified on

1:24,000-scale USGS topographic maps.

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
COUNTY4		Char	20	Name of fourth county, if necessary, where the sample watershed is located. Identified on 1:24,000-scale USGS topographic maps.
C0316	μeq L^{-1}	Num	9.3	Calculated carbonate concentration C0316 = 60009 * (DICVAL/12011) * ALPHA2 * 33.33 where: ALPHA2 = K1*K2 / ((H**2) + (H * K1) + (K1*1(2)) where: Ki1= 4.463 x 10**-7 K(2 = 4.6881 x 10**11 H = 10 **(-PHSTVL) (See Appendix A, Section A..16)
DATSMP		Num	DATE7	Date of sampling visit by field crew.
DICE11	mg L^{-1}	Num	9.3	Air-equilibrated dissolved inorganic carbon measured at the analytical laboratory using a carbon analyzer on an unfiltered, unacidified aliquot that was purged for 20 minutes with 300 ppm $C0_2$ air (also, see PHEQ11) (EPA method 415.2, modified).
DICI11	mgL^{-1}	Num	9.3	Dissolved inorganic carbon measured at the analytical laboratory using a carbon analyzer on an unfiltered, unacidified aliquot (EPA method 415.2, modified).

TABLE B-1. NSS-I DATA DICTIONARY (continued)

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
DICVAL	mg L^{-1}	Num	9.3	Dissolved inorganic carbon measured at the processing laboratory using a carbon analyzer on a closed-headspace sample collected and transported in a 60-mi syringe (EPA method 415.2, modified).
DO_IS	mg L^{-1}	Num	8.1	Dissolved oxygen measured in situ at the field sampling site (Hagley et al., 1988).
DOC11	mg L^{-1}	Num	9.3	Dissolved organic carbon measured at the analytical laboratory in a filtered, acidified (H_250_4) sample using a carbon analyzer (EPA method 415.2).
DRPCDE		Num	2.0	Numeric code used to indicate whether a specific observation is considered interest or noninterest in the target population of reaches represented in the NSS. This variable was created to allow the selective exclusion of noninterest observations in statistical analyses. These values are encoded as follows:

Drop Code	Exclusion Criteria Description
0	NSS-I target observation for both the upstream and downstream sample site.
1	Alternate node of noninterest site (useful in analyses which require a data set of reaches for which both the upstream and downstream sites were sampled, e.g., estimates of interpolated reach length).
2	Noninterest sites (Kaufmann et al., 1988; Section 3.2). Observation not included in NSS-I target population estimates (because of tidal influence, intermittent flow conditions, high conductivity, etc.)
3	Site is acidic due to acid mine drainage

(Kaufmann et al., 1988; subsection 9.3.1).

4 Pilot Survey nonapring index data (water samples not collected between March 15 and May 15) (Kaufmann et al., 1988; subsection 1.3.2.1).

5 NSS-I special interest sites (Kaufmann et al., 1988; Section 2.6).

13 Combination of DRPCDE 1 and 3 (i.e., an alternate node of a noninterest site and is impacted by acid mine drainage).

Note: Values with DRPCDE values> 1 were excluded when generating estimates of the NSS-I target population presented in the final NSS-I report (Kaufmann et al., 1988).

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
ELEV	m	Num	7.2	Elevation of visited sampling site. Measured on 1:24,000-scale USGS topographic maps.
FE11	mg L^{-1}	Num	9.3	Dissolved iron measured at the analytical laboratory on a filtered, acidified aliquot using flame MS or inductively coupled plasma emission spectroscopy (ICPES) (EPA method 236.1).
FE16	μmol L^{-1}	Num	9.3	FE16 = FEl11* 1000/55.84
FTL11	mg L^{-1}	Num	9.3	Total dissolved fluoride measured at the analytical laboratory in a filtered, unacidified aliquot using an ion-sensitive electrode (ISE) (EPA method 340.2, modified).
FTL16	μeq L^{-1}	Num	9.3	FTL16 = FTL11 * 52.64
GRADE	%	Num	8.2	Reach gradient; based on the difference in elevation and distance between the upstream and downstream sampling sites. From measurements made on 1:24,000-scale USGS topographic maps. GRADE = (upstream sample site elevation (m) - downstream sample site elevation (m)) / sampled length (m) * 100

				Sampled length = L2 * 1000
HC0316	μeq L^{-1}	Num	8.3	Calculated bicarbonate concentration: HC0316 = 61017 * (DICVAL/12011) * ALPHAi1* 16.39 where: ALPHA1 = (H * K1) / ((H**2) + (H * K1) +K1*K2)) where: K1= 4.4463 x 10**-7 K2 = 4.6881 x 10**-11 H = 10**-PHSTVL (See Appendix A, Section A. 16)

SAS Variable Name	Units Format	Type	Suggested Format Width	Variable Definition
H16	μeq L^{1}	Num	9.4	Calculated hydronium concentration. H16 (10**-PHSTVL) * (10**6)
Ku	mg L^{-1}	Num	9.3	Dissolved potassium measured on a filtered, acidified aliquot at the analytical laboratory using flame MS (EPA method 258.1".
K16	μeq L^{1}	Num	9.3	K16 = K11 * 25.57
LABNAM		Char	20	Name of the analytical laboratory that performed the ana!yses. GLOBAL = Global Laboratories NYSDOH = New York State Department of Health
LAT_STD	Decimal Degrees	Num	9.4	Latitude of the visited sample site. Measured from 1:24,000- scale USGS topographic maps.
LON_STD	Decimal Degrees	Num	9.4	Longitude of the visited sample site. Measured from 1:24,000-scale USGS topographic maps.
L2	km	Num	9.3	Length of the reach segment between the visited upstream and downstream field sampling sites (see also RCH_LN). Measured from I :24,000-scale USGS topographic maps. 12 was <u>not</u> used in making NSS population estimates of length.
		MAP1 Char	32	Name of the 1:24,000-scale USGS map showing the sample site and watershed location.
MAP2		Char	32	Name of the second 1:24,000-scale USGS map, if necessary, showing the sample watershed location.
MAP3		Char	32	Name of the third 1:24,000-scale USGS map, if necessary, showing the sample watershed location.
MAP4		Char	32	Name of the fourth 1:24,000-scale USGS map, if necessary, showing the sample watershed location.
MAP5		Char	32	Name of the fifth 1:24,000-scale USGS map, if necessary showing the sample watershed location.

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
MAP6		Char	32	Name of the sixth 1:24,000-scale USGS map, if necessary, showing the sample watershed location.
MG11	mg L^{-1}	Num	9.3	Dissolved magnesium determined on a filtered, acidified aliquot at the analytical laboratory using flame MS or ICPES (EPA method 242.1).
MG16	μeq L^{-1}	Num	9.3	MG16 = MG11 * 82.26
MN11	mg L^{-1}	Num	9.3	Dissolved manganese measured at the analytical laboratory using flame MS or ICPES on a filtered, acidified aliquot (EPA method 243.1).
MN16	μmol L^{-1}	Num	9.3	MN16 = MN11 * 1000 / 54.938.
NA11	mg L^{-1}	Num	9.3	Dissolved sodium measured on a filtered acidified aliquot at the analytical laboratory using flame MS (EPA method 273.1).
NA16	μeq L^{-1}	Num	9.3	NA16 = NA11* 43.50.
NH411	mg L^{-1}	Num	9.3	Ammonium measured at the analytical laboratory in a filtered, acidified (H_2SO~ aliquot using automated colorimetry (phenate) (EPA method 350.1).
NH416	μeq L^{-1}	Num	9.3	NH416 = NH411 * 55.44.
NODE		Char	1.0	The variable used to identify sampling locations or sample sites on a given reach. These locations are identified as either upstream (U) or downstream (L) ends or nodes. The point below the confluence of two reaches defines an upstream node and the point above the confluence of two reaches defines the downstream node.

Note: Water samples were collected from an area of the

reach considered well mixed when sampling below a confluence.

SAS Variable Name	Units! Format	Type	Suggested Format Width	Variable Definition
NOTSAM		Char	30	Reason, when appropriate, why a stream was not sampled. Based on comments and decisions of field sampling crews.
N0311	mg L^{-1}	Num	9.3	Nitrate ion measured at the analytical laboratory on a filtered, unacidified aliquot using ion chromatography (ASTM, 1984; O'Dell et a., 1984).
NO316	μeq L^{-1}	Num	9.3	N0316 = NO311 * 16.13.
OHI 6	μeq L^{-1}	Num	9.3	Calculated Hydroxide ion concentration: 0H16 = (10**(PHSTVL-14)) * (10**6).
ORGION	μeq L^{-1}	Num	9.3	Estimated organic anion concentration (Oliver Model; Oliver, 1983). ORGION = [(10 **(-PK)) * DOC11 * 10] / [(10**(-PK) + (10**-PHAC11)] where PK = 0.96 + 0.9 * PHAC11 - 0.039 * (PHAC11 ** 2)
PH_CLO	pH units	Num	4.2	Measured only in NSS-l Pilot Survey. Sample pH measured at the field sampling site in a closed container with a portable pH meter (Beckman pHI-21) and glass combination electrode (Orion Ross Model 8104) (Knapp et al., 1988). Measured only in Pilot Survey.
PH_R	pH units	Num	4.2	Sample pH measured at the field sampling site in an open container with a portable meter (Beckman pHI-21) and glass combination electrode (Orion Ross Model 8104) (Hagley et al., 1988).
PHAC11	pH units	Num	4.2	Initial pH from the acidity titration prior to the addition of base titrant. Measured at the analytical laboratory

on an unfiltered, unacidified aliquot stirred in a CO_2-free vessel (EPA method 150.1).

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
PHAL11	pH units	Num	4.2	Initial pH from the alkalinity titration prior to the addition of acid titrant. Measured at the analytical laboratory on an unfiltered, unacidified aliquot, in an open (exposed to air) vessel (EPA method 150.1).
PHEQ11	pH units	Num	4.2	Air-equilibrated sample pH measured at the analytical laboratory on an unfiltered, unacidified aliquot aerated with 300 ppm $C0_2$ air for 20 minutes (EPA method 150.1).
PHSTVL	pH units	Num	4.2	Sample pH measured at the processing laboratory with a portable pH meter (Beckman pHl-21) and glass combination electrode (Orion Ross Model 8104) on a closed-headspace sample. Variable used to make NSS population estimates of pH (EPA method 150.1).
PTD11	mg L^{-1}	Num	9.3	Total dissolved phosphorus measured at the analytical laboratory using automated colorimetric (phosphomolybdate) methods. Determined on a filtered aliquot (measured only in Phase I survey) (USGS 1A600-78, modified).
PTD16	μmol L^{-1}	Num	9.3	PTD16 = PTD11 * 1000 / 30.974
PTL11	mg L^{-1}	Num	9.3	Total phosphorus measured at the analytical laboratory using automated colorimetric (phosphomolybdate

					methods). Determined on an unfiltered aliquot (measured only in the Pilot Survey) (USGS 1-4600-78, modified).
PTL16	μmol L^{-1}		Num	9.3	PTL16 = PTL11 *1000 / 30.974
QUAD		Char	30		Name of the 1:250,000-scale USGS topographic map on which the sample site appears.
RCH_HW		Num	2.0		Number of headwater reaches upstream of the mapped sample location on 1:250,000-scale maps (Shreve order from 1:250,000-scale maps).

SAS Variable Name	Units! Format	Suggested Format Type	Width	Variable Definition
RCH_ID		Char	8	Reach identification code (e.g., 3B041016). An 8-digit code containing three fields that indicate the (1) NSS-l subregion (3B), the 1 :250,000-scale map ID (041), special interest site designation (0 = Routine sites, 9 = Special interest sites), and the gnd~ot number (16). *Note:* The first two characters of this variable, which indicate mapped subregion areas, mav not accurately identify the appropriate NSS-I subregion for a given reach (see Appendix A, Section A.15).
RCH_LN	km	Num	8.3	Length of the mapped reach segment between the upstream and downstream ends (associated with site selection process) identified on 1:250,000-scale maps, but measured on 1 :24,OOO~scale USGS topographic maps. Variable used in making NSS population estimates of reach length. *Note:* This value is the mapped length of the reach as originally mapped and not a measure of the distance between visited sampling points. It is not unusual for this value to be different from 12, the distance between visited sampling points.
SAM_ID		Char	3.0	Sample identification code within a batch, assigned at the processing laboratory. The combination of BAT_ID and SAM_ID form a unique sample identifier.
SAMCOD		Char	3.0	Sample type code, or combination of codes, used to identify the sample as follows (note that precipitation event codes, E, are always listed first, if present): D = Duplicate sample DA = Routine/Duplicate average E = Suspected precipitation event influence (episode)

EDA = RoutinelDuplicate average influenced by
precipitation event
ER = Routine sample influenced by precipitation event
R = Routine
SY = Synthesized chemistry values

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
SAMRN		Num	1.0	Sample visit number identifying the visit number to the sample site.
SHRE75		Num	3.0	Shreve stream order measured from 1 :24_1 000-scale maps.
SI0211	mg L^{-1}	Num	9.3	Silica measured at the analytical laboratory using automated colorimetry (molybdate blue) (USGS 1-2700-78).
SIO216	μmol L^{-1}	Num	9.3	SIO216 = SIO211 * 1000 / 60.084
SIT_CLS		Char	6.0	Reach or observation noninterest code. Used to distinguish impacts that result in a sampling location being considered noninterest or nontarget. A1 = Acid mine drainage affects upper node. A2 = Acid mine drainage affects lower node. A3 = Acid mine drainage affects both nodes. C1 = Conductance greater than 500 μSIcm at upper node C2 = Conductance greater than 500 μS/cm at lower node C3 = Conductance greater than 500 ~S/cm at both nodes. I 1 = Intermittent flow at upper node. l 2 = Intermittent flow at lower node. L3 = Intermittent flow at both nodes. 03 = No evident channel at both nodes (swamp$_1$ or lake). R = Random sample miss; no alternate node, reach dropped. S1 = Special case at upper node. S2 = Special case at lower node. S3 = Special case at both nodes. T3 = Both nodes tidally influenced. *Note:* A given record may have multiple values. When multiple noninterest conditions occurred on a reach, these codes are combined, (i.e., high conductivities found at both sample sites in

addition to acid mine drainage evidence at a downstream end would be classified as "A2C3").

SOBC	μeq L^{-1}	Num	9.3		Sum of base cations. SOBC = NA16 + K16 + MG16 + CA16

)

SAS Variable Name	Unitsl Format	Type	Suggested Format Width	Variable Definition
SO411	mg L^{-1}	Num	9.3	Sulfate measured at the analytical laboratory using ion chromatography (ASTM, 1984; O'Dell et a., 1984).
SO416	μeq L^{-1}	Num	9.3	SO416 = SO411 * 20.82
STATE I		Char	2	Two-letter postal abbreviation for the state in which the sample site and watershed are found. Identified on 1:24,000-scale USGS topographic maps.
STATE2		Char	2	Two-letter postal abbreviation for a second state, if any, in which the sample watershed is found. Identified on 1:24,000- scale USGS topographic maps.
STRATUM		Num	1.0	Statistical stratum (from the site selection process) for the reach (Section 2.3.2). 1 = Regular 2 = Low ANC 3 = Small a_1
STRA75		Num	3.0	Strahler order measured from 1:24,000-scale maps (Strahler, 1957).
STRA250		Num	2.0	Strahler order measured from 1:250,000-scale maps (Strahler, 1957). Measured only in Phase-I Survey.
STRM_ID		Char	9.0	Site identification code concatenates RCH_ID and NODE. *Note:* The first two characters of this variable,

				which indicate mapped subregion areas, may not accurately identify the appropriate NSS-I subregion for a given reach (see Appendix A, Section A.15).
STRMDP	m	Num	3.1	Representative depth at field sampling site.
STRMNAM		Char	30	Stream name.

SAS Variable Name	Units/ Format	Type	Suggested Format Width	Variable Definition
STRMWD	m	Num	4.1	Representative stream width at field sampling site.
SUB_ID		Char	3.0	Subregion identification code (See Appendix A, Section A.15): 1 D = Poconos/Catskills 2CN = Northern Appalachians 2BN = Valley and Ridge 3B = Mid-Atlantic Coastal Plain = Southern Blue Ridge 3A = Piedmont 2X = Southern Appalachians 2D = OzarkslOuachitas 3C = Florida SI = Special Interest Sites
TIMSMP	HH:MM	Num	TIMES	Time at which water samples were collected.
TMPSTR	^{0}C	Num	9.3	Stream temperature measured in situ at time of sampling.
TURVAL	NTU	Num	9.3	Turbidity measured at the processing laboratory using a nephelometer (EPA method 180.1).
W		Num	12.6	The statistical weighting factor used iri making population estimates calculated as: W = WC * (64/Max(A1,0.2)).
WC		Num	12.6	Conditional stage II sampling weight.

Table B-2. NSS-l DATABASE FLAGS

FLAGS GENERATED VIA ANION/CATION BALANCE EVALUATION:

AO Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to unknown cause.

Al Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to unmeasured anions/cations not considered in %IBD calculation.

A2 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to anion (flag suspect anion) contamination.

A3 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to cation contamination.

A4 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to unmeasured organic protolytes (fits Oliver Model).

A5 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to possible analytical error - anion concentration too high (flag suspect anion).

A6 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to possible analytical error - cation concentration too low (flag suspect cation).

A7 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to possible analytical error - anion concentration too low (flag suspect anion).

A8 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to possible analytical error - cation concentration too high (flag suspect cation).

A9 Anion/Cation % Ion Balance Difference (%IBD) was outside criteria due to possible analytical error - alkalinity (ANC) measurement.

FLAGS GENERATED VIA QC BLANK SAMPLE EVALUATION:

B0 External (field) blank is above expected criteria for pH, DIC, DOC, specific conductance, ANC, and BNC determinations.

B1 Internal (lab) blank is >2 x CRDL for DIC, DOC, and specific conductance determinations.

B2 External (field) blank is above expected criteria and contributed >20% to sample concentrations. CThis flag is not used for pH, DIC, DOC, specific conductance, ANC, and BNC determinations.)

B3 Internal (lab) blank is >2 x CRDL and contributes >10% to the sample concentrations. This flag is not used for DIC, DOC, and specific conductance determinations.)

B4 Potential negative sample bias based on internal (laboratory) blank data.

B5 Potential negative sample bias based on external (field) blank data.

FLAGS GENERATED VIA CONDUCTANCE BALANCE EVALUATION:

C0 % Conductance Difference (%CD) was outside criteria due to unknown cause.

Cl % Conductance Difference (%CD) was outside criteria due to possible analytical error - anion concentration too high (flag suspect anion).

C2 % Conductance Difference (%CD) was outside criteria due to anion contamination.

C3 % Conductance Difference (%CD) was outside criteria due to cation contamination.

C4 % Conductance Difference (%CD) was outside criteria due to unmeasured organic ions (fits Oliver Model).

C5 % Conductance Difference (%CD) was outside criteria due to possible analytical error in specific conductance measurement.

C6 % Conductance Difference (%CD) was outside criteria due to possible analytical error - anion concentration too low (flag suspect anion).

C7 % Conductance Difference (%CD) was outside criteria due to unmeasured anions/cations (other anions/cations not considered in %CD calculation).

C8 % Conductance Difference (%CD) was outside criteria due to possible analytical error - cation concentration too low (flag suspect cation).

FLAGS GENERATED VIA DUPLICATE PRECISION EVALUATION:

C9 % Conductance Difference (%CD) was outside criteria due to possible analytical error - cation concentration too high (flag suspect cation).

D2 External (field) duplicate precision exceeded the maximum expected % Relative Standard Deviation (%RSD), and both the routine and duplicate sample concentrations were greater than 10 times the Contract Required Detection Limit (CROL).

D3 Internal (lab) duplicate precision exceeded the maximum contract required % Relative Standard Deviation (%RSD), and both the routine and duplicate sample concentrations were >10 x Contract Required Detection Limit (CRDL).

FLAGS THAT IDENTIFY SUSPECT FIELD DATA:

F0 % Conductance difference (%CD) exceeded criteria when in situ field conductance value was replaced.

Fl Hillman/Kramer protolyte analysis program indicated field pH problem when stream site pH value was replaced.

F2 Hillman/Kramer protolyte analysis program indicated unexplained problem with stream site pH or processing laboratory DIC values when stream site pH value was replaced.

F3 Hillman/Kramer protolyte analysis program indicated field problem - processing laboratory pH.

F4 Hillman/Kramer protolyte analysis program indicated field problem - processing laboratory DIC.

F5 Hillman/Kramer protolyte analysis program indicated unexplained problem with processing laboratory pH or DIC values when processing laboratory pH value was replaced.

F6 % Conductance Difference (%CD) exceeded criteria when processing laboratory (trailer) specific conductance value was replaced.

FLAGS GENERATED VIA HOLDING TIME EVALUATION:

H0 The maximum holding time criteria were not met.

No H1 'Date Analyzed' data were submitted for reanalysis data.

FLAG GENERATED VIA DETECTION LIMIT EVALUATION

L1 Instrumental Detection Limit (IDL) exceeded Contract Required Detection Limit (CRDL) and sample concentration was <10 x IDL

FLAG GENERATED VIA CONTRACT SPECIFICATION ASSESSMENTS:

M0 Value obtained using a method that was outside criteria as specified by contract.

TABLE B-2. NSS-I DATABASE FLAGS (continued)

FLAGS GENERATED VIA OC AUDIT SAMPLE EVALUATION:

N0 Audit sample value exceeded upper control limit.

N1 Audit sample value was below the control limit.

FLAGS GENERATED VIA PROTOLYTE EVALUATION:

P0 Lab problem - initial pH from alkalinity (ANC) titration.

P1 Lab problem - initial pH from acidity (<ANC) titration.

P2 Lab problem - unexplained - initial pH from ANC or BNC titration.

P3 Lab problem - initial DIC determination.

P4 Lab problem - air-equilibrated pH or DIC determinations.

P5 Lab problem - unexplained - initial pH from ANC or BNC titrations or initial DIC determinations.

P6 Lab problem - alkalinity (ANC>) determination.

P7 Lab problem - C02-Acidity (BNC) determination.

FLAGS GENERATED VIA QC CALIBRATION REFERENCE SOLUTIONS:

Q1 Quality Control Check Sample (QCCS) was above contractual criteria.

Q2 Quality Control Check Sample (QCCS) was below contractual criteria.

Q3 Insufficient number of QCCS were measured.

Q4 No Quality Control Check Sample (QCCS) was analyzed.

Q5 Detection Limit QCCS was not 2 to 3 times Contract Required Detection Limit (CRDL) and measured value was not within 20% of the theoretical concentration.

TABLE B-2. NSS-l DATABASE FLAGS (continued)

FLAGS GENERATED VIA DATA VALIDATION:

Ul Value considered error. Substitution made with observation from same stream.

U2 Value considered error. Substitution made based on known relationships with other variables.

WI Unusual value in context of subregion chemistry, but reconcilable based on site chemistry.

W2 Unusual value in context of subregion and on-site chemistry, but not replaced.

FLAGS USED TO IDENTIFY VALUES FOUND DURING VERIFICATION:

X0 Irreconcilable but confirmed based on QA review.

Xl Extractable Al concentration is greater than total Al concentration by 0.010 mg/L where extractable Al> 0.015 mg/l

X2 Irreconcilable but confirmed data - possible aliquot switch.

X3 Irreconcilable but confirmed data - possible gross contamination of aliquot or parameter.

X4 Irreconcilable but confirmed data - possible sample (all aliquots) switch. Values for flags X0 through X4 should not be included in any statistical analysis.

APPENDIX C
PILOT SURVEY REVISIONS

Results of the NSS-l Pilot Survey, conducted in the Southern Blue Ridge (NSS subregion 2~), are presented by Messer et al. (1986,1988). The data are contained in PILOTIDD53 and PlLOTDS4. These results are also presented by Kaufmann et al. (1988), but with some minor changes from the Pilot Survey Report. After publication of the Pilot Survey Report, revisions were made on two variables in the Pilot Survey data, Al and ALKA11. The revised data for the Southern Blue Ridge are included only in Data Set NSSIDS4.

REVISIONS TO Al (a_1)

It is important to note any change to Al, because this variable is used in calculating the final weight of each reach (the number of reaches the observation represents in the target population). During the final assessment of NSS geographic data, the Al estimates for two reaches were revised. The first was stream reach “2A0789l,” a special interest site. The Al value for this reach was revised from 5.15 mi^2 to 5.19 mi^2. Since special interest sites are not used in making statistical estimates of the target population, this update does not affect any estimates of the Southern Blue Ridge reported by Messer et al. (1986). The second Al revision, however, was for stream reach “2A0788l,” a probability sample site. The Al value was revised from 13.34 mi^2 to 6.57 mi^2. In turn, this reach's final sample weight changed by about 10, from 9.60 to *19.48,* resulting in a change in the estimated target population for the Southern Blue Ridge from 2,021 reaches to 2,031 reaches. The standard error estimate also changed from 326.7 to 326.4. This change is minor in terms of the total estimate and does not change any of the conclusions based on the results of the Pilot Survey.

REVISIONS TO ALKALINITY (ALKA11)

During the full-scale NSS-l, the method used to calculate ANC and BNC from Gran titration data was refined. All ANC values for streams in the Southern Blue Ridge were revised. The revisions are summarized as follows. All ANC values for ALKA11 are in μeq L^{-1}.

Mean ANC Change	Standard Deviation	Mean % Change	Standard Deviation	Mean Absolute Change	Standard Deviation	Mean Absolute % Change	Standard Deviation
1.85	31.22	0.9%	6.6%	7.3	30.4	3.0%	5.9%

(Figure C-1). The revised ANC values are, on the whole, slightly greater than the original values. Regression of new values on original values shows the magnitude of difference, with a slope of 1.017 and an r^2 of 0.9908 (Figure C-2a). This small adjustment is minimized even further when these data are considered in the context of the target population estimates, which are normally based on index values, each of which is the mean of three spring measurements (Figure C-2b), with a regression slope of 1.008 and an r^2 of 0.9964. Overall, the revision in ANC calculation had minimal impact on estimates of the distribution of ANC in the target population. Figure C-3 compares the cumulative distribution function (CDF) of the target population for the original Pilot Survey ANC values and the revised values.

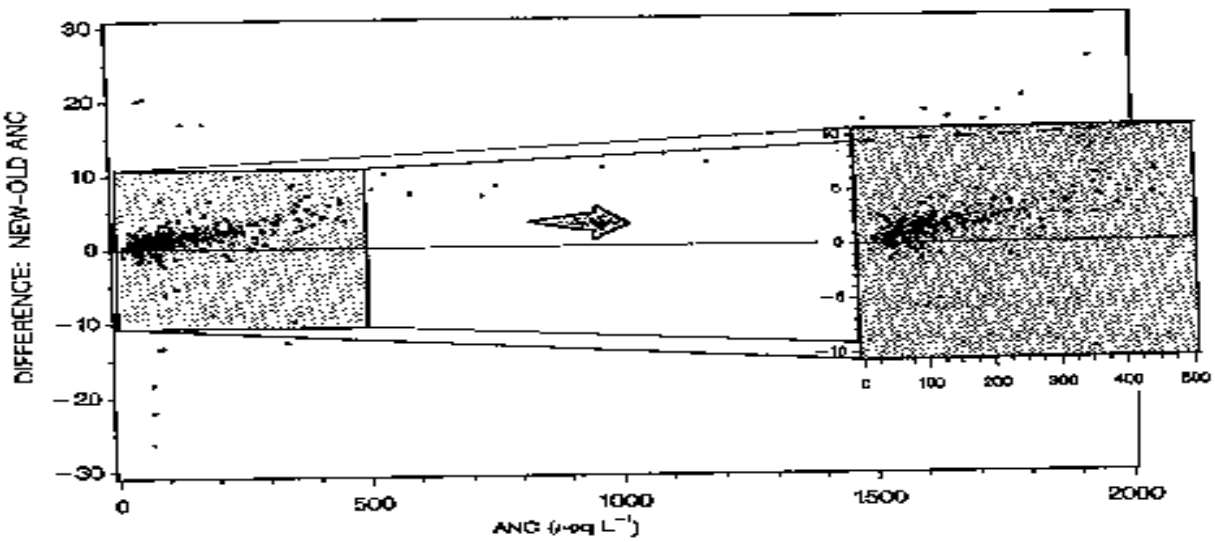


Figure C-1. Difference in original and revised ANC values versus original ANC.

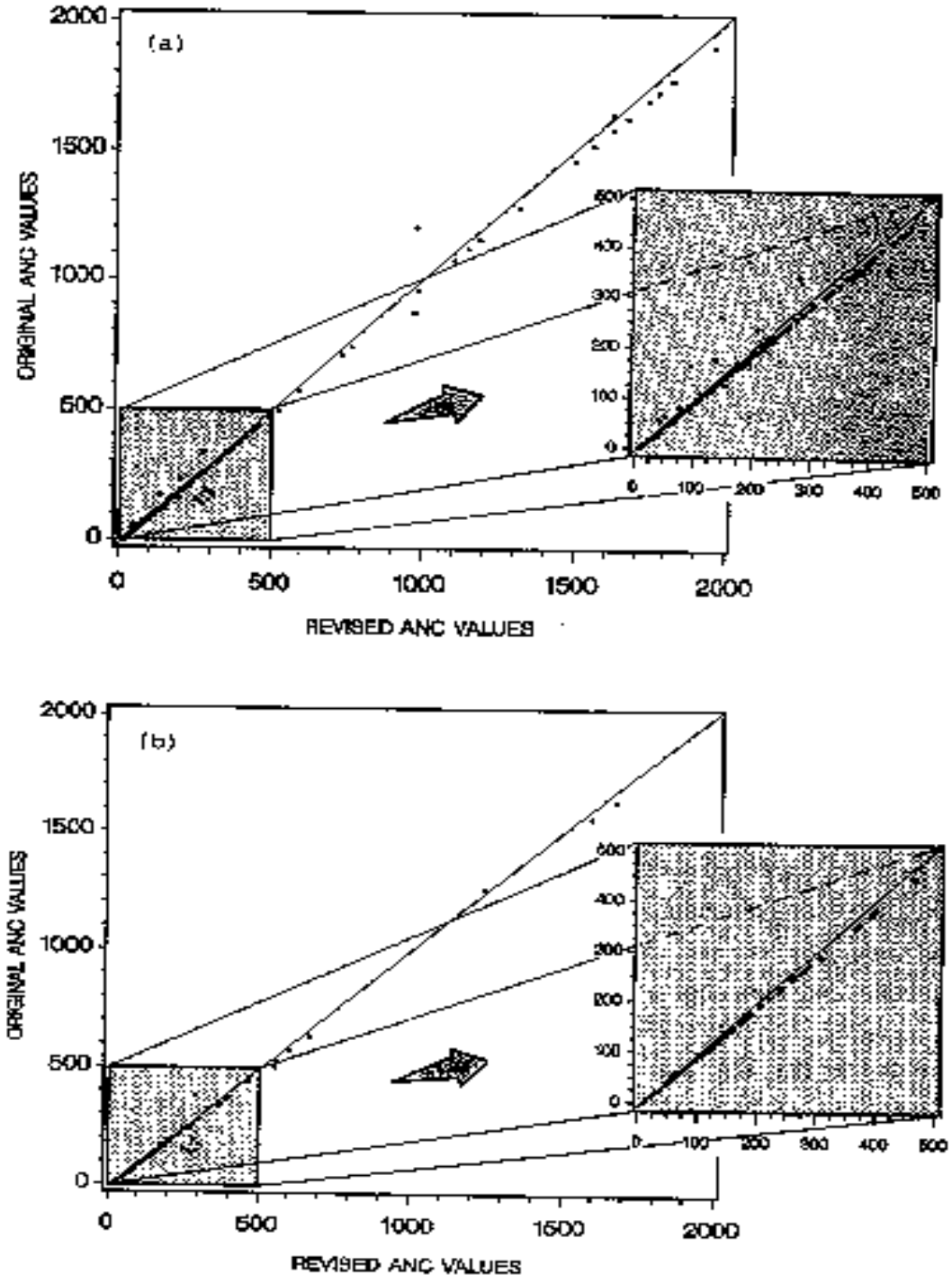
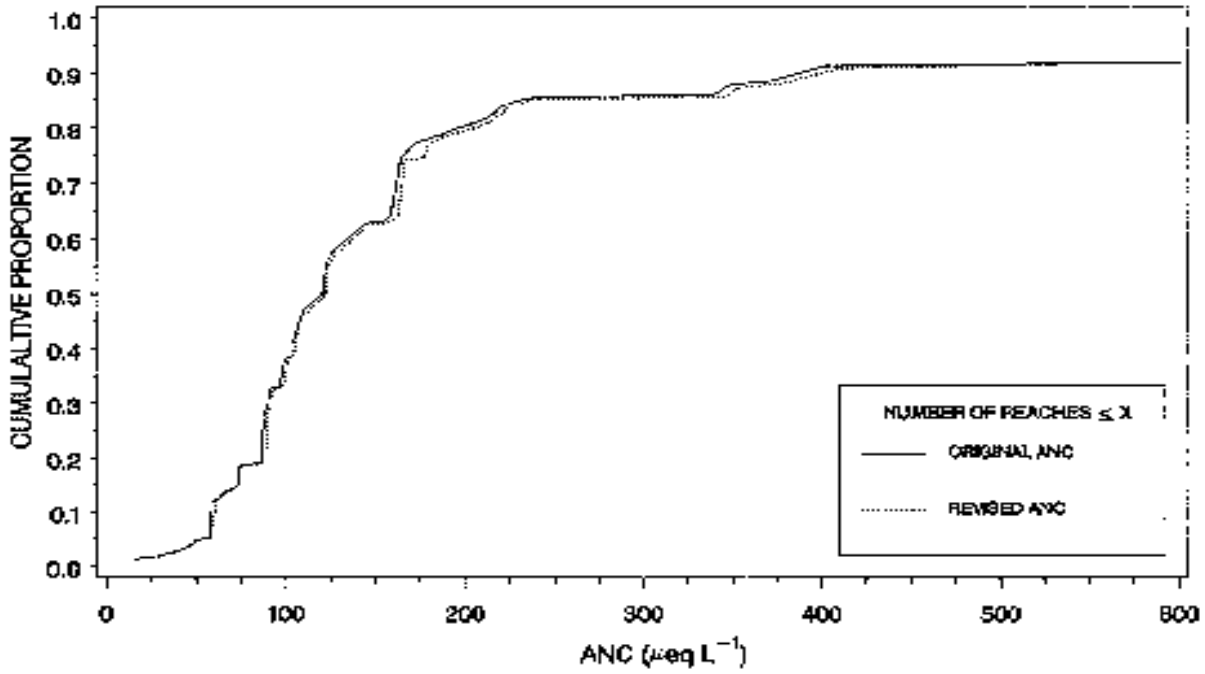


Figure C-2. NSS-I Pilot Survey original versus revised ANC values for (a) all samples, and (b) NSS-I Pilot Survey
original versus revised ANC values using averaged “indexed” observations.



Estimate Based on	Min.	20th %	4Oth %	Median	6Oth %	80th %	Max.	Mean	SD
Original ANC	16.2	86.6	102.6	119.6	134.3	197.7	1710.5	252.02	399.14
Revision ANC	16.5	88.3	104.4	121.6	137.5	206.7	1730.5	257.18	403.10

Figure C-3. ANC distribution for NSS subregion 2As (Southern Blue Ridge Province) target population based on original and revised ANC values.

APPENDIX D
NSS-I FIELD OBSERVATION VARIABLES
Data Set NSSFSO

FIELD SITE OBSERVATIONS

Originally categorized as “Watershed Characteristics”, this information pertains only to the area in the immediate vicinity of the sampling site. Table D-1 is a list of the variables. The field site observation data contain four categories of information:

- Descriptive Information
- Watershed Activities/Disturbances
- Bank Coverage
- Stream substrate

Descriptive Information

This includes basic identification variables and location specific information about the sampling site.

Watershed Activities/Disturbances

Field crews recorded observations of immediate watershed characteristics but did not perform an extensive field reconnaissance of the entire watershed area for disturbances which may influence reach chemistry. These data were collected to document potential watershed influences on sample chemistry and do not identify all potential disturbances in the entire watershed. A standardized format was used to record observations about potential watershed impacts and human disturbances. The approximate distance of disturbance from a sampling location was estimated to the nearest 100 feet (30 meters). Table D-2 summarizes watershed activity and disturbance variables. Although none of this information has undergone the level of quality assurance that was applied to the chemical data, such information is helpful when interpreting individual sample chemistries.

Bank Coverage

Bank coverage estimates were made for the area within 100 meters of the stream bed. This category provides an estimate of vegetation cover and type based on a coarse scale of high, medium, low, and none.

Stream Substrate

Substrate composition estimates were made at the reach sampling location. Particle size categories were based on a scheme suggested by Cummins (1962). This information is based on observations of individual field crews and is on the whole subjective. This information has not been subjected to the stringent quality assurance review that the chemical data have. Observations of substrate are difficult to review, but these data may be useful for examining the specific conditions that existed at the time of sampling.

TABLE D-1. NSS-I FIELD SITE OBSERVATIONS

SAMPLING SITE IDENTIFICATION DATA

Parameter	Variable Name	Parameter	Variable Name	Parameter	Variable Name
Stream ID code	STRM_ID	Stream name	STRMNAM	Elevation(ft)	ELEV
Latitude(ddmmss)	LAT_ST	Longitude(dddmmss)	LON_ST	Date Sampled	DATSMP
Stream depth (m)	STRMDP	1:250,000 Map name	MAP_BIG	Sample Crew ID	CRW_ID
Steam width (m)	STRMWD	1:24,000 Map name	MAP_SML	Country name	COUNTY
Field Comments	COMMO7				

STREAM BANK COVERAGE ESTIMATES WITHIN 100 METERS OF STREAM BED:

A-Absent; S-Sparse (< 25%); M-Moderate (25%-75%); H-Heavy (> 75%)

Parameter	Variable Name	Parameter	Variable Name	Parameter	Variable Name
%Conif. trees	CNTREE	%Deciduous trees	DCE	Grasses/forb	GRASS
%Rock/bare	RCKBR	Shrub cover	SHRUB	%Moss cover	MOSS
%Wetland	WETLND				

STREAM SUBSTRATE COVERAGE ESTIMATES WITHIN 100 METERS OF STREAM BED:

A-Absent; S-Sparse (< 25%); M-Moderate (25%-75%); H-Heavy (> 75%)

Parameter	Variable Name	Parameter	Variable Name	Parameter	Variable Name
%Substrate:aufwchs	AUFS	%Substrate:boulders	BOULD	%Substrate:cobble	COBBLE
%Substrate:gravel	GRAVEL	%Substrate:sand	SAND	%Substrate:silt	SILT

TABLE D-2. WATERSHED ACTIVITY/DISTURBANCE VARIABLE SUMMARY

Parameter	Variable Name	Parameter	Variable Name
Roadways Along Stream			
Distance to unpaved road (m)	UPRD_D	Distance to paved road (m)	PRD_D
Crossings Above Stream			
Distance to bridge (m)	BRDG_D		
Distance to grade (m)	GRAD_D	Distance to culvert (m)	CULV-D
Dwellings			
Distance to multiple dwellings (m)	MDWL_D	Distance to single dwellings (m)	SDWL_D
Agriculture			
Distance to cropland (m)	CROP_D	Distance to pastures (m)	PSTR_D
Distance to unfenced land (m)	UFNC_D	Distance to fenced land (m)	FENC_D
Industry			
Industry 1 (type), Distance (m)	IND1, IND1_D		
Logging			
Presence and Age, Distance (m)	LOG, LOG_D		
Fires			
Presence and Age, Distance (m)	FIRE, FIRE_D		
Mines/Quarries			
Type, Distance (m)	MNQR, MNQR_D		
Impoundments			
Above site (type), Distance (m)	IMPA, IMPA_D	Below site (type, Distance (m)	IMPB, IMPB_D
Livestock			
Livestock (type), Distance from (m)	LIVE, LIVE_D		
Other			
Other disturbances near site, Distance (m)	OTH, OTH_D		

APPENDIX E

CARD IMAGE FORMAT DEFINITION

Only the final NSS-l data sets (Data Set 4 for the Pilot Survey and the full-scale NSS-l Survey) are provided as both SAS4ormatted files and as 80-column ASCII card-image files. This includes data sets NSSIDS4, SBRSYN, and PILOTDS4. The data for NSSIDS4 have been divided into two data sets (NSSIDS4A and NSSIDS4B), so that each will fit on a 1.2 M floppy diskette. Data Set NSSIDS4A contains data for streams in subregions 1D, 2Bn, 2Cn, and 3B. Data Set NSSIDS4B contains data for streams in subregions 2As, 2D, 2X, 3A, and 3C, and for all special interest sites. The formats for all four data sets are listed in Tables F-1, F-2, and F3. Numeric variables were transferred to the card-image files using the suggested variable width listed in Table B-1. Dates are in DDMMMYY format and times are in HH:MM format (24-h clock) for all data card-image data sets. Missing numeric variables are represented as -999. These values should be removed prior to any data analysis.

TABLE E-1. Card-image Format Definition, NSS Data Sets NSSIDS4A and NSSIDS4B

Card #	Variable Name	Type	Variable Format	Start	Column End	Column label	Card #
1	A1	Num	7.3	0	7	DIRECT WATERSHED AREA (SQ MI)	1
1	A1PRIME	Num	7.3	9	16	UPDATED (1989) AI (SQ MI)	1
1	A2	Num	7.3	18	25	WS AREA TO MAPPED UPPER NODE (SQ MI)	1
1	A3	Num	7.3	27	34	WS AREA TO MAPPED HEADWATER (SQ MI)	1
1	A4	Num	7.3	36	43	WS AREA BETWEEN U/L SAMPLE SITE (SQ SQ KM)1	
1	A5	Num	7.3	45	52	WS AREA TO UPPER SAMPLE SITE (SQ (SQ KM)1	
1	ACCO11	Num	9.3	54	63	BASE NEUTRALIZING CAPACITY (UEQ/L)1	
1	ALDS16	Num	9.3	65	74	MONOMERIC CPCV) ALUMINUM (UMOL/L)1	
2	ALEX16	Num	9.3	0	9	EXTRACTABLE (MIBK) ALUMINUM (UMOL/L)2	
2	ALINOR	Num	9.3	11	20	INORG. MONOMERIC ALUMINUM (UMOL/L)2	
2	ALKA11	Num	9.3	22	31	ACID NEUTRALIZING CAPACITY (UEQ/L)2	
2	ALOR16	Num	9.3	33	42	ORG. MONOMERIC CPCV) ALUMINUM (UMOL/L)	
					2		
2	ALTL16	Num	9.3	44	53	TOTAL ALUMINUM (UMOL/L)	2
2	ANDEF	Num	9.3	55	64	ANION DEFICIT, CATSUM-ANSUM (UEQ/L)	2
2	ANSUM	Num	9.3	66	75	SUM OF ANIONS (UEQ/L)	2
3	A_WS	Num	7.3	0	7	WS AREA TO MAPPED NODE (SQ KM)	3
3	CA16	Num	9.3	9	18	CALCIUM (UEQ/L)	3
3	CATSUM	Num	9.3	20	29	SUM OF CATIONS (UEQ/L)	3
3	CL16	Num	9.3	31	40	CHLORIDE (UEQ/IL)	3
3	C0316	Num	9.3	42	51	CARBONATE (UEQ/L)	3
3	COLVAL	Num	8.0	53	61	COLOR VALUE (PCU)	3
3	COND11	Num	9.3	63	72	CONDUCTANCE -ANALYTICAL LAB- (US/CM)3	
4	CONIS	Num	9.3	0	9	IN-SITU CONDUCTANCE (US/CM)	4
4	CONVAL	Num	9.3	11	20	CONDUCTANCE -PROCESS. LAB- (US/CM)4	
4	COUNTYI	Char	15.0	22	37	COUNTY NAME	4
4	COUNTY2	Char	15.0	39	54	COUNTY NAME	4
4	COUNTY3	Char	15.0	56	71	COUNTY NAME	4
5	COUNTY4	Char	15.0	0	15	COUNTY NAME	5
5	DATSMP	Num	7.0	17	24	DATE SAMPLED	5
5	DICE11	Num	9.3	26	35	AIR EQUIL. 015. IMORG. CARBON (MG/L)	5
5	DICI11	Num	9.3	37	46	INITIAL 015. INORGANIC CARBON (MG/L)	5
5	DICVAL	Num	9.3	48	57	DIS. INORG. CARBON -PROCESS.LAB- (MG/L)5	
5	DOC11	Num	9.3	59	68	DIS. ORGANIC CARBON (MG/L)	5
6	DO_ IS	Num	9.3	0	9	IN-SITU DISSOLVED OXYGEN (MG/L)	6
6	DRPCDE	Num	2.0	11	13	SITE EXCLUSION CODE (0,1,2,3,4,5,13)	6
6	ELEV	Num	7.2	15	22	SAMPLE SITE ELEVATION CM)	6
6	FE16	Num	9.3	24	33	IRON (UMOL/L)	6
6	FTLI6	Num	9.3	35	44	TOTAL FLUORIDE (UEQ/L)	6
6	GRADE	Num	8.2	46	54	STREAM REACH GRADIENT CX)	6
6	H16	Num	9.4	56	65	HYDROGEN ION ACTIVITY (UEQ/L)	6
6	HCO316	Num	9.3	67	76	BICARBONATE (UEQ/L)	6
7	K16	Num	9.3	0	9	POTASSIUM (UEQ/L)	7
7	L2	Num	9.3	11	20	LENGTH BETWEEN U/L SAMPLE SITES (KM)	7
7	LASNAM	Char	6.0	22	28	CHEMICAL ANALYSIS LABORATORY NAME7	
7	LAT_STD	Num	9.4	30	39	SAMPLE SITE LATITUDE (DECIMAL FORM)7	
7	LON_STD	Num	9.4	41	50	SAMPLE SITE LONGITUDE (DECIMAL FORM)	7
8	MAP1	Char	32.0	O	32	1:24,000 SCALE MAP NAME	8
8	MAP2	Char	32.0	34	66	1:24,000 SCALE MAP NAME	8
9	MAP3	Char	32.0	O	32	1:24,000 SCALE MAP NAME	9
9	MAP4	Char	32.0	34	66	1:24,000 SCALE MAP NAME	9
10	MAPS	Char	32.0	O	32	1:24,000 SCALE MAP NAME	10
10	MAP6	Char	32.0	34	66	1:24,000 SCALE MAP NAME	10
10	MG16	Num	9.3	68	77	MAGNESIUM (UEQ/L)	10
11	MN16	Num	9.3	O	9	MANGANESE (UMOL/L)	11
11	NAI6	Num	9.3	11	20	SODIUM (UEQ/L)	11
11	NH416	Num	9.3	22	31	AMMONIUM (UEQ/L)	11
11	N0316	Num	9.3	33	42	NITRATE (UEQ/L)	11
11	NODE	Char	1.0	44	45	REACH SAMPLE POSITION (U=UPPER,L=LOWER)11	
11	NOTSAM	Char	30.0	47	77	REASON NOT SAMPLED	11
12	0H16	Num	9.3	O	9	HYDROXIDE (UEQ/L)	12
12	ORGION	Num	9.3	11	20	CALCULATED ORGANIC ANIONS (UEQ/L)12	
12	PHAC11	Num	4.2	22	26	INITIAL PH, ACIDITY TITRATION	12
12	PHAL11	Num	4.2	28	32	INITIAL PH, ALKALINITY TITRATION	12
12	PHEQ11I	Num	4.2	34	38	AIR EQUILIBRATED LAB PH	12
12	PHSTVL	Num	4.2	40	44	CLOSED SYSTEM PH -PROCESS. LAB-	12
12	PH_CLO	Num	4.2	46	50	FIELD PH, CLOSED CONTAINER -PILOT ONLY12	
12	PH_R	Num	4.2	52	56	FIELD PH, OPEN SYSTEM	12
12	PT516	Num	9.3	58	67	TOTAL DISSOLVED PHOSPHOROUS (UMOL/L)	12
12	PTLI6	Num	9.3	69	78	TOTAL PHOSPHOROUS (UMOL/L)	12
13	QUAD	Char	30.0	0	30	1:250,000 SCALE MAP NAME	13
13	RCH_HW	Num	2.0	32	34	SHREVE ORDER -1:250,000 SCALE MAP	13

TABLE E-1. Card–image Format Definition, NSS Data Sets NSSIDS4A and NSSIDS4B (continued)

Card #	Variable Name	Type	Variable Format			Start	Column End	Column label	Card #
13	RCH_ID		Char	8.0		36	44	REACH IDENTIFICATION CODE	13
13	RCH_LN		Num	9.3		46	55	LENGTH OF MAPPED BLUE LINE REACH (KM)	13
13	SAMCOD	Char	3.0		57	60	SAMPLE TYPE (D,DA,E,EDA,ER,NS,SY,R)		13
13	SAMRN		Num	1.0		62	63	SAMPLE VISIT NUMBER (0,1,2,3,4)	13
13	SHRE75		Num	3.0		65	68	SHREVE ORDER -1:24,000 SCALE MAP	13
14	SIO16	Num	9.3		0	9	DISSOLVED SILICA (UMO/IL)		14
14	SIT_CLS		Char	6.0		11	17	SITE CHARACTERISTIC CODE	14
14	S0416	Num	9.3		19	28	SULFATE (UEQ/L)		14
14	SOBC	Num	9.3		30	39	SUM OF BASE CATIOMS (UEQ/L)		14
14	STATEl	Char	2.0		41	43	STATE (TWO CHARACTER CODE)		14
14	STATE2		Char	2.0		45	47	STATE (TWO CHARACTER CODE)	14
14	STRA75		Num	9.3		49	58	STRANLER ORDER -1:24,000 SCALE MAP	14
14	STRA25O	Num	9.3		60	69	STRAHLER ORDER -1:250,000 SCALE MAP		14
14	STRATUM	Num	1.0		71	72	STRATUM (1=REG.,2=LOW ANC,3=SMALL Al)		14
14	STRMDP		Num	3.1		74	77	STREAM DEPTH CM)	14
15	STRMMAM	Char	30.0		0	30	STREAM NAME		15
15	STRMWD	Num	4.1		32	36	STREAM WIDTH (M)		15
15	STRM_ID	Char	9.0		38	47	STREAM/SITE IDENTIFICATION CODE		15
15	SUB_ID		Char	3.0		49	52	SUBREGION IDENTIFICATION CODE	15
15	TIMSMP		Num	5.0		54	59	TIME SAMPLED (HH:MM)	15
15	TMPSTR		Num	9.3		61	70	STREAM TEMPERATURE (DEG C)	15
16	TURVAL		Num	9.3		0	9	TURBIDITY (MTU)	16
16	W	Num	12.6		11	23	REACH WEIGHTING FACTOR		16
16	WC	Num	12.6		25	37	STAGE II CONDITIONAL WEIGHT		16

Card #	Variable Name	Type	Variable Format		Start	Column End	Column label	Card #
1	A1	Num	7.3	0	7	DIRECT WATERSHED AREA (SQ MI)		1
1	A1PRIME	Num	7.3	9	16	UPDATED (1989) AI (SQ MI)		1
1	A2	Num	7.3	18	25	WS AREA TO MAPPED UPPER NODE (SQMI) 1		
1	A3	Num	7.3	27	34	WS AREA TO MAPPED HEADWATER (SQ MI) 1		
1	A4	Num	7.3	36	43	WS AREA BETWEEN U/L SAMPLE SITE (SQ SQKM) 1		
1	A5	Num	7.3	45	52	WS AREA TO UPPER SAMPLE SITE (SQ (SQKM) 1		
1	ALEXl6		Num	9.3	54	63	EXTRACTABLE (MIBK) ALUMINUM (UMOL/L) 1	
1	ALKA11		Num	9.3	65	74	ACID NEUTRALIZING CAPACITY (UEQ/L) 1	
2	ALOR16		Num	9.3	0	9	ORG. MONOMERIC (PCV) ALUMINUM (UMOL/L) 2	
2	ANDEF		Num	9.3	11	20	ANION DEFICIT, CATSUM-ANSUM (UEQ/L) 2	
2	ANSUM		Num	9.3	22	31	SUM OF ANIONS (UEQ/L)	2
2	A_WS		Num	7.3	33	40	WS AREA TO MAPPED NODE (SQ KM) 2	
2	CA16	Num	9.3	42	51	CALCIUM (UEQ/L)		2
2	CATSUM		Num	9.3	53	62	SUM OF CATIONS (UEQ/L)	2
2	CLI6	Num	9.3	64	73	CHLORIDE (UEQ/L)		2
3	CONO11		Num	9.3	0	9	CONDUCTANCE -ANALYTICAL LAB- (US/CM) 3	
3	COUNTY1	Char	15.0	11	26	COUNTY NAME		3
3	DICI11		Num	9.3	28	37	INITIAL DIS. INORGANIC CARBON (MG/L) 3	
3	DOC11		Num	9.3	39	48	DIS. ORGANIC CARBON (MG/L)3	
3	DRPCDE		Num	2.0	50	52	SITE EXCLUSION CODE (0,1,2,3,4,5,13) 3	
3	ELEV	Num	7.2	54	61	SAMPLE SITE ELEVATION (N)		3
3	FE16	Num	9.3	63	72	IRON (UMOL/L)		3
4	FTLI6	Num	9.3	0	9	TOTAL FLUORIDE (UEQ/L)		4
4	GRADE		Num	8.2	11	19	STREAM REACH GRADIENT (X)4	
4	H16	Num	9.4	21	30	HYDROGEN ION ACTIVITY (UEQ/L)		4
4	HC0316		Num	9.3	32	41	BICARBONATE (UEQ/L)	4
4	K16	Num	9.3	43	52	POTASSIUM (UEQ/L)		4
4	L2	Num	9.3	54	63	LENGTH BETWEEN /IL SAMPLE SITES (KM) 4		
4	LAT_STD	Num	9.4	65	74	SAMPLE SITE LATITUDE (DECIMAL FORM)4		
5	LON_STD	Num	9.4	0	9	SAMPLE SITE LONGITUDE (DECIMAL FORM) 5		
5	MAP1	Char	32.0	11	43	1:24,000 SCALE MAP NAME		5
5	MAP2	Char	32.0	45	77	1:24,000 SCALE MAP NAME		5
6	MAP3	Char	32.0	0	32	1:24,000 SCALE MAP NAME		6
6	MAP4	Char	32.0	34	66	1:24,000 SCALE MAP NAME		6
7	MAPS		Char	32.0	0	32	1:24,000 SCALE MAP NAME	7
7	MG16	Num	9.3	34	43	MAGNESIUM (UEQ/L)		7
7	MN16	Num	9.3	45	54	MANGANESE (UMOL/L)		7
7	NA16	Num	9.3	56	65	SODIUM (UEQ/L)		7
7	NH416		Num	9.3	67	76	AMMONIUM (UEQ/L)	7
8	NO316		Num	9.3	0	9	NITRATE (UEQ/L)	8
8	NODE		Char	1.0	11	12	REACH SAMPLE POSITION (U=UPPER,L=LOWER) 8	
8	PHSTVL		Num	4.2	14	18	CLOSED SYSTEM PH -PROCESS. LAB- 8	
8	PTL16		Num	9.3	20	29	TOTAL PHOSPHOROUS (UMOL/L)8	
8	QUAD		Char	30.0	31	61	1:250,000 SCALE MAP NAME	8
8	RCH_HW	Num	2.0	63	65	SHREVE ORDER -1:250,000 SCALE MAP8		
8	RCH_ID		Char.	8.0	67	75	REACH IDENTIFICATION CODE8	
9	RCH_LN		Num	9.3	0	9	LENGTH OF MAPPED BLUE LINE	

						REACH (KM)		9	
9	SANCOD	Char	3.0		11	14	SAMPLE TYPE (D,DA,E,EDA,ER,NS,SY,R)9		
9	SAMRN		Num	1.0		16	17	SAMPLE VISIT NUMBER (0,1,2,3,4)9	
9	SHRE75		Num	3.0		19	22	SHREVE ORDER -1:24,000 SCALE MAP	9
9	SIO216		Num	9.3		24	33	DISSOLVED SILICA (UMOL/L)	9
9	SO416		Num	9.3		35	44	SULFATE (UEQ/L)	9
9	SOBC		Num	9.3		46	55	SUM OF BASE CATIONS (UEQ/L)9	
9	STATE1		Char	2.0		57	59	STATE (TWO CHARACTER CODE)9	
9	STRA75		Num	9.3		61	70	STRAHLER ORDER -1:24,000 SCALE MAP	9
9	STRATUM	Num	1.0		72	73	STRATUM (1=REG. ,2-=LOW ANC,3-=SMALL Al)	9	
9	STRMDP		Num	3.1		75	78	STREAM DEPTH (N)	9
10	STRMNAM		Char	30.0		0	30	STREAM NAME	10
10	STRMWD		Num	4.1		32	36	STREAM WIDTH (N)	10
10	STRM_ID		Char	9.0		38	47	STREAM/SITE IDENTIFICATION CODE	10
10	SUB_ID		Char	3.0		49	52	SUBREGION IDENTIFICATION CODE	10
10	W		Num	12.6		54	66	REACH WEIGHTING FACTOR	10
11	WC		Num	12.6		0	12	STAGE II CONDITIONAL WEIGHT	11

Card #	Variable Name	Type	Variable Format		Start	Column End	label	Card #
1	A1	Hum	7.3		0	7	DIRECT WATERSHED AREA (SQ HI)	1
1	A1PRIME	Num	7.3		9	16	UPDATED (1989) Al (SQ HI)	1
1	A2	Hum	7.3		18	25	WS AREA TO MAPPED UPPER NODE (SQ MI)	1
1	A3	Hum	7.3		27	34	WS AREA TO HAPPED HEADWATER (SQ MI)	1
1	A4	Hum	7.3		36	43	WS AREA BETWEEN U/L SAMPLE SITE (SQ KM)1	
1	A5	Hum	7.3		45	52	WS AREA TO UPPER SAMPLE SITE (SQ (SQ KM)1	
1	ACCO11		Hum	9.3		54	63	BASE NEUTRALIZING CAPACITY (UEQ/L)1	
1	ALEX16		Hum	9.3		65	74	EXTRACTABLE (MIBK) ALUMINUM (UMOL/L)1	
2	ALKA11		Hum	9.3		0	9	ACID NEUTRALIZING CAPACITY (UEQ/L)2	
2	ALORl6	Hum	9.3		11	20	ORG. MONOMERIC (PCV) ALUMINUM (UMOL/L)	2
2	ALTL16	Hum	9.3		22	31	TOTAL ALUMINUM (UMOL/L)	2
2	ANDEF	Hum	9.3		33	42	ANION DEFICIT, CATSUM-ANSUM (UEQ/L)	2
2	ANSUM		Hum	9-3		44	53	SUM OF ANIONS (UEQ/L)	2
2	A WS	Hum	7.3		55	62	WS AREA TO MAPPED NODE (SQ KM)	2
2	CA16	Hum	9.3		64	73	CALCIUM (UEQ/L)	2
3	CATSUM		Hum	9.3		0	9	SUM OF CATIONS (UEQ/L)	3
3	CL16	Hum	9.3		11	20	CHLORIDE (UEQ/L)	3
3	C0316	Hum	9.3		22	31	CARBONATE (UEQ/L)	3
3	COLVAL		Hum	8.0		33	41	COLOR VALUE (PCU)	3
3	COND11		Hum	9.3		43	52	CONDUCTANCE -ANALYTICAL LAB- (US/CM)3	
3	CONIS	Hum	9.3		54	63	IN-SITU CONDUCTANCE (US/CM)	3
4	COUNTY1	Char	15.0		0	15	COUNTY NAME	4
4	DATSMP		Hum	7.0		17	24	DATE SAMPLED	4
4	DICE11	Hum	9.3		26	35	AIR EQUIL. DIS. INORG CARBON (HG/L)	4
4	DICI11	Hum	9.3		37	46	INITIAL DIS. INORGANIC CARBON (HG/L)	4
4	DICVAL		Hum	9.3		48	57	DIS. INORG. CARBON -PROCESS.LAB- (MG/L)4	
4	DOC11	Hum	9.3		59	68	015. ORGANIC CARBON (HG/L)	4
5	DO_IS	Hum	9.3		0	9	IN-SITU DISSOLVED OXYGEN (MG/L)	5
5	DRPCDE		Hum	2.0		11	13	SITE EXCLUSION CODE (0,1 2 3 4 5 13)	5
5	ELEV	Hum	7.2		15	22	SAMPLE SITE ELEVATION (H)	5
5	FE16	Hum	9.3		24	33	IRON (UMOL/L)	5
5	FTL16	Hum	9.3		35	44	TOTAL FLUORIDE (UEQ/L)	5
5	GRADE	Hum	8.2		46	54	STREAM REACH GRADIENT (X)	5
5	H16	Hum	9.4		56	65	HYDROGEN ION ACTIVITY (UEQ/L)	5
5	HC0316		Hum	9.3		67	76	BICARBONATE (UEQ/L)	5
6	K16	Hum	9.3		0	9	POTASSIUM (UEQ/L)	6
6	L2	Hum	9.3		11	20	LENGTH BETWEEN U/L SAMPLE SITES (KM)	6
6	LASNAM		Char	6.0		22	28	CHEMICAL ANALYSIS LABORATORY NAME6	
6	LAT_STD	Hum	9.4		30	39	SAMPLE SITE LATITUDE (DECIMAL FORM)	6
6	LON~STD	Hum	9.4		41	50	SAMPLE SITE LONGITUDE (DECIMAL FORM)	6
7	MAP_	Char	32.0		0	32	1:24,000 SCALE MAP NAME	7
7	MAP2	Char	32.0		34	66	1:24,000 SCALE MAP NAME	7
8	MAP3	Char	32.0		0	32	1:24,000 SCALE MAP NAME	8
8	MAP4	Char	32.0		34	66	1:24,000 SCALE MAP NAME	8
9	MAPS	Char	32.0		0	32	1:24,000 SCALE MAP NAME	9
9	MG16	Hum	9.3		34	43	MAGNESIUM (UEQ/L)	9
9	MH16	Hum	9.3		45	54	MANGANESE (UMOL/L)	9
9	NA16	Hum	9.3		56	65	SODIUM (UEQ/L)	9

9	NH416	Hum	9.3		67	76	AMMOMIUM (UEQ/L)	9
10	N0316	Hum	9.3		0	9	NITRATE (UEQ/L)	10
10	NODE	Char	1.0		11	12	REACH SAMPLE POSITION (U=UPPER,L-LOWER)	10
10	NOTSAM		Chat	30.0		14	44 REASON NOT SAMPLED	10
10	OH16	Hum	9.3		46	55	HYDROXIDE (UEQ/L)	10
10	ORGION		Hum	9.3		57	66 CALCULATED ORGANIC ANIONS (UEQ/L)	10
10	PHAC11		Hum	4.2		68	72 INITIAL PH, ACIDITY TITRATION	10
10	PHAL11		Hum	4.2		74	78 INITIAL PH, ALKALINITY TITRATION	10
11	PHEQ11		Hum	4.2		0	4 AIR EQUILIBRATED LAB PH	11
11	PHSTVL		Hum	4.2		6	10 CLOSED SYSTEM PH -PROCESS. LAB-	11
11	PH_CLO		Hum	4.2		12	16 FIELD PH, CLOSED CONTAINER -PILOT ONLY	11
11	PH_R	Hum	4.2		18	22	FIELD PH, OPEN SYSTEM	11
11	PTLI6	Hum	9.3		24	33	TOTAL PHOSPHOROUS (UMOL/L)	11
11	QUAD	Chat	30.0		35	65	1:250,000 SCALE MAP NAME	11
11	RCH_HW	Hum	2.0		67	69	SHREVE ORDER -1:250,000 SCALE MAP	11
12	RCH_ID		Char	8.0		0	8 REACH IDENTIFICATION CODE	12
12	RCH_LN		Hum	9.3		10	19 LENGTH OF MAPPED BLUE LINE REACH (KM)	12
12	SANCOD	Char	3.0		21	24	SAMPLE TYPE (D,DA,E,EDA,ER,NS,SY,R)	12
12	SAMRN		Hum	1.0		26	27 SAMPLE VISIT NUMBER (0,1,2,3,4)	12

TABLE E-3. Card-image Format Definition, NSS Data Det PILOTDS4

Card #	Variable Name	Type	Variable Format			Start	Column End	Column label	Card #
12	SHRE75		Num	3.0		29	32	SHREVE ORDER -1:24,000 SCALE MAP	12
12	S10216	Num	9.3		34	43	DISSOLVED SILICA (UMOL/L)		12
12	SIT CLS		Char	6.0		45	51	SITE CHARACTERISTIC CODE	12
12	SO4T6	Num	9.3		53	62	SULFATE (UEQ/IL)		12
12	SOBC	Num	9.3		64	73	SUM OF BASE CATIONS (UEQ/L)		12
12	STATE1		Char	2.0		75	77	STATE (TWO CHARACTER CODE)	12
13	STATE2		Char	2.0		0	2	STATE (TWO CHARACTER CODE)	13
13	STRA75		Num	9.3		4	13	STRAHLER ORDER -1:24,000 SCALE MAP13	
13	STRATUM	Num	1.0		¶5	16	STRATUM (1--REG.,2=LOW ANC,3--SMALL A1)		13
13	STRMDP		Num	3.1		18	21	STREAM DEPTH CM)	13
13	STRMMAM	Char	30.0		23	53	STREAM NAME		13
13	STRMWD	Num	4.1		55	59	STREAM WIDTH (N)		13
13	STRM_ID	Char	9.0		61	70	STREAM/SITE IDENTIFICATION CODE		13
13	SUB_ID		Char	3.0		72	75	SUBREGION IDENTIFICATION CODE	13
14	TIMSMP		Num	5.0		0	5	TIME SAMPLED (HH:MM)	14
14	TMPSTR		Num	9.3		7	16	STREAM TEMPERATURE (DEG C)	14
14	TURVAL		Num	9.3		18	27	TURBIDITY (MTU)	14
14	W	Num	12.6		29	41	REACH WEIGHTING FACTOR		14
14	WC	Mum	12.6		43	55	STAGE II CONDITIONAL WEIGHT		14

TABLE E-4. Card-image Listing (First Five Observations), Data Set NSSIDS4A, U.S. EPA National Stream Survey

```
0         1         2         3         4         5         6         7         8
01234567890123456789012345678901234567890123456789012345678901234567890123456789
0

   0.980     0.980    21.250     0.000     2.564    54.903    43.900
0.6191         1
     0.185     0.011   740.400         0.608     1.594    13.679
1176.559         2
  57.576   848.300   1190.238    166.157         5.373          5
125.000         3
    87.000   125.200  CHAUTAUQUA
            4
                23APR86          9.220     9.370     8.943         1.870
            5
   11.200    0   410.55          0.179     2.316     0.61      0.0126
721.503         6
   18.487     2.108  NYSDOH     42.2150   79.0992
            7
KENNEDY, NY 1979                  CHERRY CREEK, NY 1954
            8
HAMLET, NY 1954                   GERRY, NY 1979
            9

205.650         10
      0.055   116.580         1.209      43.067  L
           11
      0.794    18.531     7.58  7.56    7.75  7.90     -999  8.24     0.084    -
999.000         12
BUFFALO 1962                      4  1D022009       2.172  R    1   19
           13
     56.088            237.348    1189.017  N Y          4.000    2.000   3
0.3             14
CLEAR CREEK                       7.0   1D022009L   1D   12:25    10.500
           15
      1.200        65.306122         1.000000
           16
     0.980      0.980   21.250     0.000    2.564    54.903    36.200
0.334           1
      0.089       0.000   1324.200        0.552     0.519     8.474
1789.925        2
   57.570      1312.370  1798.399     237.528     8.316       10   186.000
            3
    118.000         187.600  CHAUTAUQUA
            4
                     0BMAY86    14.400    14.900    15.210    0.843
            5
      10.600       0    410.55       0.125    2.421    0.61    0.0138   1224.260
            6
      22.962          2.108  NYSDOH   42.2150   79.0992
            7
KENNEDY, NY 1979                  CHERRY CREEK, NY 1954
            8
HAMLET, NY 1954           GERRY, NY 1979
            9
                                                                          315.878
           10
      0.055    146.160    1.015    52.261  L
           11
      0.724      8.370   7.73  7.73 8.29  7.86   -999   7.98     0.132          -
999.000         12
BUFFALO 1962                   4  10022009      2.172  R   2    19
           13
    27.794           264.414  1797.370  N          4.000    2.000       3  0.3
            14
CLEAR CREEK                    7.0   1D022009L  1D   7:05    8.700
           15
      0.370      65.306122    1.000000
           16
    0.980      0.980  21.250  0.000     2.564    54.903    42.000     0.678
            1
      0.222         0.000    763.100     0.778     1.371    38.376   1183.386
```

2
55.032 878.240 1221.762 162.772 3.706 10 126.400
3
89.000 121.400 CHAUTAUQUA
4
23APR86 9.100 9.050 9.231 1.610
5
10.900 0 423.35 0.215 2.421 0.61 0.0186 735.925
6
18.768 2.108 NYSDOH 42.2264 79.1156
7
KENNEDY, NY 1979 CHERRY CREEK, NY 1954
8
HAMLET, NY 1954 GERRY, NY 1979
9
213.053
10
0.091 110.490 1.192 44.841 U
11
0.537 15.965 7.63 7.65 7.84 7.73 -999 7.92 0.107 -
999.000 12
BUFFALO 1962 4 10022009 2.172 R 1 19
13 51.262 233.184 1220.552 NY
4.000 2.000 3 0.2 14
CLEAR CREEK 6.0 1D022009U 1D 13:25
10.500 15
0.800 65.306122 1.000000
16
0.980 0.980 21.250 0.000 2.564 54.903 54.100
0.359 1
0.170 0.000 1277.000 0.567 0.852 49.396
1770.054 2
55.032 1347.300 1819.450 226.808 4.356 10
180.200 3
113.000 182.200 CHAUTAUQUA
4
08MAY86 15.200 15.700 15.570 2.420
5
9.800 0 423.35 0.090 2.316 0.61 0.0263
1221.937 6
22.962 2.108 NYSDOH 42.2264 79.1156
7
KENNEDY, NY 1979 CHERRY CREEK, NY 1954
8
HAMLET, NY 1954 GERRY, NY 1979
9

306.007 10
0.036 142.245 0.909 49.842 U
11
0.380 24.030 7.74 7.73 8.23 7.58 -999 7.69 0.042 -
999.000 12
BUFFALO 1962 4 1D022009 2.172 R 2 19
13
36.615 264.414 1818.514 NY 4.000 2.000 3
0.2 14
CLEAR CREEK 6.0 10022009U 10 6:15 9.800
15
0.350 65.306122 1.00000

0 1 2 3 4 5 6 7
8
012345678901234567890123456789012345678901234567890123456789012345678901234567
8901234567890

TABLE E-4. Card-image Listing (First Five Observations), Data Set NSSIDS4A, U.S. EPA National Stream Survey (continued)

```
0         1         2         3         4         5                   6         7         8
01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890

  18.470  18.470   0.000           0.91    45.243    2.564           91.500    0.574     1
    0.111    0.000          1020.900      0.949      3.37     29.261    1537.182  2
  47.837  1087.820          1566.443    210.164      4.607      15     162.000   3
    97.000   142.400          CATTARAUGUS
                                                                                4
                  23APR86              12.500    12.900    12.330           3.070     5
    11.800  0  386.47           0.609    2.790              1.06    0.0200    980.334   6
    25.826   13.807    NYSDOH    42.1561  78.9656                              7
RANDOLPH, NY 1979                    NEW ALBION, NY 1963                              8
                                                                                        9
                                                                                282.152   10
      0.255    169.650    0.976           53.552  L                                    11
      0.501     30.284  7.32   7.36         8.05  7.70  -999   7.92.    0.158   -999.000  12
BUFFALO 1962                                1  1D022010    14.419   R 1   28          13
    56.421           285.234          1565.447  NY          4.000
                                                  1.000  1   0.2     14
ELM CREEK                                   7.0  1D022010L  10   9:30    6.500     15
      1.300     59.738953          17.240288                                        16


0         1         2         3         4         5         6         7         8
01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
```