



## Project Summary

# The National Surface Water Survey National Stream Survey Phase I - Pilot Survey: Summary of Quality Assurance Data Results

Sevdâ K. Drousé

A primary objective of the National Surface Water Survey and, thus, of the National Stream Survey is to ensure that the data collected are scientifically sound and of known quality. An extensive quality assurance program has been established in support of this objective. To evaluate the effectiveness of the quality assurance program, several types of quality assurance and quality control samples were collected and analyzed during a pilot survey that was conducted prior to the initiation of National Stream Survey Phase I field activities. This report presents a statistical analysis of results obtained for field duplicate samples, blank samples, and audit samples used in the pilot survey. The results show that even overall estimated within-batch precision was adequate to meet the analytical data quality objectives established for the National Stream Survey and that detection limit goals were achieved at the contract analytical laboratories. The observed system decision limits and system detection limits, however, must be considered in interpreting the pilot study data and data from future surveys that employ similar sampling, processing, and analytical methods.

This report was submitted in partial fulfillment of contract number

68-03-3249 by Lockheed Engineering and Management Services Company, Inc., under the sponsorship of the U.S. Environmental Protection Agency. This report covers a period from March 1, 1985, to July 16, 1985, and field work completed as of July 16, 1985.

*This Project Summary was developed by EPA's Environmental Monitoring Systems Laboratory, Las Vegas, NV, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).*

### Introduction

The National Stream Survey (NSS) Phase I - Pilot Survey was conducted in the Southern Blue Ridge Province of the United States between March 1 and July 16, 1985. This study was performed prior to the initiation of the full-scale NSS field activities as part of the National Surface Water Survey (NSWS), under the administration of the National Acid Precipitation Assessment Program (NAPAP), Task Group E (Aquatic Effects).

One of the objectives of the NSWS is to ensure that the data collected are scientifically sound and are of known quality. An extensive quality assurance (QA) program has been established in support of this objective. To evaluate the

effectiveness of the QA program and to maximize the confidence in the resulting data, several types of QA and quality control (QC) samples were collected and analyzed during the pilot survey. The purpose of this report is to summarize the resulting QA and QC sample data.

### **Activities of the National Stream Survey Phase I - Pilot Survey**

The pilot survey covered a probability sample of 54 stream reaches drawn from the population of stream reaches in the Southern Blue Ridge Province. The characteristics measured include geographic, physical, and chemical variables.

Water samples were collected by three sampling crews at the downstream node of each reach on three occasions during the spring (March 17 through April 30) at approximately biweekly intervals and on one occasion in the summer (May 30 through July 16). Samples also were collected at the upstream node of 17 reaches during the final spring sampling and of 54 reaches during the summer sampling.

The samples were filtered and preserved, aliquots were prepared, and analyses (pH, dissolved inorganic carbon, true color, and turbidity) were performed at the field station located in Sylva, North Carolina. Field sampling and field laboratory activities are illustrated in Figure 1.

All variables were measured by one contract analytical laboratory using extensively reviewed techniques and protocols. Activities and results were subjected to a high degree of quality control and quality assurance, from sample collection to the final disposition of the data in the data base. All samples were required to be analyzed within the specified holding times established for each variable to ensure the integrity of the samples. Over the course of the study, 389 routine samples were collected from 61 stream reaches (54 probability samples and 7 special interest samples) and a total of 759 samples (routine stream samples, field blanks, field duplicates, audits, contract analytical laboratory duplicates, and matrix spikes) were analyzed.

### **Evaluation of Data Quality**

During the pilot survey, QA and QC samples were used to judge the quality of data produced by field sampling, field laboratory, and contract analytical laboratory activities. For the pilot survey, only one field laboratory and one

contract analytical laboratory were used to process and analyze samples. Therefore, interlaboratory bias was not a consideration in analyzing the data. QA samples were used to evaluate the overall performance of these activities and to establish precision estimates. The QC samples were used to ensure that instruments were operating properly and that data-gathering activities were performed according to established guidelines.

Data quality objectives (DQOs) were set for the analytical data in terms of precision (expressed as relative standard deviation), accuracy (expressed as maximum absolute bias, in percent), and detectability (expressed as a required detection limit). The DQOs for representativeness, completeness, and comparability of data were special concerns for the pilot survey; the effectiveness of subsequent Phase I surveys depends on the statistical validity of pilot samples and on the legitimacy of extrapolating pilot survey results. In light of the survey design and the quality of data obtained, the pilot survey results are sufficiently representative of the stream populations and the study areas (that is, accurately and precisely reflective of their characteristics) to allow other planned Phase I activities to be meaningfully assessed. An assessment of feasibility also required an in-depth statistical analysis of the data, which was facilitated by data completeness (the quantity of acceptable data actually collected relative to the total quantity that was attempted) of more than 99 percent; the initial DQO was 90 percent. Comparability (the similarity within and among data sets) was assured by requiring use of standard protocols for collecting, processing, handling, and analyzing samples and of a uniform set of units and data forms for reporting data.

Evaluation of the QA and QC sample data was an ongoing process during and following the pilot survey. A substantial part of this evaluation process was the statistical analysis of the verified QA and QC sample data. The results of this statistical analysis are presented in this report.

### **Conclusions and Recommendations**

Results indicate that, in general, the QA program was successful in assuring that the data collected during the NSS Phase I - Pilot Survey were consistent, reliable, and of known and verifiable

quality. The data and experience obtained from the NSS Phase I - Pilot Survey will improve the full-scale NSS Phase I survey to be conducted in the spring of 1986. Recommendations for accomplishing these improvements also are provided in this section.

### **Conclusions**

1. Duplicate samples allowed evaluation of within-batch precision for the overall survey as calculated from field duplicates and of intralaboratory (or analytical) precision as calculated from trailer (field laboratory) and contract analytical laboratory duplicates. Intralaboratory precision goals for most variables were achieved even with the field duplicate samples; thus, overall estimated within-batch precision was adequate to meet the analytical data quality objectives established at the beginning of the survey.
2. Detection limit goals were achieved for all parameters in the contract analytical laboratory. However, analysis of the data for a few of the low-concentration variables indicated that background sources of contamination caused the decision limit to be significantly higher than the required detection limit. Thus, the achieved values must be evaluated in light of background values obtained for field blanks. If the background value from sample collection and handling is higher than the laboratory detection limit, obtaining extremely low detection limits in the laboratory is meaningless. The system decision limit (the lowest instrument signal that can be distinguished from the background) and system detection limit (the lowest concentration that can be meaningfully measured above the system decision limit) are more representative indicators of data quality for variables that are not instrumental or method detection limits.
3. Evaluation of data from synthetic air samples analyzed during the pilot survey (measured versus theoretical concentrations) shows that the NSS pilot survey sample collection and analysis system accurately measures the parameters of interest in most cases. However, three variables (iron, extractable aluminum, and initial [air-equilibrated] dissolved inorganic carbon) did not show clo

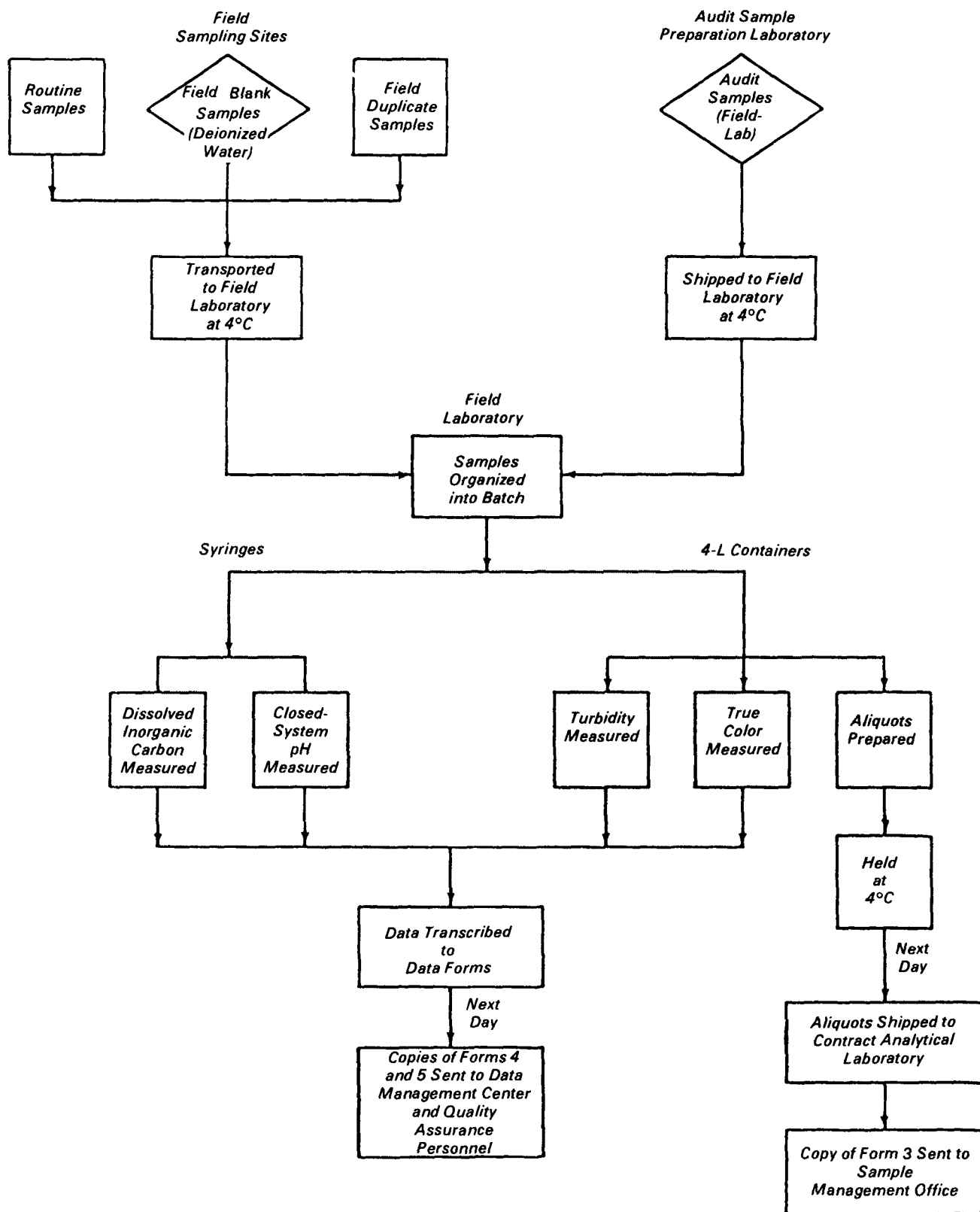


Figure 1. Flowchart of field sampling and field laboratory activities for the NSS Phase I - Pilot Survey.

agreement between measured and theoretical values.

4. It was not possible to quantify *absolute* accuracy using field and laboratory audit synthetic sample data because the true concentrations of these samples are not known. Due to the complex equilibria and analyte incompatibilities, it is not possible to adjust every analyte concentration to any desired level. It is also not possible to predict theoretical values for acid-neutralizing capacity, base-neutralizing capacity, specific conductance, and pH.
5. Analysis of matrix spike data shows that no matrix effect was apparent in the stream samples.
6. A comparison between precision estimated from field and contract analytical laboratory duplicate sample data and precision estimated from audit sample data should not be made because precision estimated from duplicate data represents a wide range of concentrations, whereas precision estimated from audit data represents a specific concentration range.
7. On the basis of QA and QC samples used in the NSS Phase I - Pilot Survey, two components of variability can be estimated for all parameters except pH, dissolved inorganic carbon, true color, and turbidity. Three components can be estimated for pH and dissolved inorganic carbon, one can be estimated for true color and turbidity.

## Recommendations

Recommendations for future surveys include:

1. Using synthetic and natural audit samples with at least five different compositions (from near the detection limit to throughout the expected ranges) in order to maximize the confidence of obtaining accurate measurements of synthetic and natural audit samples.
2. Using only laboratory synthetic audit samples because they bypass the filtration and sample processing steps at the field laboratory.
3. Using  $\text{NH}_4$  acetate rather than the more volatile  $\text{NH}_4\text{Cl}$  for the MIBK extraction of aluminum to minimize contamination of samples by ammonia in the field laboratory.
4. Pouring the aluminum aliquot in the hood to minimize contamination by aluminum-containing dust in the field laboratory.
5. Using a Plexiglas panel to separate the acid-washed filtering apparatus from the filtering apparatus that is used to process the nitrate aliquot that is not acid washed.

Recommendations for data users to consider include:

1. Subtraction of the blank levels from the measured routine stream sample concentrations for those variables with high background concentrations.
2. Incorporation of additional QA and QC samples at each step of sample processing and additional measurements at the sampling site and at the field laboratory to partition components of variability. (Such additional procedures would increase the cost of the QA program and may be logistically difficult.)

## Procedures

The NSS Phase I - Pilot Survey QA program used several types of QA and QC samples as described in the following paragraphs.

QA samples (field blanks, field duplicates, and audits) introduced in the field or at the field laboratory were analyzed at the field laboratory and at the contract analytical laboratory. These samples were used to evaluate overall method performance for field sampling, field laboratory procedures, and contract analytical laboratory procedures, as well as to evaluate overall data quality. The QA samples were "double blind" to the contract analytical laboratory (i.e., the laboratory did not know the origin, identity, or composition of the samples). Consequently, the contract analytical laboratory processed and analyzed QA samples as it would any stream water sample.

The supporting QC samples (laboratory blanks, laboratory duplicates, check standards, and matrix spikes) allowed field samplers, field laboratory personnel, and contract analytical laboratory personnel to identify and correct local problems (e.g., reagent contamination or faulty instrument performance) as they occurred. In some cases, the same sample served as a QA sample and as a QC sample. The types of QA and QC samples used during the NSS Phase I - Pilot Survey are described below.

## Blank Samples

Blank samples were used to identify contamination problems and instrument drift. They also provided data that were used to determine system detection

limits, system decision limits, quantitation limits, and method and instrument detection limits.

## Duplicate Samples

Duplicate samples were used to determine sample homogeneity and to estimate overall method precision, which includes effects of collection, handling, processing, and analyses. The differences between the amounts of overall (field duplicate) and analytical (laboratory duplicate) within-batch precision were used to indicate whether or not data variability resulted from sample collection, processing, and handling.

## Audit Samples

Audit samples were used to estimate overall precision and accuracy of measurement system and to provide information about the quality of the routine stream sample data.

## Methods and Results

### Blank Samples

During the pilot survey, 71 field blanks were processed and analyzed. The field laboratory did not routinely analyze field blanks and did not analyze field blank for pH or DIC because (1) there were significant detection limit problems with measurements performed in the field laboratory and (2) pH and DIC measurements were highly variable in field blanks because they tend to absorb CO<sub>2</sub> from the air during preparation and analysis.

A statistical evaluation of the verification data yielded a system decision limit and an estimated system detection limit for each variable. No problems were encountered for most of the variables. However, analysis of the data for some of the low-concentration variables indicated that background sources of contamination caused the decision limit to be significantly higher than the required detection limit. The data should consider this factor when evaluating these data for ammonium, aluminum, dissolved inorganic carbon, base neutralizing capacity, calcium chloride, dissolved inorganic carbon, phosphorus, and nitrate.

The contract analytical laboratory required to determine and to report instrumental and method detection limits. The laboratory calculated these limits three times the standard deviation of nonconsecutive calibration and reagent blank analyses (the parametric method assumes normal distribution).

Instrumental detection limits for each required variable were also estimated from the detection limit quality control check sample measured in the contract analytical laboratory. At the field laboratory, instrumental detection limits were estimated from the calibration blanks for DIC measurement. Instrumental detection limits are not applicable to pH, true color, and turbidity measurements.

### **Duplicate Samples**

In total, 98 field duplicate samples were collected during the NSS Phase I - Pilot Survey. The data from routine-duplicate sample pairs were analyzed to provide an overall estimate of the within-batch precision, which includes the effects of sample collection, processing, and analysis on data variability. This estimate did not include the effect of among-batch variation that may have been caused by day-to-day differences such as different calibration curves.

For variables other than pH, precision for a routine-duplicate pair is reported as the percent relative standard deviation (%RSD). A valid summary statistic for the precision overall routine-duplicate pairs for variables other than pH is the pooled or root-mean-square (RMS) of the %RSD values. For pH variables, the summary statistic is the RMS of the standard deviation values of the routine-duplicate pairs.

Because large values of  $\text{RMS}\% \text{RSD}$  and  $\text{RMS}_{\text{standard deviation}}$  indicate poor precision between routine-duplicate pairs for a given variable at low concentrations, very small differences between routine and duplicate samples result in large %RSD values. An objective technique for determining the overall precision for a variable uses only those routine-duplicate pairs whose mean is greater than 10 times the standard deviation of field blank samples (the quantitation limit). For each variable measured during the NSS pilot survey, the precision was estimated twice. For the first estimate, only those routine-duplicate pairs whose mean was greater than zero were included in the calculation. For the second estimate, only those routine-duplicate pairs whose mean was greater than the quantitation limit were included.

The results indicate that overall within-batch precision of pH measurement was better than the required 0.1 pH unit. For certain variables (Ca, Mg, K,  $\text{SO}_4^{2-}$ , and specific conductance), reported

measurements were far above the required detection limits. These results indicate, as does the pH measurement, that for each variable the overall within-batch precision was better than the intralaboratory precision goal.

Some analytes (Mn, Fe, organic monomeric Al, total extractable Al, total Al,  $\text{NO}_3^-$ ,  $\text{SiO}_2$ ,  $\text{NH}_4^+$  and total P) were characterized by low concentrations (at or below the detection limit) in many of the samples. Thus, an estimate of the true precision was difficult to ascertain. As the instrumental detection limit is approached, the relative variability in the analysis increases.

The quantitation limit is the concentration above which relative precision stabilizes. The larger variation and effects of low concentrations of analytes at the detection limit must be considered when interpreting the precision of values that are less than the quantitation limit. Variability also could be confounded by the fact that the field duplicate sample is collected 15 to 30 minutes after the routine stream sample is collected. Unlike lake samples, stream samples are collected from flowing waters; therefore, homogeneity of the stream routine and duplicate samples may be in question.

Duplicate analyses identified as trailer duplicates were performed once per batch in the field laboratory for all measurements (DIC, pH, turbidity, and true color). The observed precision for pH and DIC was within the intralaboratory precision (RMS %RSD) goals established for the study. Because most of the streams sampled were of low turbidity and were colorless, and because true color is a coarse measurement (read to the nearest 5 PCU), precision for these variables was expected to exceed the intralaboratory precision goals.

The results obtained at the contract analytical laboratory indicate that the observed analytical precision for all variables (including pH) was better than the intralaboratory precision goal. These data may have been biased because the laboratory analyst chose which sample to duplicate and knew that the QC procedures required that the precision goal be achieved. The results, however, are an indication of the precision that was achieved within the contract analytical laboratory when the method QC requirements were followed.

For each of the variables except total Al, Fe, and  $\text{NH}_4^+$  the estimated analytical laboratory within-batch precision was better than the intralaboratory precision goal.

The measurement of within-batch precision, the portion of the total data variability that occurred during chemical analyses of the samples collected and processed on a given day, was based on results for duplicate pairs with means greater than 10 times the standard deviation of the calibration blanks (for all variables) and reagent blanks (for  $\text{SiO}_2$  and total Al) analyzed by the analytical laboratory.

Two components of variability could be estimated for all variables. Three components could be estimated for pH and DIC; one component could be estimated for true color and turbidity. The overall precision estimates from field duplicate pairs are generally slightly higher than the analytical precision from contract analytical laboratory duplicate pairs. This may be a result of variability in the field procedure or a result of real differences in the measurement of water collected in successive samples.

### **Audit Samples**

Natural (field and laboratory) and synthetic (field and laboratory) audit samples processed and analyzed during the survey were used to estimate among-batch precision within a specific concentration range. Natural and synthetic field audits both indicate variability resulting from processing and analysis, but not from sample collection; laboratory audits provide information about precision of analytical results without involving potential effects of field laboratory processing. Natural and synthetic audit samples also were used to judge the performance of the laboratory in analyzing individual sample batches. Relative interlaboratory bias could not be determined from the audit samples because only one contract analytical laboratory was used during the pilot survey.

A total of 46 field natural audit samples and 17 laboratory natural audit samples were analyzed during the pilot survey. The synthetic audit samples used during the pilot survey all had low concentrations of analytes. A total of 36 synthetic samples were used.

All samples appeared to be stable throughout the survey; however, the synthetic audit samples showed a greater than desired variance for some variables, which could be associated with audit sample preparation. On an individual basis the synthetic audit samples appeared to be stable from the time of preparation to the time of analysis. In general, the relative precision estimated from laboratory synthetic audit samples

was marginally better than that from field synthetic audit samples indicating that the processing and handling of the aliquots in the field laboratory did not introduce more than the slight variability that might be expected from such additional activities. For the NSS Phase I - Pilot Survey data, however, the natural audit samples can be expected to provide a better estimate of the among-batch precision.

If the concentration is near or below the quantitation limit (among-batch precision), a high %RSD is expected because large relative errors may occur at low concentrations. For example, some analytes (total P,  $\text{NH}_4^+$  and Mn) were present at low concentrations in the natural stream waters sampled during the pilot survey.

Among-batch precision for the field laboratory and the contract analytical laboratory was compared for pH and DIC measurements. Precision was estimated from results for the natural audit samples and field laboratory low-concentration synthetic audit samples. The overall precision estimated for field laboratory pH and DIC analyses appears to be close to that for the contract analytical laboratory. The slightly better precision that was observed in the field laboratory measurements could be a result of (1) a more controlled closed system in the field laboratory and (2) the field laboratory personnel knowing which samples were audits.

The comparison of the measured concentrations of analytes in the synthetic audit samples with the theoretical concentrations provides valuable information on relative accuracy of a measurement system. Accuracy is defined as a measure of the closeness of an individual measurement or the average of a number of measurements to the true value. For most variables there is reasonable agreement between the measured and theoretical values. The most notable exceptions are iron and total extractable aluminum, which essentially are not measured in the field synthetic audit samples but which are present in the laboratory synthetic audits. These two analytes probably were removed during filtration in the field laboratory.

A high percent recovery for total aluminum in field synthetic audit samples was most likely due to the contamination that resulted from the large amount of dust at the field laboratory early in the survey. Measured values for initial dissolved inorganic carbon (DIC) were consistently above the theoretical

concentration. The synthetic audit samples probably absorbed atmospheric  $\text{CO}_2$ , and this process increased the DIC concentration. The high measured values for dissolved organic carbon (DOC) could have resulted from contamination caused by airborne volatile organic carbons. The background levels of DOC in blanks was also high.

Average percent recoveries could not be determined for ANC, BNC, pH, specific conductance, and air-equilibrated DIC because the levels of these variables are determined by those of other variables that make up the synthetic audit samples; these variables also are related and affect the analytical results of one another.

---

*Sevdâ Droué is with Lockheed Engineering and Management Services  
Company, Las Vegas, NV 89119.*

*Robert Schonbrod is the EPA Project Officer (see below).*

*The complete report, entitled "The National Surface Water Survey, National  
Stream Survey, Phase I—Pilot Survey, Summary of Quality Assurance Data  
Results," (Order No. PB 88-140 298/AS; Cost: \$14.95, subject to change)  
will be available only from:*

*National Technical Information Service  
5285 Port Royal Road  
Springfield, VA 22161  
Telephone: 703-487-4650*

*The EPA Project Officer can be contacted at:*

*Environmental Monitoring Systems Laboratory  
U.S. Environmental Protection Agency  
Las Vegas, NV 89193-3478*

United States  
Environmental Protection  
Agency

Center for Environmental Research  
Information  
Cincinnati OH 45268

---

Official Business  
Penalty for Private Use \$300

EPA/600/S8-87/057

0000329 PS

U S ENVIR PROTECTION AGENCY  
REGION 5 LIBRARY  
230 S DEARBORN STREET  
CHICAGO IL 60604