United States
Environmental Protection
Agency

Office of Research and
Development
Washington, D.C. 20460

EPA/600/R-97/144
March 1998

EPA Environmental Technology Verification Report

Field Portable X-ray Fluorescence Analyzer

HNU Systems SEFA-P

SITE
SUPERFUND INNOVATIVE TECHNOLOGY EVALUATION

ETV
Environmental Technology
Verification Program

061CMB98

EPA/600/R-97/144
March 1998

Environmental Technology Verification Report

Field Portable X-ray Fluorescence Analyzer

HNU Systems SEFA-P

U.S. Environmental Protection Agency
Office of Research and Development

Notice

The information in this document has been funded wholly or in part by the U.S. Environmental Protection Agency (EPA) under Contract No. 68-CO-0047 to PRC Environmental Management, Inc. This work supports the Superfund Innovative Technology Evaluation Program administered by the National Risk Management Research Laboratory, Cincinnati, Ohio. This demonstration was conducted under the Monitoring and Measurement Technologies Program which is managed by the National Exposure Research Laboratory–Environmental Sciences Division, Las Vegas, Nevada. It has been subjected 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 of specific products.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Office of Research and Development
Washington, D.C. 20460

ENVIRONMENTAL TECHNOLOGY VERIFICATION PROGRAM

ENVIRONMENTAL TECHNOLOGY VERIFICATION PROGRAM
VERIFICATION STATEMENT

TECHNOLOGY TYPE:	**FIELD PORTABLE X-RAY FLUORESCENCE ANALYZER**
APPLICATION:	**MEASUREMENT OF METALS IN SOIL**
TECHNOLOGY NAME:	**SEFA-P ANALYZER**
COMPANY:	**HNU SYSTEMS, INC.**
ADDRESS:	**160 CHARLEMONT ST.** **NEWTON HIGHLANDS, MA 02161-9987**
PHONE:	**(617) 964-6690**

The U.S. Environmental Protection Agency (EPA) has created a program to facilitate the deployment of innovative technologies through performance verification and information dissemination. The goal of the Environmental Technology Verification (ETV) Program is to further environmental protection by substantially accelerating the acceptance and use of improved and more cost-effective technologies. The ETV Program is intended to assist and inform those involved in the design, distribution, permitting, and purchase of environmental technologies. This document summarizes the results of a demonstration of the HNU SEFA-P Analyzer.

PROGRAM OPERATION

The EPA, in partnership with recognized testing organizations, objectively and systematically evaluates the performance of innovative technologies. Together, with the full participation of the technology developer, they develop plans, conduct tests, collect and analyze data, and report findings. The evaluations are conducted according to a rigorous demonstration plan and established protocols for quality assurance. The EPA's National Exposure Research Laboratory, which conducts demonstrations of field characterization and monitoring technologies, selected PRC Environmental Management, Inc., as the testing organization for the performance verification of field portable X-ray fluorescence (FPXRF) analyzers.

DEMONSTRATION DESCRIPTION

In April 1995, the performance of seven FPXRF analyzers was determined under field conditions. Each analyzer was independently evaluated by comparing field analysis results to those obtained using approved reference methods. Standard reference materials (SRM) and performance evaluation (PE) samples also were used to independently assess the accuracy and comparability of each instrument.

The demonstration was designed to detect and measure a series of inorganic analytes in soil. The primary target analytes were arsenic, barium, chromium, copper, lead, and zinc; nickel, iron, cadmium, and antimony were secondary analytes. The demonstration sites were located in Iowa (the RV Hopkins site) and Washington (the ASARCO site). These sites were chosen because they exhibit a wide range of concentrations for most of the target metals and are located in different climatological regions of the United States; combined, they exhibit three distinct soil types: sand, clay, and loam. The conditions at these sites are representative of those environments under which the technology would be expected to operate. Details of the demonstration, including a data summary and

discussion of results, may be found in the report entitled "Environmental Technology Verification Report, Field Portable X-ray Fluorescence Analyzer, HNU Systems SEFA-P." The EPA document number for this report is EPA/600/R-97/144.

The EPA Method 6200 was tested and validated using the data derived from this demonstration. This method may be used to support the general application of FPXRF for environmental analysis.

TECHNOLOGY DESCRIPTION

This analyzer operates on the principle of energy dispersive X-ray fluorescence spectroscopy where the characteristic energy components of the excited X-ray spectrum are analyzed directly by an energy proportional response in an X-ray detector. Energy dispersion affords a highly efficient, full-spectrum measurement which enables the use of low intensity excitation sources (such as radioisotopes) and compact battery-powered, field-portable electronics. The FPXRF instruments are designed to provide rapid analysis of metals in soil. This information allows investigation and remediation decisions to be made on-site and reduces the number of samples that need to be submitted for laboratory analysis. In the operation of these instruments, the user must be aware that FPXRF analyzers do not respond well to chromium and that detection limits may be 5 to 10 times greater than conventional laboratory methods. As with all field collection programs, a portion of the samples should be sent to a laboratory for confirmatory analyses.

The SEFA-P Analyzer can use up to three radioactive sources with a lithium-drifted silicon detector to analyze a large number of metals in a variety of matrices. The SEFA-P is a transportable (weighs about 50 pounds) analyzer that operates in the intrusive mode (it only measures samples in cups). The SEFA-P Analyzer was able to measure all 10 target analytes for this demonstration. Most of the sample analysis data were collected after the actual demonstration. The SEFA-P Analyzer supplied by the developer experienced a malfunction when the source holder locked into place, not allowing sample exposure and analysis. This incident occurred at the start of the demonstration at the ASARCO site. HNU was unable to supply a replacement unit in time to continue the demonstration; an EPA-owned SEFA-P Analyzer was subsequently used to analyze a subset of 100 demonstration samples. The samples chosen represented all three soil types, a wide range of concentrations from each of the sites, and included all of the PE and SRM samples from the demonstration. The SEFA-P Analyzer with three sources cost $49,000 at the time of the demonstration.

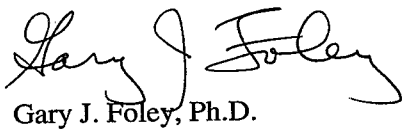
VERIFICATION OF PERFORMANCE

The performance characteristics of the SEFA-P Analyzer include the following:

- **Detection limits:** Precision-based detection limits were determined by collecting 10 replicate measurements on site-specific soil samples with metals concentrations 2 to 5 times the expected MDLs. The results were 120 milligrams per kilogram (mg/kg) for antimony and lead, 225 mg/kg for copper, and 360 mg/kg for arsenic. No values were reported for cadmium, chromium, or nickel due to an insufficient number of samples in the target concentration range. Values for iron, zinc, and barium were 900, 990, and 1150 mg/kg, respectively.
- **Throughput:** Average throughput was found to be 7 to 8 analyses per hour using a live count time of 240 seconds. This rate only represents the analysis time since different personnel were used to prepare the samples.
- **Drift:** A quantitative assessment of drift was performed using a calibration check standard which was analyzed at the beginning and end of each day. The drift RSD values for the mean recovery of the target analytes ranged from 0 to 35 percent.
- **Completeness:** The SEFA-P Analyzer produced results for 100 of the 100 samples for a completeness of 100 percent. However, prior to the mechanical failure in the field, the SEFA-P Analyzer was to have analyzed 630 samples.
- **Blank results:** None of the target analytes were reported above the precision-based method detection limits in the lithium carbonate blanks.

- **Precision:** The goal of the demonstration was to achieve relative standard deviations (RSD) less than 20 percent at analyte concentrations 5 to 10 times the method detection limits. The RSD values for antimony, barium, copper, and lead were less than 8 percent. The number of samples limited a complete assessment of this parameter; no values were reported for arsenic, cadmium, chromium, iron, nickel, or zinc.
- **Accuracy:** Accuracy was assessed using site-specific soil PE samples and soil SRMs. The SEFA-P Analyzer reported 7 of 33 or 21.2 percent of all analytes in the site-specific soil PE samples and 3 of 18 or 16.7 percent of the soil SRMs within the quantitative acceptance range of 80 - 120 percent.
- **Comparability:** This demonstration showed that the SEFA-P Analyzer produced data that exhibited a $\log_{10}$-$\log_{10}$ linear correlation to the reference data. The coefficient of determination (r^2) measured using a Compton calibration represents the degree of correlation between the reference and field data. For this demonstration, the coefficient of determination was 0.89 for antimony, 0.95 for arsenic, 0.73 for barium, 0.64 for cadmium, 0.35 for chromium, 0.92 for copper, 0.76 for iron, 0.97 for lead, and 0.89 for zinc. No value was reported for nickel due to limited sample data.
- **Data quality levels:** Using the demonstration derived precision RSD results and the coefficient of determination as the primary qualifiers, using data generated from the Compton ratio calibrations, the SEFA-P Analyzer produced definitive level data for copper, antimony, and lead; data of quantitative screening level were produced for barium. Without adequate precision or comparability data, levels for the other six analytes could not be assigned.

The results of the demonstration show that the HNU SEFA-P portable X-ray fluorescence analyzer can provide useful, cost-effective data for environmental problem-solving and decision-making. Undoubtedly, it will be employed in a variety of applications, ranging from serving as a complement to data generated in a fixed analytical laboratory to generating data that will stand alone in the decision-making process. As with any technology selection, the user must determine what is appropriate for the application and the project data quality objectives.



Gary J. Foley, Ph.D.
Director
National Exposure Research Laboratory
Office of Research and Development

Foreword

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the Agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, the EPA's Office of Research and Development (ORD) provides data and science support that can be used to solve environmental problems and to build the scientific knowledge base needed to manage our ecological resources wisely, to understand how pollutants affect our health, and to prevent or reduce environmental risks.

The National Exposure Research Laboratory (NERL) is the Agency's center for the investigation of technical and management approaches for identifying and quantifying risks to human health and the environment. Goals of the Laboratory's research program are to develop and evaluate technologies for the characterization and monitoring of air, soil, and water; support regulatory and policy decisions; and provide the science support needed to ensure effective implementation of environmental regulations and strategies.

The EPA's Superfund Innovative Technology Evaluation (SITE) Program evaluates technologies for the characterization and remediation of contaminated Superfund and Resource Conservation and Recovery Act (RCRA) corrective action sites. The SITE Program was created to provide reliable cost and performance data to speed the acceptance of innovative characterization and monitoring technologies.

Effective measurement and monitoring technologies are needed to assess the degree of contamination at a site, to provide data which can be used to determine the risk to public health or the environment, to supply the necessary cost and performance data to select the most appropriate technology, and to monitor the success or failure of a remediation process. One component of the SITE Program, the Monitoring and Measurement Technologies Program, demonstrates and evaluates innovative technologies to meet these needs.

Candidate technologies can originate from within the federal government or from the private sector. Through the SITE Program, developers are given the opportunity to conduct a rigorous demonstration of their technology's performance under realistic field conditions. By completing the evaluation and distributing the results, the Agency establishes a baseline for acceptance and use of these technologies. The Monitoring and Measurement Technologies Program is managed by ORD's Environmental Sciences Division in Las Vegas, Nevada.

Gary J. Foley, Ph.D.
Director
National Exposure Research Laboratory
Office of Research and Development

Abstract

In April 1995, the U.S. Environmental Protection Agency (EPA) conducted a demonstration of field portable X-ray fluorescence (FPXRF) analyzers. The primary objectives of this demonstration were (1) to determine how well FPXRF analyzers perform in comparison to a standard reference method, (2) to identify the effects of sample matrix variations on the performance of FPXRF, (3) to determine the logistical and economic resources needed to operate these analyzers, and (4) to test and validate an SW-846 draft method for FPXRF analysis. The demonstration design was subjected to extensive review and comment by the EPA's National Exposure Research Laboratory, EPA Regional and Headquarters Superfund technical staff, the EPA's Office of Solid Waste–Methods Section, and the technology developers.

Two sites were used for this demonstration: RV Hopkins and the ASARCO Tacoma Smelter. RV Hopkins is an active steel drum recycling facility and the location of a former battery recycling operation. It is located in Davenport, Iowa. The ASARCO site is a former copper and lead smelter and is located in Tacoma, Washington. The samples analyzed during this demonstration represented three distinct soil textures: sand, loam, and clay. The reference methods used to evaluate the comparability of data were EPA SW-846 Methods 3050A and 6010A, "Acid Digestion of Sediments, Sludges, and Soils" and "Inductively Coupled Plasma-Atomic Emission Spectroscopy," respectively.

The FPXRF analyzers tested in this demonstration were designed to provide rapid, real-time analysis of metals concentrations in soil samples. This information will allow investigation and remediation decisions to be made on-site more efficiently, and can reduce the number of samples that need to be submitted for confirmatory analysis. Of the seven commercially available analyzers tested, one is manufactured by Niton Corporation (the XL Spectrum Analyzer); two are manufactured by TN Spectrace (the TN 9000 and TN Pb Analyzer); two are manufactured by Metorex Inc. (the X-MET 920-P Analyzer and the X-MET 920-MP Analyzer); one is manufactured by HNU Systems, Inc. (the SEFA-P Analyzer); and one is manufactured by Scitec Corporation (the MAP Spectrum Analyzer). The X-MET 940, a prototype FPXRF analyzer developed by Metorex, was given special consideration and replaced the X-MET 920-P for part of the RV Hopkins sample analyses. This environmental technology verification report (ETVR) presents information regarding the SEFA-P Analyzer. Separate ETVRs have been published for the other analyzers demonstrated.

The SEFA-P Analyzer uses three separate excitation sources and, therefore, can analyze a wide range of elements. The count times for each source can be varied; the average sample throughput for this demonstration was approximately 7 to 8 samples per hour. The analyzer's software allows the use of either a Compton or empirical type calibration to quantitate analyte concentrations. Generally, the Compton calibration-based data exhibited a higher degree of comparability to the reference data relative to the empirical calibration-based data. Based on the Compton calibration, the SEFA-P provided definitive level (equivalent to reference data) data quality for antimony, lead, and copper; and quantitative screening level (not equivalent to reference data, but correctable by collecting confirmatory

samples) quality data for barium. Because of an instrument malfunction, only a subset of the original samples could be analyzed. These samples represented all three soil types and all of the PE and SRM materials but did not provide sufficient data in all concentration ranges to determine data quality levels for arsenic, chromium, zinc, nickel, iron, or cadmium. The field-based method detection limits generally exceeded the developer's listed method detection limits. In most cases, field-based method detection limits were higher than the precision-based or reference method detection limits.

This study showed that the analyzer produced data that exhibit a $\log_{10}$-$\log_{10}$ linear correlation. Through regression analysis of $\log_{10}$ transformed data, the analyzer's data can be corrected to be even more comparable to reference data. Unless a user has regulatory approval, confirmatory (reference) sampling and data correction is recommended when using this analyzer for site characterization or remediation monitoring.

This demonstration found that the analyzer was generally simple to operate. It was designed to operate in the intrusive mode only. The SEFA-P Analyzer is an effective tool for field-based analysis of metals contamination in soil. This analyzer can be expected to identify contaminated areas allowing investigation and remediation decisions to be made more efficiently on-site which may reduce the number of samples that need to be submitted for confirmatory analysis. Ownership and operation of this analyzer may require specific licensing by state nuclear regulatory agencies. There are special radiation safety training requirements and costs associated with this type of license.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

α	alpha
β	beta
AC	alternating current
Am^{241}	americium-241
CCB	continuing calibration blank
CCV	continuing calibration verification
Cd^{109}	cadmium-109
CI	confidence interval
CLP	Contract Laboratory Program
cm	centimeter
cm^2	square centimeter
cm^3	cubic centimeter
CRM	certified reference material
CRT	cathode ray tube
DC	direct current
%D	percent difference
EPA	Environmental Protection Agency
ERA	Environmental Resource Associates
ETVR	environmental technology verification report
eV	electron volt
Fe^{55}	iron-55
FPXRF	field portable X-ray fluorescence
ICAL	initial calibration
ICB	initial calibration verification
ICP-AES	inductively coupled plasma-atomic emission spectroscopy
ICS	interference check standard
ICV	initial calibration verification
IDL	instrument detection limit
IDW	investigation-derived waste
keV	kiloelectron volt
LCD	liquid crystal display
LCS	laboratory control samples
log_{10}	base 10 logarithm
LRL	lower reporting limit
MCA	multichannel analyzer
mCi	millicurie
MDL	method detection limit
mg/kg	milligram per kilogram
mL	milliliter
mm	millimeter

MMTP	Monitoring and Measurement Technologies Program
MRI	Midwest Research Institute
NERL-ESD	National Exposure Research Laboratory—Environmental Sciences Division
NIST	National Institute of Standards and Technology
OSW	Office of Solid Waste
PAL	performance acceptance limit
PARCC	precision, accuracy, representativeness, completeness, and comparability
PC	personal computer
PE	performance evaluation
PI	prediction interval
ppm	part per million
PRC	PRC Environmental Management, Inc.
QA	quality assurance
QAPP	quality assurance project plan
QC	quality control
r	correlation coefficient
r^2	coefficient of determination
RAM	random access memory
RCRA	Resource Conservation and Recovery Act
ROI	region of interest
RPD	relative percent difference
RSD	relative standard deviation
RTC	Resource Technology Corporation
SD	standard deviation
Si(Li)	silicon lithium
SITE	Superfund Innovative Technology Evaluation
SOP	standard operating procedure
SRM	standard reference material
TC	toxicity characteristic
USGS	United States Geological Survey
XRF	X-ray fluorescence

Acknowledgments

The U.S. Environmental Protection Agency (EPA) wishes to acknowledge the support of all those who helped plan and conduct this demonstration, interpret data, and prepare this report. In particular, for demonstration site access and relevant background information, Tom Aldridge (ASARCO) and Harold Abdo (RV Hopkins); for turnkey implementation of this demonstration, Eric Hess, Patrick Splichal, and Harry Ellis (PRC Environmental Management, Inc.); for editorial and publication support, Suzanne Ladish, Anne Witebsky, Karen Bollinger, and Ed Hubert (PRC Environmental Management, Inc.); for technical and peer review, Paula Hirtz, David Farnam, and Alan Byrnes (PRC Environmental Management, Inc.); for SEFA-P analyzer operation, Darrell Hamilton (PRC Environmental Management, Inc.); for sample preparation, Scott Schulte, Keith Brown, and Curt Enos (PRC Environmental Management, Inc.); for EPA project management, Stephen Billets, National Exposure Research Laboratory–Environmental Sciences Division (NERL-ESD); and for peer review, Sam Goforth (independent consultant), John Wallace (Wallace Technologies), Shirley Wasson (National Risk Management Research Laboratory), Brian Schumacher (National Exposure Research Laboratory), and Bill Engelmann (National Exposure Research Laboratory). In addition, we gratefully acknowledge the participation of Oliver Fordham, EPA Office of Solid Waste; Piper Peterson, EPA Region 10; Brian Mitchell, EPA Region 7; and John Moore, HNU Systems, Inc.

Section 1
Executive Summary

In April 1995, the U.S. Environmental Protection Agency (EPA) sponsored a demonstration of field portable X-ray fluorescence (FPXRF) analyzers. The primary objectives of this demonstration were to evaluate these analyzers for: (1) their analytical performance relative to standard analytical methods, (2) the influence of sample matrix variations (texture, moisture, heterogeneity, and chemical composition) on performance, (3) the logistical and economic resources needed to operate these technologies in the field, and (4) to test and validate an SW-846 draft method for FPXRF analysis. Secondary objectives for this demonstration were to evaluate FPXRF analyzers for their reliability, ruggedness, cost, range of usefulness, and ease of operation.

This demonstration was intended to provide users a reference measure of performance and to act as a guide for the application of this technology. In this demonstration, the reference methods for evaluating the comparability of data were SW-846 Methods 3050A and 6010A, "Acid Digestion of Sediments, Sludges, and Soils" and "Inductively Coupled Plasma-Atomic Emission Spectroscopy (ICP-AES)," respectively.

The EPA requested that PRC Environmental Management, Inc. (PRC) assist in the planning, execution, and reporting on a demonstration of FPXRF analyzers. This demonstration was conducted under the EPA's Superfund Innovative Technology Evaluation (SITE) Program and managed by the National Exposure Research Laboratory-Environmental Sciences Division (NERL-ESD) under the Monitoring and Measurement Technologies Program (MMTP), Las Vegas, Nevada.

The FPXRF analyzers tested in this demonstration were designed to provide rapid, real-time analysis of metals concentrations in soil samples. This information will allow investigation and remediation decisions to be made on-site more efficiently, and should reduce the number of samples that need to be submitted for confirmatory analysis. Of the seven commercially available analyzers evaluated, one is manufactured by Niton Corporation (the Niton XL Spectrum Analyzer); two are manufactured by Metorex Inc. (the X-MET 920–P Analyzer and the X-MET 920–MP Analyzer); two are manufactured by TN Spectrace (the TN 9000 and the TN Pb Analyzer); one is manufactured by HNU Systems, Inc. (the SEFA-P Analyzer); and one is manufactured by Scitec Corporation (the MAP Spectrum Analyzer). The X-MET 940, a prototype FPXRF analyzer developed by Metorex, was given special consideration and replaced the X-MET 920-P for part of the RV Hopkins sample analyses. This environmental technology verification report (ETVR) presents information relative to the SEFA-P. Separate ETVRs will be published for the other analyzers that were demonstrated.

The target analytes for this demonstration were selected from the Resource Conservation and Recovery Act's (RCRA) Toxicity Characteristic (TC) list, analytes known to have a high aquatic toxicity

and analytes likely to produce interferences for the FPXRF analyzers. The primary analytes for these comparisons were arsenic, barium, chromium, copper, lead, and zinc; nickel, iron, cadmium, and antimony were secondary analytes. In this demonstration, the SEFA-P reported all of the target analytes.

To demonstrate the analyzers, hazardous waste sites in Iowa (the RV Hopkins site) and in Washington (the ASARCO site) were selected. These sites were chosen because they exhibit a wide range of concentrations for most of the target analytes, are located in different climatological regions of the United States, and combined they exhibit three distinct soil textures: sand, loam, and clay.

This demonstration found that the SEFA-P Analyzer was simple to operate. It was designed to analyze prepared samples. The developer provided a training course for the technology operator that was similar to that which would be provided to a purchaser of the equipment. The training encompassed enough FPXRF theory and hands-on analyzer use to allow the operator to manipulate the data collection software, calibrate the analyzer, and adjust instrument parameters such as count times and target analytes.

The SEFA-P reported the analysis of 100 soil samples. A serious failure occurred at the start of the demonstration and data could only be obtained when NERL-ESD donated its SEFA-P Analyzer for demonstration. Out of concern for including all the original developers and due to time and budgetary constraints, it was decided to test the SEFA-P on a reduced data set after the formal field test was completed. The analyzer was field portable, and it could operate on battery power or on alternating current (AC). The analyzer required an auxiliary computer to process and store data. Downloading of data to both paper and electronic format was accomplished without difficulty.

The SEFA-P Analyzer uses three radioactive sources and, therefore, can quantitate a wide range of analytes. The analyzer reported all of the target analytes for this demonstration (arsenic, barium, copper, cadmium, chromium, lead, antimony, nickel, iron, and zinc). It can use different count times for each source. During the demonstration, the analyzer's throughput averaged 7 to 8 samples per hour.

The analyzer used both Compton and empirical calibrations to quantitate analyte concentrations for this demonstration. Generally, the Compton-based data was more comparable to the reference data than the empirical calibration-based data. Based on the Compton calibrations, the SEFA-P Analyzer produced data meeting definitive level (equivalent to reference data) quality criteria for lead, copper, and antimony and produced quantitative screening level (not equivalent to reference data but correctable with confirmatory sample analysis) quality data for barium. The subset of samples analyzed in this abbreviated demonstration did not provide sufficient data to recommend a quality level for the remaining six target analytes.

The SEFA-P Analyzer exhibited precision similar to the reference methods. Field-based method detection limits (MDL) were higher than the precision-based or developer-provided MDLs. Site or soil texture did not affect data comparability. Correcting the raw data using regression analysis resulted in a measurable improvement in average relative bias and average relative accuracy for the analyzer.

Based on the performance of the SEFA-P Analyzer, this demonstration found it to be an effective tool for characterizing the concentration of select metals in soil samples. As with all FPXRF analyzers, unless a user has regulatory approval, confirmatory (reference) sampling and data correction is recommended when using this analyzer for site characterization and remediation monitoring.

Section 2
Introduction

This environmental technology verification report (ETVR) presents information from the demonstration of the HNU Systems SEFA-P Analyzer. This analyzer was developed by HNU Systems, Inc. to perform elemental analyses (metals quantitation) in the petroleum and petrochemical industry, the mining and minerals industry, and the environmental field. This analyzer uses a silicon drifted-lithium (SiLi) detector with multiple radioactive sources to quantitate metals concentrations. The analyzer is designed to operate in an intrusive mode. In this mode of operation, a soil sample is physically collected, homogenized, dried or sieved, and placed into a sample cup. The sample cup is then placed into an analysis chamber on the probe and a measurement is taken.

This section provides general information about the demonstration including the purpose, objectives, and design. Section 3 presents and discusses the quality of the data produced by the reference methods against which the analyzer was evaluated. Section 4 discusses the HNU SEFA-P Analyzer's capabilities, reliability, throughput, accuracy, precision, comparability to reference methods, and other performance factors. Section 5 discusses the potential applications of the analyzer, presents a method for data correction, and suggests a framework for a standard operating procedure (SOP). Section 6 lists references cited in this ETVR.

Demonstration Background, Purpose, and Objectives

The demonstration was conducted under the Monitoring and Measurement Technologies Program (MMTP), a component of the SITE Program. MMTP is managed by NERL-ESD, Las Vegas, Nevada. The goal of the MMTP is to identify and demonstrate new, innovative, and commercially available technologies that can sample, identify, quantify, or monitor changes in contaminants at hazardous waste sites. This includes those technologies that can be used to determine the physical characteristics of a site more economically, efficiently, and safely than conventional technologies. The SITE Program is administered by the National Risk Management Research Laboratory, Cincinnati, Ohio.

The purpose of this demonstration was to provide the information needed to fairly and thoroughly evaluate the performance of FPXRF analyzers to identify and quantify metals in soils. The primary objectives were to evaluate FPXRF analyzers in the following areas: (1) their accuracy and precision relative to conventional analytical methods; (2) the influence of sample matrix variations (texture, moisture, heterogeneity, and chemical composition) on their performance; (3) the logistical and economic resources needed to operate these analyzers; and (4) to test and validate an SW-846 draft method for FPXRF analysis.

Secondary objectives for this demonstration were to evaluate FPXRF analyzers for their reliability, ruggedness, cost, range of usefulness, data quality, and ease of operation. The performances of the FPXRF analyzers were not compared against each other. Instead, the performance of each analyzer was independently and individually compared to that of conventional analytical methods commonly used in regulatory enforcement or compliance activities. In addition, each analyzer's performance was assessed relative to measurements of standard reference materials (SRM), performance evaluation (PE) samples, and other quality control (QC) samples.

A special request was made by Mr. Oliver Fordham, the demonstration's technical advisor, EPA Office of Solid Waste (OSW), for Midwest Research Institute (MRI) to analyze some of the soil samples to validate the performance of draft Method 3052 "Microwave Assisted Acid Digestion of Ash and Other Siliceous Wastes." Thirty percent of the soil samples were extracted using draft Method 3052 and then analyzed by Method 6010A. The data generated from the draft Method 3052 and Method 6010A analysis were not used for comparative purposes to the FPXRF data in this demonstration.

Reference Methods

To assess the performance of each analyzer, FPXRF data were compared to reference data. The reference methods used for this assessment were EPA SW-846 Methods 3050A/6010A, which are considered the standards for metals analysis in soil for environmental applications. For the purposes of this demonstration, the term "reference" was substituted for "confirmatory" since the data were used as a baseline for comparison. In accordance with Federal Acquisition Regulations, the Midwest Research Institute (MRI) was awarded a subcontract to analyze soil samples using the reference methods. The award was made based on MRI's costs, ability to meet the demonstration's quality assurance project plan (QAPP) requirements, and as the only commercial laboratory identified that could perform all the analyses in the required timeframe.

Method 3050A is the standard acid extraction procedure for determining metals concentrations in soil samples. It is not a total digestion method, and it potentially does not extract all the metals in a soil sample. Method 6010A is the standard method used to analyze Method 3050A extracts. Both of these methods are described in Section 3.

High quality, well documented reference laboratory results were essential for meeting the objectives of the demonstration. For a true and accurate assessment, the reference methods had to provide a known level of data quality. For all measurement and monitoring activities conducted by the EPA, the Agency requires that data quality parameters be established based on the end uses of the data. Data quality parameters usually include five indicators of data quality referred to as the PARCC parameters: precision, accuracy, representativeness, completeness, and comparability. In addition, method detection limits (MDLs) are often used to assess data quality.

Reference methods were evaluated using the PARCC parameters to establish the quality of data generated and to ensure that the comparison of FPXRF analyzers to reference methods was acceptable. The following paragraphs provide definitions of each of the PARCC parameters.

Precision refers to the degree of mutual agreement between replicate measurements and provides an estimate of random error. Precision is often expressed in terms of relative standard deviation (RSD) between replicate samples. The term relative percent difference (RPD) is used to provide this estimate of random error between duplicate samples.

Accuracy refers to the difference between a sample result and the reference or true value. Bias, a measure of the departure from perfect accuracy, can be estimated from the reference or true value. Accuracy and bias for the reference laboratory were assessed by evaluating calibration standard linearity, method blank results and the percent recoveries of matrix spike samples, laboratory control samples (LCS), standard reference materials (SRMs), and PE samples.

Representativeness refers to the degree to which data accurately and precisely measures the conditions and characteristics of the parameter of interest. Representativeness for the reference laboratory was ensured by executing consistent sample collection procedures including sample locations, sampling procedures, sample storage, sample packaging, sample shipping, sampling equipment decontamination, and proper laboratory sample handling procedures. Representativeness was ensured by using the appropriate reference method at its optimum capability to provide results that represented the most accurate and precise measurement it was capable of achieving. The combination of the existing method requirements supplemented by the demonstration QAPP provided the guidance to assure optimum performance of the method. Representativeness was assessed by evaluating calibration standards, method blank samples, duplicate samples, and PE samples.

Completeness refers to the amount of data collected from a measurement process compared to the amount that was expected to be obtained. For the reference data, completeness referred to the proportion of valid, acceptable data generated.

Comparability refers to the confidence with which one data set can be compared to another. Data generated from the reference methods should provide comparable data to any other laboratory performing analysis of the same samples with the same analytical methods. Comparability for the reference methods was achieved through the use of standard operating procedures (SOPs), EPA-published analytical methods, and the demonstration QAPP. QC samples that were used to evaluate comparability include: calibration standards, method blank samples, matrix spike samples, replicate samples, LCSs, SRMs, and PE samples.

Site Selection

PRC conducted a search for suitable demonstration sites between September and November 1994. The following criteria were used to select appropriate sites:

- The site owner had to agree to allow access for the demonstration.
- The site had to have soil contaminated with some or all of the target heavy metals. (Slag, ash, and other deposits of mineralized metals would not be assessed during the demonstration.)
- The site had to be accessible to two-wheel drive vehicles.
- The site had to exhibit one or more of the following soil textures: sand, clay, or loam.
- The site had to exhibit surface soil contamination.
- The sites had to be situated in different climatological environments.

PRC contacted NERL-ESD, regional EPA offices, state environmental agencies, metals fabrication, and smelting contacts to create an initial list of potential demonstration sites. PRC received considerable assistance from the EPA RCRA and Superfund branches in Regions 4, 6, 7, 8, 9, and 10. PRC also

contacted the Montana Department of Health and Environment, the Nevada Bureau of Mines and Geology, the Oklahoma Department of Environmental Quality, the Arizona Department of Environmental Quality, the Missouri Department of Natural Resources, the Arizona Bureau of Geology, and the New Mexico Bureau of Mines and Mineral Resources. PRC surveyed its offices in Kansas City, Kansas; Atlanta, Georgia; Denver, Colorado; Dallas, Texas; Albuquerque, New Mexico; Helena, Montana; Chicago, Illinois; Seattle, Washington; and San Francisco, California, for information regarding potential sites. These PRC offices have existing RCRA, Superfund, or Navy environmental contracts that allow access to regional, state, and federal site information. PRC also used the Record of Decision Scan database (Morgan and others 1993) to search for appropriate sites.

PRC screened 46 potential sites based on the site-selection criteria with the assistance of the various contacts listed above. Based on this screening, PRC and EPA determined that the RV Hopkins and ASARCO sites met most of the site-selection criteria, and therefore, would be the sites used for the demonstration.

The ASARCO site consists of 67 acres of land adjacent to Commencement Bay. The site is marked by steep slopes leading into the bay, a slag fill that was used to extend the original shoreline, a cooling water pond, and the various buildings associated with the smelting process. Partial facility demolition was conducted in 1987. Most of the buildings were demolished between 1993 and 1994. The only buildings remaining are the Fine Ore Building, the Administrative Building, and a Maintenance Garage.

Past soil sampling results targeted four general areas of the site as acceptable candidates for this demonstration: the plant administration area, the former cooling pond, the 1987 demolition area, and certain off-site residential areas adjacent to the smelter stack. Previous sampling has shown surficial soils to be more contaminated than subsurface soils. Arsenic, copper, and lead are the predominant contaminants in the local soils. The highest arsenic concentrations were found in the soils around the former arsenic kitchen, along with cadmium and mercury. The soils around the former cooling pond contained the highest copper concentrations and high levels of silver, selenium, barium, and chromium. Lead concentrations are highest northeast of the arsenic plant.

Much of the smelter site is covered with artificial fill material of varying thickness and composition. Two general types of fill are found on-site: granular and slag. The composition of the granular fill material ranges from sand to silt with demolition debris and slag debris mixed throughout. The slag fill is a solid, fractured media restricted to the plant site. The surface soil in the plant administration area has a layer of slag particles on top, ranging from 1 to 3 inches thick. Surficial material in the parking lot area and southwest of the stack is mostly of glacial origin and composed of various mixtures of sand, gravel, and cobbles. The soils around the former cooling pond are fine-grained lacustrine silts and clays. Alluvium upgradient of the former cooling pond has been almost entirely covered with granular fill material. Generally, soils in the arsenic kitchen and stack hill areas are sand mixed with gravel or sandy clay mixed with cobbles. No slag was analyzed as part of this demonstration.

The RV Hopkins site is located in the west end of Davenport, Iowa. The facility occupies approximately 6.7 acres in a heavy industrial/commercial zoned area. Industrial activities in the area of the RV Hopkins property included the manufacture of railroad locomotive engines during the mid-1800's. The RV Hopkins property was a rock quarry during the late 1800's. Aerial surveys beginning in 1929 show that the rock quarry occupied the majority of the site initially, gradually decreasing until it was completely filled by 1982. It was reported that the site was used to dispose of demolition debris, automotive, and scrap metal. The site also has been used by a company that recycled lead acid batteries.

RV Hopkins began operating as a drum reconditioner in 1951 across the street from its current location. In 1964, the site owner reportedly covered the former quarry area of the site with foundry sand. No foundry sand was analyzed as part of this demonstration. RV Hopkins receives between 400 and 600 drums per day for reconditioning, accepting only drums that meet the definition of "empty" according to 40 Code of Federal Regulations 261.7. Most of the drums received at the facility come from the paint, oil, and chemical industries. The surrounding area is reported to be underlain by Devonian-aged Wapsipinicon Limestone, and gray-green shale, lime mud, and sand stringers dating back to the Pennsylvanian age.

The RV Hopkins property is composed of five buildings: the office and warehouse, a warehouse used to store drums of hazardous waste and a waste pile, a manufacturing building, a drum reclamation furnace, and a cutting shed. The office and the warehouse are located on the southwest corner of the site. Areas investigated on each site include the furnace area, the old and new baghouses, the former drum storage area on the north end of the facility, the former landfill, and a drainage ditch. Major contaminants include barium, lead, chromium, and zinc, as well as lesser concentrations of other metals, such as copper and nickel, pesticides, and volatile organic compounds.

Based on historical data, the most concentrated contaminants in the furnace area are chromium, lead, and zinc. The highest concentrations of these elements are at the furnace entrance, as opposed to the furnace exit. The concentrations of lead are higher in the old baghouse than in the new, while the new baghouse exhibits a higher concentration of chromium, as well as high iron, lead, and barium concentrations. The former landfill has concentrations of barium, chromium, lead, nickel, and zinc greater than 1,000 mg/kg. Lead is the most prevalent contaminant in the former drum storage area with lesser concentrations of barium, chromium, and zinc.

Predemonstration Sampling

Predemonstration sampling was conducted at both sites between December 5 and 14, 1994. These sampling events had the following objectives:

- To provide data on, or verify, the extent of surface contamination at each site and to locate optimum sampling areas for the demonstration.
- To allow the developers to analyze samples from the demonstration sites in advance of the demonstration, and if necessary, refine and recalibrate their technologies and revise their operating instructions.
- To evaluate samples for the presence of any unanticipated matrix effects or interferences that might occur during the demonstration.
- To check the quality assurance (QA) and QC procedures of the reference laboratory.

One hundred soil samples were analyzed on each site by the FPXRF analyzers during the predemonstration sampling activities. The samples represented a wide range in the concentration of metals and soil textures. Thirty-nine samples were submitted for reference method analysis using EPA SW-846 Methods 3050A/6010A. Twenty-nine of these samples were split and sent to the developers. Nine field duplicates were collected and submitted for reference method analysis to assess proposed sample homogenization procedures. One purchased PE sample also was submitted to the reference laboratory to provide an initial check of its accuracy.

Additionally, three samples representing low, medium, and high concentrations were collected at each site. These samples were dried, ground, and then analyzed by six independent laboratories before

the demonstration began to create site-specific PE samples. These samples were analyzed with laboratory-grade X-ray fluorescence (XRF) analyzers.

Experimental Design

The experimental design of this demonstration was developed to meet the primary and secondary objectives stated above, and was approved by all demonstration participants prior to the start of the demonstration. The design is detailed in the demonstration plan and is summarized below.

Approximately 100 soil samples were collected from each of three target soil textures: clay, loam, and sand. This variety of soil textures allowed the examination of the effect of soil texture on data comparability. Splits of these samples were analyzed by all of the FPXRFs and by the reference methods.

The SEFA-P Analyzer is designed to operate in the intrusive mode. The overall sampling and analysis procedure was designed to reflect a common application of FPXRF analyzers. The sample preparation procedure used in this demonstration is illustrated in Figure 2-1. Since this instrument operates in the intrusive mode only, this discussion of the experimental design will be limited to intrusive sample preparation and analysis. Furthermore, of the two intrusive sample preparations, the SEFA-P only analyzed samples from the final preparation step (intrusive-prepared).

After sample collection, homogenization was monitored by adding 1 to 2 grams of sodium fluorescein salt (which fluoresced when it was exposed to ultraviolet light) to the sample bag. During the predemonstration, it was determined that sodium fluorescein did not affect the FPXRF or reference method analyses. Sample homogenization took place by kneading the sample and sodium fluorescein salt in a plastic bag for 2 minutes. After this period, the sample preparation technician examined the sample under ultraviolet light to assess the distribution of sodium fluorescein throughout the sample. If the fluorescein salt was not evenly distributed, the homogenization and checking process were repeated until the sodium fluorescein was evenly distributed. This monitoring process assumed that even distribution of sodium fluorescein was indicative of good sample homogenization. The effectiveness of this homogenization is discussed later in this section.

The homogenized sample material was then passed through a No. 10 mesh sieve (2-millimeter (mm) openings) and approximately 10 grams of this material was placed in a sample cup for analysis as an intrusive-unprepared sample. Replicate measurements were collected from 4 percent of these samples to assess analyzer precision. These data represented FPXRF intrusive mode measurements on soils with no sample preparation (intrusive-unprepared). Sample material from this preparation step was submitted to the reference laboratory for analysis.

Following the intrusive-unprepared procedure, a portion of the soil sample was dried in a convection oven at 110 °C for 1 hour and ground with a mortar and pestle until it passed through a No. 40 stainless-steel sieve (0.425-mm openings). These samples were then analyzed as intrusive-prepared. Four percent of these samples underwent replicate measurements to evaluate analyzer precision. These replicate measurements were made on the same soils as in the intrusive-unprepared precision measurements. This data represented FPXRF intrusive measurements on prepared soils (intrusive-prepared). One hundred of these samples were the only ones analyzed by the SEFA-P.

These preparation procedures allowed for the evaluation of the effects of sample preparation on FPXRF comparability to reference data.

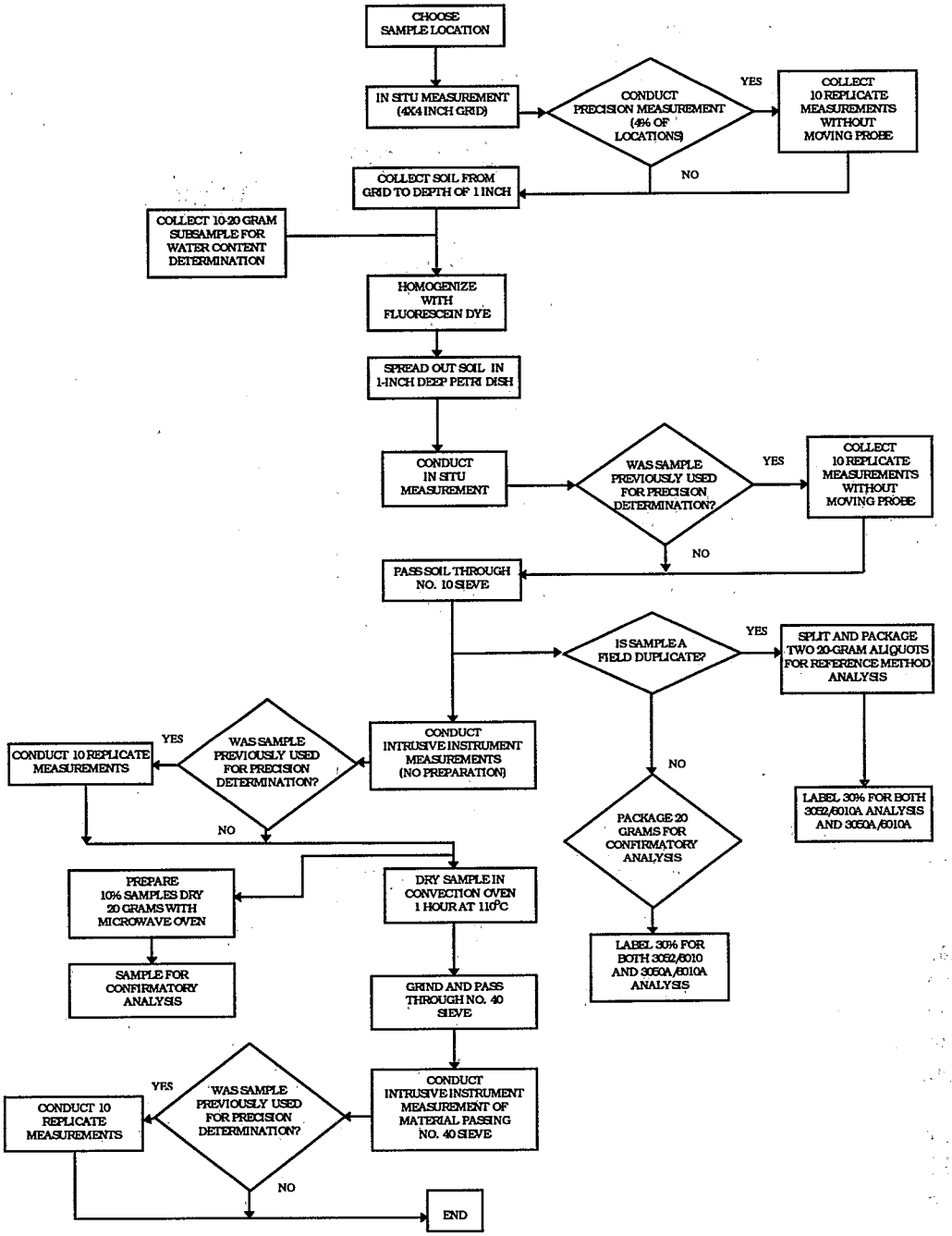


Figure 2-1. Sample Preparation and Analysis: This flowchart depicts the handling procedures for each sample taken during the demonstration.

Qualitative Factors

There are a number of factors important to data collection that are difficult to quantify and must be evaluated qualitatively. These are considered qualitative factors. One such factor was the amount of training required to operate a given FPXRF analyzer. To assess this factor, PRC operators were trained by the developers on how to operate their respective FPXRF analyzers. All operators met or exceeded the developers' minimum requirements for education and previous experience. Demonstration procedures were designed to simulate routine field conditions as closely as possible. The developers trained the operators using their respective operator training manuals. Based on this training and field experience, the operators prepared a subjective evaluation assessing the training and technology operation during the demonstration (Section 4).

Many analytical methods exhibit significant "operator effects," in which individual differences in sample preparation or operator technique result in a significant effect on the numerical results. To reduce the possible influence of operator effects, a single operator was used to operate each FPXRF analyzer. While this reduced some potential error from the evaluation, it did not allow the analyzers to be evaluated for their susceptibility to operator-induced error. A single operator was used to analyze all of the samples at both sites during this demonstration. Sample preparation variation effects were minimized in the field by using the same personnel to prepare samples. To eliminate the influence of operator effects on the reference method analysis, only one reference laboratory was used to analyze the samples. Based on this design, there can be no quantitative estimate of the "operator" effect.

Quantitative Factors

Many factors in this demonstration could be quantified by various means. Examples of quantitative factors evaluated during this demonstration include analyzer performance near regulatory action levels, effects of sample preparation, effects of microwave sample drying, count times, health and safety considerations, costs, and interferences.

The data developed by the FPXRF analyzers were compared to reference data for the following primary analytes: arsenic, barium, chromium, copper, lead, and zinc; and for the following secondary analytes: nickel, iron, cadmium, and antimony. The SEFA-P reported all of these analytes.

Evaluations of analyzer data comparability involved examining the effects of each site, soil texture, and sample preparation technique (Table 2-1). Two sites were sampled for this demonstration and therefore two site variables were examined (RV Hopkins and ASARCO sites). These sites produced samples from three distinct soil textures, and therefore, three soil variables were examined (clays, sands, and loams). Four sample preparation steps were used: (1) *in situ*-unprepared, (2) *in situ*-prepared, (3) intrusive-unprepared, and (4) intrusive-prepared. These variables were nested as follows: each site was divided into RV Hopkins and ASARCO data sets; the RV Hopkins data represented the clay soil texture, and the ASARCO data was divided into sand and loam soil textures; then each soil texture was subdivided by the four soil preparations. These variables allowed the examination of particle size and homogenization effects on data comparability. These effects were believed to have the greatest potential impact on data comparability.

Of greatest interest to users is analyzer performance near action levels. For this reason, samples were approximately distributed as follows: 25 percent in the 0 - 100 mg/kg range, 50 percent in the 100 - 1,000 mg/kg range, and 25 percent in the greater than 1,000 mg/kg range. The lower range tested analyzer performance near MDLs; the middle range tested analyzer performance in the range of many

action levels for inorganic contaminants; and the higher range tested analyzer performance on grossly contaminated soils. All samples collected for the demonstration were split between the FPXRF analyzers and reference laboratory for analysis. Metal concentrations measured using the reference methods were considered to represent the "true" concentrations in each sample. Where duplicate samples existed, concentrations for the duplicates were averaged and the average concentration was considered to represent the true value for the sample pair. This was specified in the demonstration plan. If one or both samples in a duplicate pair exhibited a nondetect for a particular target analyte, that pair of data was not used in the statistical evaluation of that analyte. The reference methods reported measurable concentrations of target analytes in all of the samples analyzed.

Table 2-1. Performance and Comparability Variables Evaluated

Variables		
Site Name (100)	Soil Texture (100)	Preparation Step [237]
ASARCO (68)	Sand (31)	intrusive-prepared [79[a]]
	Loam (37)	intrusive-prepared [79[a]]
RV Hopkins (32)	Clay (32)	intrusive-prepared [79[a]]

Notes: [a] Includes the PE samples and SRMs.

( ) Total number of sample points.

[ ] Total number of measurements taken.

In addition to the quantitative factors discussed above, the common FPXRF sample preparation technique of microwave drying of samples was evaluated. Sample temperatures during this procedure can be high enough to melt some mineral fractions in the sample or combust organic matter. Several metals that present environmental hazards can volatilize at elevated temperatures. Arsenic sublimes at 188 °C, within the potential temperature range achieved during microwave drying of samples. To assess this effect, 10 percent of the homogenized, crushed, oven-dried, and sieved samples were split and heated in a microwave oven on high for 3 minutes. This time was chosen to approximate common microwave drying times used in the field. These split samples were then submitted for reference analysis. The reference data for these samples were compared to the corresponding reference data produced from the convection oven-dried sample. These data showed the effects of the microwave drying variable on analyte concentration. This was a minor variable and it was only evaluated for the reference laboratory in an attempt to identify any effect on data comparability.

Another quantitative variable evaluated was the count time used to acquire data. During the formal sample quantitation and precision measurement phase of the demonstration, the count times were set by the developers and remained constant throughout the demonstration. Count times can be tailored to produce the best results for specific target analytes. The developers, however, selected count times that produced the best compromise of results for the entire suite of target analytes. To allow a preliminary assessment of the effect of count times, select soil samples were analyzed in replicate using count times longer and shorter than those set by the developers. This allowed the evaluation of the effects of count times on analyzer performance. Since sample throughput can be affected by adjusting count times, operators used only the developer-specified count times throughout the demonstration.

An important health and safety issue during the demonstration was the effectiveness of radioactivity shielding of each FPXRF analyzer. Occasional radiation readings were quantitatively made with a gamma ray detector near each analyzer to assess the potential for exposure to radiation.

A compilation of the costs associated with the use of each FPXRF analyzer was another important evaluation factor. Cost includes analyzer purchase or rental, expendable supplies, such as liquid nitrogen and sample cups, and nonexpendable costs, such as labor, licensing agreements for the radioactive sources, operator training costs, and disposal of investigation-derived waste (IDW). This information is provided to assist a user in developing a project cost analysis.

Factors that could have affected the quantitative evaluations included interference effects and matrix effects. Some of these effects and the procedures used to evaluate their influence during this demonstration are summarized below:

- Heterogeneity: For *in situ*-unprepared measurements, heterogeneity was partially controlled by restricting measurements within a 4-by-4-inch area. For measurements after the initial point-and-shoot preparation, heterogeneity was minimized by sample homogenization. This effect was evaluated through the sample preparation data.
- Particle Size: The effect of particle size was evaluated with the two intrusive sample preparations. Theoretically, precision and accuracy should increase as particle size decreases and becomes uniform. Since only the intrusive samples were analyzed, this factor was not evaluated.
- Moisture Content: It has been suggested that major shifts in sample moisture content can affect a sample's relative fluorescence. This effect could not be evaluated as thoroughly as planned because of the small difference in sample moisture content observed at the two sites.
- Overlapping Spectra of Elements: Interferences result from overlapping spectra of metals that emit X-rays with similar energy levels. The reference method analysis provided data on the concentration of potential interferants in each sample.

Evaluation of Analyzer Performance

Metals concentrations measured by each analyzer were compared to the corresponding reference laboratory data, and to other QA/QC sample results. These comparisons were conducted independently for each target analyte. These measurements were used to determine an analyzer's accuracy, data quality level, method precision, and comparability to reference methods. PE samples and SRM samples were used to assess analyzer accuracy. Relative standard deviations (RSD) on replicate measurements were used to determine analyzer precision. These data were also used to help determine the data quality of each FPXRF analyzer's output. The data comparability and quality determination was primarily based on a comparison of the analyzer's data and the reference data. Linear regression and a matched pairs t-test were the statistical tools used to assess comparability and data quality.

A principal goal of this demonstration was the comparison of FPXRF data and the reference laboratory data. EPA SW-846 Methods 3050A/6010A were selected as the reference methods because they represent the regulatory standard against which FPXRF is generally compared. In comparing the FPXRF data and reference data, it is important to recognize that, while similar, the process by which the data are obtained is not identical. While there is significant overlap in the nature of the samples being measured, there are also major differences. These differences, or "perspectives," allow the user to characterize the same sample in slightly different ways. Both have a role in site characterization and remediation. It is important to consider these differences and the measurement error intrinsic to each method when comparing the FPXRF method against a reference analytical method.

The reference methods chosen for this purpose involve wet chemical analysis and partial digestion of approximately 1 to 2 grams of sample (approximately 0.25 cubic centimeters (cm^3) depending on sample bulk density). The digestion process extracts the most acid-soluble portion of the sample, which represents the material from most surfaces, and clay and carbonate minerals. Since the digestion is not complete, the less acid-soluble components are not digested and are not included in the analysis. These components may include the coarser-grained quartz, feldspar, lithic components, and certain metal complexes. In contrast, FPXRF analyzers generally produce X-ray excitation in an area of approximately 3 centimeters squared (cm^2) to a depth of approximately 2.5 centimeters (cm). This equates to a sample volume of approximately 7.5 cm^3. X-rays returning to the detector are derived from all matrix material including the larger-grained quartz, feldspar, lithic minerals, metal complexes, and organics. Because the FPXRF method analyzes all material, it represents a total analysis in contrast to the reference methods, which represent a select or partial analysis. This difference can result in FPXRF concentrations that are higher than corresponding reference data when metals are contained within nonacid soluble complexes or constituents. It is important to note that if metals are contained in nonacid soluble complexes, a difference between the FPXRF analyzers and the reference methods is not necessarily due to error in the FPXRF method but rather to the inherent differences in the nature of the analytical methods.

The comparison of FPXRF data and the reference data employs linear regression as the primary statistical tool. Linear regression analysis intrinsically contains assumptions and conditions that must be valid for the data set. Three important assumptions to consider include: (1) the linearity of the relationship, (2) the confidence interval and constant error variance, and (3) an insignificant measurement error for the independent variable (reference data).

The first assumption requires that the independent variable (reference data) and the dependent variable (FPXRF data) are linearly related and are not described by some curvilinear or more complex relationship. This linearity condition applies to either the raw data or mathematical transformations of the raw data. Figure 2-2 illustrates that FPXRF data and reference data are, in fact, related linearly and that this assumption is correct.

The second assumption requires that the error be normally distributed, the sum to equal zero, be independent, and exhibit a constant error variance for the data set. Figure 2-2 illustrates that for raw data, this assumption is not correct (at higher concentrations the scatter around the regression line increases), but that for the logarithmic transformation (shown as a log-log plot) of the data, this assumption is valid (the scatter around the regression line is relatively uniform over the entire concentration range). The change in error distribution (scatter) evident in the untransformed data results in the disproportionate influence of large data values compared with small data values on the regression analysis.

The use of least squares linear regression has certain limitations. Least squares regression provides a linear equation, which minimizes the squares of the differences between the dependent variable and the regression line. For data sets produced in this demonstration, the variance was proportional to the magnitude of the measurements. That is, a measurement of 100 parts per million (ppm) may exhibit a 10 percent variance of 10 ppm, while a 1,000 ppm measurement exhibits a 10 percent variance of 100 ppm. For data sets with a large range in values, the largest measurements in a data set exert disproportionate influence on the regression analysis because the least squares regression must account for the variance associated with the higher valued measurements. This can result in an equation that has minimized error for high values, but almost neglects error for low values because their influence in minimizing dependent variable error is small or negligible. In some cases, the resulting equations, biased by high-value data, may lead to inappropriate conclusions concerning data quality. The range of the data examined for the

analyzers spanned between 1 and 5 orders of magnitude (e.g., 10 - 100,000 ppm) for the target analytes. This wide range in values and the associated wide range in variance (influenced by concentration) created the potential for this problem to occur in the demonstration data set. To provide a correlation that was equally influenced by both high and low values, logarithms (log_{10}) of the dependent and independent variables were used, thus, scaling the concentration measurements and providing equal weight in the least squares regression analysis to both small and large values (Figure 2-2). All statistical evaluations were carried out on log_{10} transformed data.

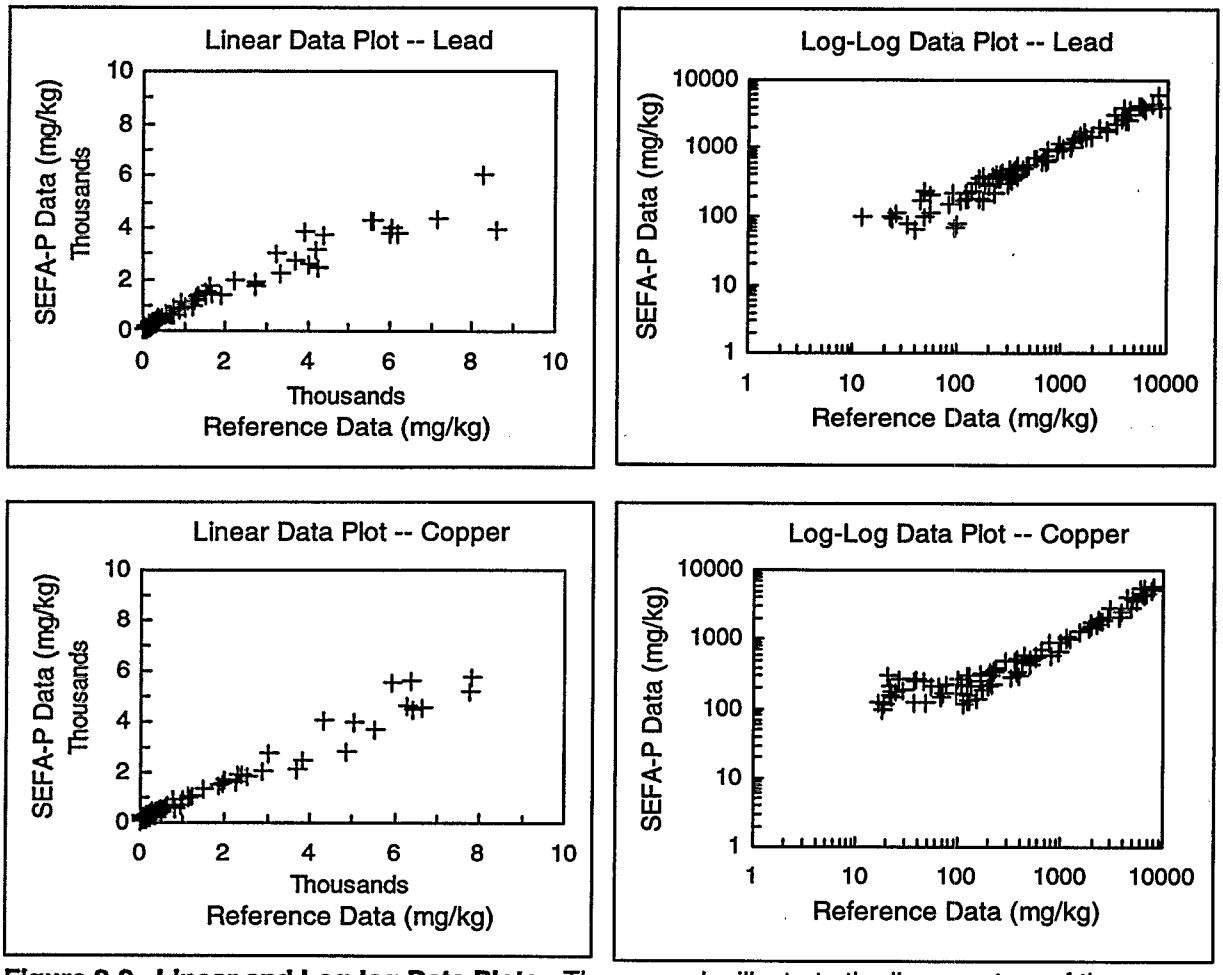


Figure 2-2. Linear and Log-log Data Plots: These graphs illustrate the linear nature of the relationship between the FPXRF data and the reference data. The linear data plots illustrate the concentration dependence of this relationship with increased scatter at higher concentrations. The log-log plots eliminate this concentration dependence effect. Scatter is relatively constant over the entire plot.

The third assumption, requiring an insignificant measurement error in the reference data, was not true for all analytes. The consequences of measurement error varied depending on whether the error is caused by the reference methods or the FPXRF method. If the error is random or if the error for the reference methods is small compared to the total regression error, then conventional regression analysis can be performed and the error becomes a part of the random error term of the regression model. This error (based on the log_{10} transformed data) is shown in the regression summary tables in Section 4 as the

"standard error." In this case, deviations from perfect comparability can be tied to an analyzer's performance. If the error for the reference methods is large compared to the total error for the correlation of the FPXRF and the reference data, then deviations from perfect comparability might be due in part to measurement error in the reference methods.

It is a reasonable assumption that any measurement errors in either the reference or FPXRF methods are independent of each other. This assumption applies to either the raw data or the $\log_{10}$ transformed data. Given this assumption, the total regression error is approximately the sum of the measurement error associated with the reference methods and the measurement error associated with the FPXRF method. The reference methods' precision is a measure of independent variable error, and the mean square error expressed in the regression analysis is a relative measure of the total regression error that was determined during the regression analysis. Precision data for the reference methods, obtained from RPD analyses on the duplicate samples from each site, for each analyte, indicated the error for the reference methods was less than 10 percent of the total regression error for the target analytes. Subsequently, 90 percent of the total measurement error can be attributed to measurement error associated with the analyzers. Based on this interpretation, the reference data does allow unambiguous resolution of data quality.

The comparison of the reference data to the FPXRF data is referred to as an intermethod comparison. All reference and QA/QC data were generated using an EPA-approved definitive level analytical method. If the data obtained by an analyzer were statistically similar to the reference methods, the analyzer was considered capable of producing definitive level data. As the statistical significance of the comparability decreased, an analyzer was considered to produce data of a correspondingly lower quality. Table 2-2 defines the criteria that determined the analyzer's level of data quality (EPA 1993).

Table 2-2. Criteria for Characterizing Data Quality

Data Quality Level	Statistical Parameter[a,b]
Definitive Level	r^2 = 0.85 to 1.0. The precision (RSD) must be less than or equal to 10 percent and inferential statistics must indicate that the two data sets are statistically similar.
Quantitative Screening Level	r^2 = 0.70 to 1.0. The precision (RSD) must be less than 20 percent but the inferential statistics indicate that the data sets are statistically different.
Qualitative Screening	r^2 = less than 0.70. The precision (RSD) is greater than 20 percent. The data must have less than a 10 percent false negative rate.

Notes:
[a] The statistical tests and parameters are discussed later in the "Intermethod Comparison" subsection in Section 4.
[b] The regression parameters apply to either raw or $\log_{10}$ transformed data sets. The precision criteria apply to only the raw data.
r^2 Coefficient of determination.
RSD Relative standard deviation.

Data from this demonstration were used to assign analyzer data into one of three data quality levels as follows: (1) definitive, (2) quantitative screening, and (3) qualitative screening. The first two data quality levels are defined in EPA guidance (1993). The qualitative screening level criteria were defined in the demonstration plan (PRC 1995) to further differentiate the screening level data.

Definitive level data are considered the highest level of quality. These data are usually generated by using rigorous, well-defined analytical methods such as those approved by the EPA or ASTM. The data

is analyte-specific with full confirmation of analyte identity and concentration. In addition, either analytical or total measurement error can be determined. Definitive data may be generated in the field, as long as the QA/QC requirements are satisfied.

Quantitative screening data provides confirmed analyte identification and quantification, although the quantification may be relatively imprecise. It is commonly recommended that at least 10 percent of the screening data be confirmed using analytical methods and QA/QC procedures and criteria associated with definitive data. The quality of unconfirmed screening data cannot be determined.

Qualitative screening level data indicates the presence or absence of contaminants in a sample, but does not provide reliable concentration estimates. The data may be compound-specific or specific to classes of contaminants. Generally, confirmatory sampling is not required if an analyzer's operation is verified with one or more check samples.

At the time of this demonstration, an approved EPA method for FPXRF did not exist. As part of this study, PRC prepared a draft Method 6200 "Field Portable X-Ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment." The draft method has been submitted for inclusion in Update 4 of SW-846 scheduled for approval in FY-97. For purposes of this demonstration, the absence of a current EPA-approved final method did not preclude the analyzers' data from being considered definitive. The main criterion for data quality level determination was the comparability of each analyzer's data to that produced by the reference methods, as well as to analyzer-specific criteria such as precision.

The comparability data set for the SEFA-P consisted of 100 matched pairs of FPXRF and reference data. This data set was analyzed as a whole and then subdivided and analyzed with respect to each of the variables listed in Table 2-1. This nesting of variables allowed for an independent assessment of the potential influence of each variable on comparability.

To obtain an adequate data set to evaluate the performance of the analyzers, a total of 315 soil samples was analyzed by the reference laboratory. These samples were to be analyzed by the FPXRF analyzers for each of the four sample preparation steps. This produced 1,260 data values for each analyzer, 630 in each mode in-situ or intrusive. Seventy of the 315 samples submitted to the reference laboratory were split and submitted as field duplicates to assess the sample homogenization process. Thirty-three of the 315 samples were also split and microwave-dried, then submitted for reference method analysis to assess the effect of microwave drying. Of the 315 samples submitted for reference method analysis, 215 were collected from the ASARCO site and 100 were collected from the RV Hopkins site. Approximately twice as many samples were collected at the ASARCO site because two of the target soil textures (sands and loams) were found there. Only one target soil texture (clay) was found at the RV Hopkins site. Under the abbreviated conditions, the SEFA-P analyzed 31 ASARCO samples from the sand soil, 37 ASARCO samples from the loam soil, and 32 RV Hopkins samples from the clay soil.

Evaluation of the influence of the site and soil variables was limited to the examination of the lead and zinc data. These were the only primary analytes that exhibited a wide distribution of concentrations across all sites and soil textures. The effects of sample preparation variables were evaluated for all target analytes. If the evaluation of the influence of a given variable did not result in a better correlation, as exhibited by a higher coefficient of determination (r^2) and smaller standard error of the estimate (using $\log_{10}$ transformed data), then the influence was considered to be insignificant. However, if the correlation worsened, the cause was examined and explained. If the correlation improved, resulting in a

higher r^2 value, and reduced standard error of the estimate, then the impact of the variable was considered significant. For example, if the r^2 and standard error of the estimate for a given target analyte improved when the data set was divided into the four sample preparation steps, the sample preparation variable was determined to be significant. Once this was determined, the variables of site and soil texture were evaluated for each of the four sample preparations steps. If the site or soil texture variable improved the regression parameters for a given soil preparation, then that variable was also considered significant.

After the significant variables were identified, the impact of analyte concentration was examined. This was accomplished by dividing each variable's $\log_{10}$ transformed data set into three concentration ranges: 0 - 100 mg/kg; 100 - 1,000 mg/kg; and greater than 1,000 mg/kg. A linear regression analysis was then conducted on the three data sets. If this did not result in improved r^2 values and reduced standard errors of the estimate, the relationship between the analyzer's $\log_{10}$ transformed data and the $\log_{10}$ transformed reference data was considered linear over the entire range of concentrations encountered during the demonstration. This would mean that there was no concentration effect.

Numerous statistical tests have been designed to evaluate the significance of differences between two populations. In comparing the performance of the FPXRF analyzers against the reference methods, the linear regression comparison and the paired t-test were considered the optimal statistical tests. The paired t-test provides a classic test for comparing two populations, but is limited to analysis of the average or mean difference between those populations. Linear regression analysis provides information not only about how the two populations compare on average, but also about how they compare over ranges of values. Therefore, this statistical analysis provides information about the structure of the relationship; that is, whether the methods differ at high or low concentrations or both. It also indicates whether the FPXRF data is biased or shifted relative to the reference data.

Linear regression provides an equation that represents a line (Equation 2-1). Five linear regression parameters were considered when assessing the level of data quality produced by the FPXRF analyzers. This assessment was made on the $\log_{10}$ transformed data sets. The five parameters were the y-intercept, the slope of the regression line, standard error of the estimate, the correlation coefficient (r), and r^2. In linear regression analysis, the r provides a measure of the degree or strength of the correlation between the dependent variable ($\log_{10}$ transformed FPXRF data), and the independent variable ($\log_{10}$ transformed reference data). The r^2 provides a measure of the fraction of total variation which is accounted for by the regression relation (Havlick and Crain 1988). That is, it is a measure of the scatter about a regression line and, thus, is a measure of the strength of the linear association.

$$Y = mX + b \qquad \textbf{(2-1)}$$

where

b is the y-intercept of the regression line, m is the slope of the regression line, and Y and X are the $\log_{10}$ transformed and dependent and independent variables, respectiv

Values for r vary from a value of 1 to -1, with either extreme indicating a perfect positive or negative correlation between the independent and dependent variables. A positive correlation coefficient indicates that as the independent variable increases, the dependent variable also increases. A negative correlation coefficient indicates an inverse relationship, as the independent variable increases the dependent variable decreases. An r^2 of 1.0 indicates that the linear equation explains all the variation between the FPXRF and reference data. As the r^2 departs from 1.0 and approaches zero, there is more unexplained variation, due to such influences as lack of association with the dependent variable ($\log_{10}$ transformed FPXRF data), or the influence of other independent variables.

If the regression correlation exhibited an r^2 between 0.85 and 1.0, the FPXRF data was considered to have met the first requirement for definitive level data classification (Table 2-2). The second criteria, precision was then examined and is required to be equal to or less than 10 percent RSD to retain the definitive data quality level for that analyte. If both these criteria are not satisfied, then certain inferential parameters are evaluated. First, the regression line's y-intercept and slope were examined. A slope of 1.0 and a y-intercept of 0.0 would mean that the results of the FPXRF analyzer matched those of the reference laboratory ($\log_{10}$ FPXRF=$\log_{10}$ reference). Theoretically, the more the slope and y-intercept differ from the values of 1.0 and 0.0, respectively, the less accurate the FPXRF analyzer. However, a slope or y-intercept can differ slightly from these values without that difference being statistically significant. To determine whether such differences were statistically significant, the Z test statistics for parallelism and for a common intercept was used at the 95 percent confidence level for the comparison (Equations 2-2 and 2-3) (Kleinbaum and Kupper 1978). These criteria were used to assign data quality levels for each analyte.

The matched pairs t-test was also used to evaluate whether the two sets of $\log_{10}$ transformed data were significantly different. The paired t-test compares data sets, which are composed of matched pairs of data. The significance of the relationship between two matched-pairs sets of data can be determined by comparing the calculated t-statistic with the critical t-value determined from a standard t-distribution table at the desired level of significance and degrees of freedom. To meet definitive level data quality requirements, both the slope and y-intercept had to be statistically the same as their ideal values, as defined in the demonstration plan (PRC 1995), and the data had to be statistically similar as measured by the t-test. Log_{10} transformed data meeting these criteria were considered statistically equivalent to the $\log_{10}$ transformed reference data.

Slope Test for Significant Differences **(2-2)**

$$Z = \frac{m - 1}{\sqrt{SE_m + 0}}$$

where

m is the slope of the regression line, SE is the standard error of the slope, and Z is the normal deviate test statistic.

Y-intercept Test for Significant Differences **(2-3)**

$$Z = \frac{b - 0}{\sqrt{SE_b - 0}}$$

where

b is the y-intercept of the regression equation.

If the r^2 was between 0.70 and 1, the precision (RSD) less than 20 percent, and the slope or intercept were not statistically equivalent to their ideal values, the analyzer was considered to produce quantitative screening level data quality (Table 2-2). In this case, the linear regression is usually sufficiently significant so that bias can be identified and corrected. Therefore quantitative screening data could be

mathematically corrected if a portion (10 - 20 percent) of the samples are sent to a reference laboratory. Laboratory analysis results for these samples would provide a basis for determining a correction factor.

Data placed in the qualitative screening level category exhibit r^2 values less than 0.70. These data either were not statistically similar to the reference data based on inferential statistics or they had a precision greater than 20 percent RSD. An analyzer producing data at this level is considered capable of detecting the presence or lack of contamination, above its detection limit, with at least a 90 percent accuracy rate, but is not considered suitable for reporting of concentrations.

MDLs for the analyzers were determined in two ways. One approach followed standard SW-846 protocol. In this approach, standard deviations (SD) from precision measurements for samples exhibiting contamination 5 to 10 times the estimated detection levels of the analyzers were multiplied by 3. The resultant number represented the precision-based MDL for the analyzers.

In a second approach, MDLs were determined by analysis of the low concentration outliers on the $\log_{10}$ transformed FPXRF and $\log_{10}$ transformed reference method data cross plots. These cross plots for all analytes characteristically exhibited a region below the MDL where the linearity of the relationship disintegrated. Above the MDL, the FPXRF concentrations increased linearly with increasing reference method values. Effectively, the linear correlation between the two methods abruptly changes to no correlation below the MDL. The value of the MDL was assigned by determining the concentration where the linear relationship disintegrates and reporting a value at two SDs above this concentration. This data also represented a portion of the regression line that, if included, resulted in a decrease in the correlation coefficient rather than an increase. This MDL represented a field- or performance-based value.

Deviations from the Demonstration Plan

Seven deviations were made from the demonstration plan during the on-site activities. The first dealt with the determination of the moisture content of the samples. The demonstration plan stated that a portion of the original sample would be used for determining moisture content. Instead, a small portion of soil was collected immediately adjacent to the original sample location and was used for determining moisture content. This was done to conserve sample volume needed for the reference laboratory. The moisture content sample was not put through the homogenizing and sieving steps prior to drying.

The second deviation dealt with the sample drying procedures for moisture content determination. The demonstration plan required that the moisture content samples be dried in a convection oven at 150 °C for 2 hours. Through visual observation, it was found that the samples were completely dried in 1 hour with samples heated to only 110 °C. Therefore, to conserve time, and to reduce the potential for volatilization of metals, the samples for moisture content determination were dried in a convection oven at 110 °C for 1 hour.

The third deviation involved assessing analyzer drift due to changes in temperature. The demonstration plan required that at each site, each analyzer would measure the same SRM or PE sample at 2-hour intervals during at least one day of field operation. However, since ambient air temperature did not fluctuate more than 20 °F on any day throughout the demonstration, potential analyzer drift due to changes in temperature was not assessed.

The fourth deviation involved the drying of samples with a microwave. Instead of microwaving the samples on high for 5 minutes, as described in the demonstration plan, the samples were microwaved on

high for only 3 minutes. This modification was made because the plastic weigh boats, which contained the samples, were melting and burning when left in the microwave for 5 minutes. In addition, many of the samples were melting to form a slag. PRC found (through visual observation) that the samples were completely dry after only 3 minutes of microwaving. This interval is still within common microwave drying times used in the field.

An analysis of the microwaved samples showed that this drying process had a significant impact on the analytical results. The mean RPD for the microwaved and nonmicrowaved data were significantly different at a 95 percent confidence level. This suggests that the microwave drying process somehow increases error and sample concentration variability. This difference may be due to the extreme heat and drying altering the reference methods' extraction efficiency for target analytes. For the evaluation of the effects of microwave drying, there were 736 matched pairs of data where both element measurements were positive. Of these pairs, 471 exhibited RPDs less than 10 percent. This 10 percent level is within the acceptable precision limits for the reference laboratory as defined in the demonstration QAPP. Pairs exhibiting RPDs greater than 10 percent totaled 265. RPDs greater than 10 percent may have causes other than analysis-induced error. Of these 265, 96 pairs indicated an increase in metals concentration with microwaving, and 169 pairs indicated a reduction in the concentration of metals. The RPDs for the microwaved samples were 2 to 3 times worse than the RPDs from the field duplicates. This further supports the hypothesis that microwave drying increases variability.

The fifth deviation involved reducing the percentage of analyzer precision measuring points. The demonstration plan called for 10 percent of the samples to be used for assessment of analyzer precision. Due to the time required to complete analysis of an analyzer precision sample, only 4 percent of the samples were used to assess analyzer precision. This reduction in samples was approved by the EPA technical advisor and the PRC field demonstration team leader. This eliminated 720 precision measurements and saved between 24 and 240 hours of analysis time. The final precision determinations for this demonstration were based on 48 sets of 10 replicate measurements for each analyzer.

The sixth deviation involved method blanks. Method blanks were to be analyzed each day and were to consist of a lithium carbonate that had been used in all sample preparation steps. Each analyzer had its own method blank samples, provided by the developer. Therefore, at the ASARCO site, each analyzer used its own method blank samples. However, at the RV Hopkins site, each analyzer used lithium carbonate method blanks that were prepared in the field, in addition to its own method blank samples. Both types of method blank analysis never identified method-induced contamination.

The seventh deviation involved assessing the accuracy of each analyzer. Accuracy was to be assessed through FPXRF analysis of 10 to 12 SRM or PE samples. Each analyzer measured a total of 28 SRM or PE samples. Instead, PE samples were used to evaluate the accuracy of the reference methods, and SRMs were used to evaluate the accuracy of the analyzers. This is because the PE concentrations are based on acid extractable concentrations while SRM concentrations represent total metals concentration. SRM data was used for comparative purposes for the reference methods as were PE data for the FPXRF data.

An eighth deviation specific to the SEFA-P Analyzer related to the number of samples to be analyzed. The demonstration plan anticipated that the analyzer would analyze all 630 intrusive samples. The analyzer, in fact, analyzed only 100 samples due to mechanical problems and time constraints at the ASARCO site. A replacement instrument was not immediately available. An EPA-owned substitute instrument was used to analyze the samples for the purpose of this report.

Sample Homogenization

A key quality issue in this demonstration was ensuring that environmental samples analyzed by the reference laboratory and by each of the FPXRF analyzers were splits from a homogenized sample. To address this issue, sample preparation technicians exercised particular care throughout the field work to ensure that samples were thoroughly homogenized before they were split for analysis. Homogenization was conducted by kneading the soil in a plastic bag for a minimum of 2 minutes. If after this time the samples did not appear to be well homogenized, they were kneaded for an additional 2 minutes. This continued until the samples appeared to be well homogenized.

Sodium fluorescein was used as an indicator of sample homogenization. Approximately one-quarter teaspoon of dry sodium fluorescein powder was added to each sample prior to homogenization. After mixing, the sample was examined under an ultraviolet light to assess the distribution of sodium fluorescein throughout the sample. If the fluorescent dye was evenly dispersed, homogenization was considered complete. If the dye was not evenly distributed, the homogenization mixing was continued and repeatedly checked until the dye was evenly distributed throughout the sample.

To evaluate the homogenization process used in this demonstration, 70 field duplicate sample pairs were analyzed by the reference laboratory. Sample homogenization was critical to this demonstration; it assured that the samples measured by the analyzers were as close as possible to samples analyzed by the reference laboratory. This was essential to the primary objectives of this demonstration, the evaluation of comparability between analyzer results and those of the reference methods.

The homogenization process was evaluated by determining the RPD between paired field duplicate samples. The RPDs for the field duplicate samples reflect the total error for the homogenization process and the analytical method combined (Equation 2-4). When total error was determined for the entire data set, the resultant mean RPD total (error) and 95 percent confidence interval was 9.7 ± 1.4, for all metals reported. When only the primary analytes were considered, the RPD total (error) and 95 percent confidence interval was 7.6 ± 1.2.

$$\textit{Total Measurement Error} = \sqrt{[(\textit{Sample Homogenization Error})^2 + (\textit{Laboratory Error})^2]} \qquad \textbf{(2-4)}$$

Using internal QA/QC data from 27 analyses, it was possible to determine the reference laboratory's method error. The reference analytical method precision, as measured by the 95 percent confidence interval around the mean RPDs (laboratory error) of predigestion duplicate analyses, was 9.3 ± 2.9 for the target analytes.

To determine the error introduced by the sample homogenization alone, the error estimate for the reference methods was subtracted from the total error (Equation 2-5). Based on the data presented above, the laboratory-induced error was less than or approximately equal to the total error. This indicates that the sample homogenization (preparation) process contributed little or no error to the overall sample analysis process.

$$\textit{Sample Homogenization Error} = \sqrt{[(\textit{Total Measurement Error})^2 - (\textit{Laboratory Error})^2]} \qquad \textbf{(2-5)}$$

Although the possibility for poorly homogenized samples exists under any homogenization routine, at the scale of analysis used by this demonstration, the samples were considered to be completely homogenized.

Section 3
Reference Laboratory Results

All soil samples collected from the ASARCO and RV Hopkins sites were submitted to the reference laboratory for trace metals analysis. The results are discussed in this section.

Reference Laboratory Methods

Samples collected during this demonstration were homogenized and split for extraction using EPA SW-846 Method 3050A. This is an acid digestion procedure where 1 to 2 grams of soil are digested on a hot plate with nitric acid, followed by hydrogen peroxide, and then refluxed with hydrochloric acid. One gram of soil was used for extraction of the demonstration samples. The final digestion volume was 100 milliliters (mL). The soil sample extracts were analyzed by Method 6010A.

Method 6010A provides analysis of metals using Inductively Coupled Plasma-Atomic Emission Spectroscopy (ICP-AES). This method requires that a plasma be produced by applying a radio-frequency field to a quartz tube wrapped by a coil or solenoid through which argon gas is flowing. The radio-frequency field creates a changing magnetic field in the flowing gas inside the coil, inducing a circulating eddy current on the argon gas that, in turn, heats it. Plasma is initiated by an ignition source and quickly stabilizes with a core temperature of 9,000 - 10,000 degrees Kelvin.

Soil sample extracts are nebulized, and the aerosol is injected into the plasma. Individual analytes introduced into the plasma absorb energy and are excited to higher energy states. These higher energy states have short lifetimes and the individual elements quickly fall back to their ground energy state by releasing a photon. The energy of the emitted photon is defined by the wavelength of electromagnetic radiation produced. Since many electronic transitions are possible for each individual element, several discrete emissions at different wavelengths are observed. Method 6010A provides one recommended wavelength to monitor for each analyte. Due to complex spectra with similar wavelengths from different elements in environmental samples, Method 6010A requires that interference corrections be applied for quantification of individual analytes.

Normal turnaround times for the analysis of soil samples by EPA SW-846 Methods 3050A/6010A range from 21 to 90 days depending on the complexity of the soil samples and the amount of QC documentation required. Faster turnaround times of 1 - 14 days can be obtained, but at additional cost.

Costs for the analysis of soil samples by EPA SW-846 Methods 3050A/6010A range from $150 to $350 per sample depending on turnaround times and the amount of QC documentation required. A sample turnaround of 28 days, a cost of $150 per sample, and a CLP documentation report for QC were chosen for this demonstration.

Reference Laboratory Quality Control

The reference laboratory, Midwest Research Institute (Kansas City, MO), holds certifications for performing target analyte list metals analysis with the U.S. Army Corps of Engineers-Missouri River Division, the State of California, and the State of Utah. These certifications include on-site laboratory audits, data package review audits, and the analysis of PE samples supplied by the certifying agency. PE samples are supplied at least once per year from each of the certifying agencies. The reference laboratory's results for the PE samples are compared to true value results and certifying agency acceptance limits for the PE samples. Continuation of these certifications hinges upon acceptable results for the audits and the PE samples.

The analysis of soil samples by the reference laboratory was governed by the QC criteria in its SOPs, Method 6010A, and the demonstration QAPP. Table 3-1 provides QAPP QC requirements that were monitored and evaluated for the target analytes. Method 6010A QC guidelines also are included in Table 3-1. Due to the complex spectra derived from the analysis of the demonstration samples, the QAPP QC requirements were applied only to the primary analytes. The QAPP QC requirements also were monitored and evaluated for the secondary analytes and other analytes reported by the reference laboratory. However, corrective actions were not required for the secondary analytes.

Table 3-1. Reference Laboratory Quality Control Parameters[a]

Parameter	Frequency	Reference Method Requirement	QAPP Requirement
Initial Calibration Verification (ICV) Standard	With each initial calibration	±10 percent of true value	±10 percent of true value
Continuing Calibration Verification (CCV) Standard	After analysis of every 10 samples and at the end of analytical run	±10 percent of true value	±10 percent of true value
Initial and Continuing Calibration Blanks (ICB) and (CCB)	With each continuing calibration, after analysis of every 10 samples, and at the end of analytical run	±3 standard deviations of the analyzer background mean	No target analytes at concentrations greater than 2 times the lower reporting limit (LRL)
Interference Check Standard (ICS)	With every initial calibration and after analysis of 20 samples	±20 percent of true value	±20 percent of true value
High Level Calibration Check Standard	With every initial calibration	±5 percent of true value	±10 percent of true value
Method Blanks	With each batch of samples of a similar matrix	No QC requirement specified	No target analytes at concentrations greater than 2 times the LRL
Laboratory Control Samples	With each batch of samples of a similar matrix	No QC requirement specified	80 - 120 percent recovery
Predigestion Matrix Spike Samples	With each batch of samples of a similar matrix	80 - 120 percent recovery	80 - 120 percent recovery
Postdigestion Matrix Spike Samples	With each batch of samples of a similar matrix	75 - 125 percent recovery	80 - 120 percent recovery

Table 3-1. ***Continued***

Parameter	Frequency	Reference Method Requirement	QAPP Requirement
Performance Evaluation Samples	As submitted during demonstration	No QC requirement specified	80 - 120 percent recovery within performance acceptance limits (PAL)
Predigestion Laboratory Duplicate Samples	With each batch of samples of a similar matrix	20 percent relative percent difference (RPD)[b]	20 percent RPD[c]
Postdigestion Laboratory Duplicate Samples	With each batch of samples of a similar matrix	No QC requirement specified	10 percent RPD[c]

Notes:
[a] Quality control parameters were evaluated on the raw reference data.
[b] RPD control limits only pertain to original and laboratory duplicate sample results that were greater than 10 times the instrument detection limit (IDL).
[c] RPD control limits only pertain to original and laboratory duplicate sample results that were greater than or equal to 10 times the LRL.

PRC performed three on-site audits of the reference laboratory during the analysis of pre-demonstration and demonstration samples. These audits were conducted to observe and evaluate the procedures used by the reference laboratory and to ensure that these procedures adhered to the QAPP QC requirements. Audit findings revealed that the reference laboratory followed the QAPP QC requirements. It was determined that the reference laboratory had problems meeting two of the QAPP QC requirements: method blank results and the high level calibration check standard's percent recovery. Due to these problems, these two QAPP QC requirements were widened. The QC requirement for method blank sample results was changed from no target analytes at concentrations greater than the lower reporting limit (LRL) to two times the LRL. The QC requirement for the high level calibration standard percent recovery was changed from ±5 to ±10 percent of the true value. These changes were approved by the EPA and did not affect the results of the demonstration.

The reference laboratory internally reviewed its data before releasing it. PRC conducted a QC review on the data based on the QAPP QC requirements and corrective actions listed in the demonstration plan.

Quality Control Review of Reference Laboratory Data

The QC data review focused upon the compliance of the data with the QC requirements specified in the demonstration QAPP. The following sections discuss results from the QC review of the reference laboratory data. All QC data evaluations were based on raw data.

Reference Laboratory Sample Receipt, Handling, and Storage Procedures

Demonstration samples were divided into batches of no more than 20 samples per batch prior to delivery to the reference laboratory. A total of 23 batches containing 315 samples and 70 field duplicate samples was submitted to the reference laboratory. The samples were shipped in sealed coolers at ambient temperature under a chain of custody.

Upon receipt of the demonstration samples, the reference laboratory assigned each sample a unique number and logged each into its laboratory tracking system. The samples were then transferred to the reference laboratory's sample storage refrigerators to await sample extraction.

Samples were transferred to the extraction section of laboratory under an internal chain of custody. Upon completion of extraction, the remaining samples were returned to the sample storage refrigerators. Soil sample extracts were refrigerated in the extraction laboratory while awaiting sample analysis.

Sample Holding Times

The maximum allowable holding time from the date of sample collection to the date of extraction and analysis using EPA SW-846 Methods 3050A/6010A is 180 days. Maximum holding times were not exceeded for any samples during this demonstration.

Initial and Continuing Calibrations

Prior to sample analysis, initial calibrations (ICAL) were performed. ICALs for Method 6010A consist of the analysis of three concentrations of each target analyte and a calibration blank. The low concentration standard is the concentration used to verify the LRL of the method. The remaining standards are used to define the linear range of the ICP-AES. The ICAL is used to establish calibration curves for each target analyte. Method 6010A requires an initial calibration verification (ICV) standard to be analyzed with each ICAL. The method control limit for the ICV is ±10 percent. An interference check sample (ICS) and a high level calibration check standard is required to be analyzed with every ICAL to assess the accuracy of the ICAL. The control limits for the ICS and high level calibration check standard were ±20 percent recovery and ±10 percent of the true value, respectively. All ICALs, ICVs, and ICSs met the respective QC requirements for all target analytes.

Continuing calibration verification (CCV) standards and continuing calibration blanks (CCB) were analyzed following the analysis of every 10 samples and at the end of an analytical run. Analysis of the ICS was also required after every group of 20 sample analyses. These QC samples were analyzed to check the validity of the ICAL. The control limits for the CCVs were ±10 percent of the true value. The control limits for CCBs were no target analyte detected at concentrations greater than 2 times the LRL. All CCVs, CCBs, and ICSs met the QAPP requirements for the target analytes with the exception of one CCV where the barium recovery was outside the control limit. Since barium was a primary analyte, the sample batch associated with this CCV was reanalyzed and the resultant barium recovery met the QC criteria.

Detection Limits

The reference laboratory LRLs for the target analytes are listed in Table 3-2. These LRLs were generated through the use of an MDL study of a clean soil matrix. This clean soil matrix was also used for method blank samples and LCSs during the analysis of demonstration samples. The MDL study involved seven analyses of the clean soil matrix spiked with low concentrations of the target analytes. The mean and standard deviation of the response for each target analyte was calculated. The LRL was defined as the mean plus three times the standard deviation of the response for each target analyte included in the method detection limit study. All LRLs listed in Table 3-2 were met and maintained throughout the analysis of the demonstration samples.

The reference laboratory reported soil sample results in units of milligram per kilogram wet weight. All reference laboratory results referred to in this report are wet-weight sample results.

Table 3-2. SW-846 Method 6010A LRLs for Target Analytes

Analyte	LRL (mg/kg)	Analyte	LRL (mg/kg)
Antimony	6.4	Copper*	1.2
Arsenic*	10.6	Iron	600[a]
Barium*	5.0	Lead*	8.4
Cadmium	0.80	Nickel	3.0
Chromium*	2.0	Zinc*	2.0

Notes:
[a] LRL elevated due to background interference.
* Primary analyte.
mg/kg Milligrams per kilogram.

Method Blank Samples

Method blanks were prepared using a clean soil matrix and acid digestion reagents used in the extraction procedure. A minimum of one method blank sample was analyzed for each of the 23 batches of demonstration samples submitted for reference laboratory analysis. All method blanks provided results for target analytes at concentrations less than 2 times the levels shown in Table 3-2.

Laboratory Control Samples

All LCSs met the QAPP QC requirements for all primary and secondary analytes except those discussed below.

The primary analytes copper and lead were observed outside the QC limits in one of the 23 batches of samples analyzed. Reanalysis of the affected batches was not performed by the reference laboratory. These data were qualified by the reference laboratory. Copper and lead data for all samples included in the affected batches were rejected and not used for demonstration statistical comparisons.

Concentrations of secondary analytes antimony, nickel, and cadmium were observed outside the QC limits in the LCSs. Antimony LCS recoveries were continually outside the control limits, while nickel and cadmium LCS recoveries were only occasionally outside QC limits. Antimony was a problem analyte and appeared to be affected by acid digestion, which can cause recoveries to fall outside control limits. Antimony recoveries ranged from 70 to 80 percent. Since secondary analytes were not subject to the corrective actions listed in the demonstration QAPP, no reanalysis was performed based on the LCS results of the secondary target analytes. These values were qualified by the reference laboratory. All other secondary analyte LCS recoveries fell within the QAPP control limits.

Predigestion Matrix Spike Samples

One predigestion matrix spike sample and duplicate were prepared by the reference laboratory for each batch of demonstration samples submitted for analysis. The predigestion matrix spike duplicate sample was not required by the QAPP, but it is a routine sample prepared by the reference laboratory. This duplicate sample can provide data that indicates if out-of-control recoveries are due to matrix interferences or laboratory errors.

Predigestion spike recovery results for the primary analytes arsenic, barium, chromium, copper, lead, and zinc were outside control limits for at least 1 of the 23 sample batches analyzed by the reference method. These control limit problems were due to either matrix effects or initial spiking concentrations below native analyte concentrations.

Barium, copper, and lead predigestion matrix spike recovery results were outside control limits in sample batches 2, 3, and 5. In all of these cases, the unacceptable recoveries were caused by spiking concentrations that were much lower than native concentrations of the analytes. These samples were re-prepared, spiked with higher concentrations of analytes, reextracted, and reanalyzed. Following this procedure, the spike recoveries fell within control limits upon reanalysis.

One predigestion matrix spike recovery was outside control limits for arsenic. The predigestion matrix spike duplicate sample also was outside of control limits. This sample exhibited an acceptable RPD for the recovery of arsenic in the predigestion matrix spike and duplicate. A matrix interference may have been responsible for the low recovery. This sample was not reanalyzed.

Chromium predigestion matrix spike recoveries were outside control limits in 7 of the 23 batches of samples analyzed. Five of these seven failures exhibited recoveries ranging from 67 to 78 percent, close to the low end of the control limits. These recoveries were similar in the predigestion matrix spike duplicate samples prepared and analyzed in the same batch. This indicates that these five failures were due to matrix interferences. The predigestion matrix spike duplicate samples prepared and analyzed along with the remaining two failures did not agree with the recoveries of the postdigestion matrix spike samples, indicating that these two failures may be due to laboratory error, possibly inaccuracies in sample spiking. These seven predigestion matrix spike samples were not reanalyzed.

The zinc predigestion matrix spike recovery data were outside control limits for four batches of samples analyzed. In three of the spike recovery pairs, recoveries ranged from 70 to 76 percent, close to the lower end of the control limits. The fourth recovery was much less than the lower end of the control limits. All of the predigestion matrix spike duplicate samples provided recoveries that agreed with the recoveries for the predigestion matrix spike sample recoveries indicating that the low recoveries were due to matrix effects. These predigestion matrix spikes and associated samples were not reanalyzed.

The secondary analytes, cadmium, iron, and nickel, had predigestion spike recoveries outside control limits. Cadmium spike recoveries were outside control limits six times. These recoveries ranged from 71 to 85 percent. Iron spike recoveries were outside of control limits once. Nickel spike recoveries were outside control limits four times. These recoveries ranged from 74 to 83 percent. Antimony spike recoveries were always outside control limits. No corrective action was taken for these secondary target analytes.

Demonstration sample results for all target analytes that did not meet the control limits for predigestion matrix spike recovery were qualified by the reference laboratory.

Postdigestion Matrix Spike Samples

All postdigestion matrix spike results were within the control limit of 80 - 120 percent recovery for the primary analytes.

Secondary analytes, antimony, and iron were observed outside the control limits. However, no corrective action was taken for secondary analytes as stated in the demonstration QAPP. All postdigestion

spike recoveries for target analytes met the QA/QC requirements of the QAPP and were considered acceptable.

Predigestion Laboratory Duplicate Samples

Predigestion laboratory duplicate RPD results were within the control limit of 20 percent for analyte concentrations greater than 10 times the LRL except for the following instances. RPDs for primary analytes barium, arsenic, lead, chromium, and copper were observed above the control limit in five predigestion laboratory duplicate samples. These samples were reanalyzed according to the corrective actions listed in the QAPP. The reanalysis produced acceptable RPD results for these primary analytes.

RPD results for the secondary analytes antimony, nickel, and cadmium were observed outside the control limit for a number of sample batches. No corrective action was taken for secondary analytes that exceeded the RPD control limit.

Postdigestion Laboratory Duplicate Samples

All primary analyte postdigestion laboratory duplicate RPD results were less than the 10 percent control limit for analyte concentrations greater than 10 times the LRL.

The RPDs for secondary analytes antimony and iron were observed above the 10 percent control limit in two sample batches. No corrective action was taken for secondary target analytes that exceeded the RPD control limit.

Performance Evaluation Samples

PE samples were purchased from Environmental Resource Associates (ERA). The PE samples are Priority PollutnT™/Contract Laboratory Program (CLP) QC standards for inorganics in soil. This type of sample is used by the EPA to verify accuracy and laboratory performance. Trace metal values are certified by interlaboratory round robin analyses. ERA lists performance acceptance limits (PAL) for each analyte that represent a 95 percent confidence interval (CI) around the certified value. PALs are generated by peer laboratories in ERA's InterLaB™ program using the same samples that the reference laboratory analyzed and the same analytical methods. The reported value for each analyte in the PE sample must fall within the PAL range for the accuracy to be acceptable. Four PE samples were submitted "double blind" (the reference laboratory was not notified that the samples were QC samples or of the certified values for each element) to the reference laboratory for analysis by EPA SW-846 Methods 3050A/6010A. Reference laboratory results for all target analytes are discussed later in this section.

Four certified reference materials (CRM) purchased from Resource Technology Corporation (RTC) also were used as PE samples to verify the accuracy and performance of the reference laboratory. These four CRMs were actual samples from contaminated sites. They consisted of two soils, one sludge, and one ash CRM. Metal values in the CRMs are certified by round robin analyses of at least 20 laboratories according to the requirements specified by the EPA Cooperative Research and Development Agreement. The certified reference values were determined by EPA SW-846 Methods 3050A/6010A. RTC provides a 95 percent PAL around each reference value in which measurements should fall 19 of 20 times. The reported value from the reference laboratory for each analyte must fall within this PAL for the accuracy to be considered acceptable. As with the four PE samples, the four CRMs were submitted "double blind" to the reference laboratory for analysis by EPA SW-846 Methods 3050A/6010A. The reference laboratory results for the target analytes are discussed later in the Accuracy subsection.

Standard Reference Material Samples

As stated in the demonstration plan (PRC 1995), PE samples also consisted of SRMs. The SRMs consisted of solid matrices such as soil, ash, and sludge. Certified analyte concentrations for SRMs are determined on an analyte by analyte basis by multiple analytical methods including but not limited to ICP-AES, flame atomic absorption spectroscopy, ICP-mass spectrometry, XRF, instrumental neutron activation analysis, hydride generation atomic absorption spectroscopy, and polarography. These certified values represent total analyte concentrations and complete extraction. This is different from the PE samples, CRM samples, and the reference methods, which use acid extraction that allows quantitation of only acid extractable analyte concentrations.

The reference laboratory analyzed 14 SRMs supplied by the National Institute of Standards and Technology (NIST), U.S. Geological Survey (USGS), National Research Council Canada, South African Bureau of Standards, and Commission of the European Communities. The percentage of analyses of SRMs that were within the QAPP-defined control limits of 80 - 120 percent recovery was calculated for each primary and secondary analyte.

Analyses of SRMs were not intended to assess the accuracy of EPA SW-846 Methods 3050A/6010A as were the ERA PE or RTC CRM samples. Comparison of EPA SW-846 Methods 3050A/6010A acid leach data to SRM data cannot be used to establish method validity (Kane and others 1993). This is because SRM values are acquired by analyzing the samples by methods other than the ICP-AES method. In addition, these other methods use sample preparation techniques different from those for EPA SW-846 Methods 3050A/6010A. This is one reason no PALs are published with the SRM certified values. Therefore, the SRMs were not considered an absolute test of the reference laboratory's accuracy for EPA SW-846 Methods 3050A/6010A.

The SRM sample results were not used to assess method accuracy or to validate the reference methods. This was due to the fact that the reported analyte concentrations for SRMs represent total analyte concentrations. The reference methods are not an analysis of total metals; rather they target the leachable concentrations of metals. This is consistent with the NIST guidance against using SRMs to assess performance on leaching based analytical methods (Kane and others 1993).

Data Review, Validation, and Reporting

Demonstration data were internally reviewed and validated by the reference laboratory. Validation involved the identification and qualification of data affected by QC procedures or samples that did not meet the QC requirements of the QAPP. Validated sample results were reported using both hard copy and electronic disk deliverable formats. QC summary reports were supplied with the hard copy results. This qualified data was identified and discussed in the QC summary reports provided by the reference laboratory.

Demonstration data reported by the reference laboratory contained three types of data qualifiers: C, Q, and M. Type C qualifiers included the following:

- U - the analyte was analyzed for but not detected.
- B - the reported value was obtained from a reading that was less than the LRL but greater than or equal to the IDL.

Type Q qualifiers included the following:

- N - spiked sample recovery was not within control limits.
- * - duplicate analysis was not within control limits.

Type M qualifiers include the following:

- P - analysis performed by ICP-AES (Method 6010).

Quality Assessment of Reference Laboratory Data

An assessment of the reference laboratory data was performed using the PARCC parameters discussed in Section 2. PARCC parameters are used as indicators of data quality and were evaluated using the review of reference laboratory data discussed above. The following sections discuss the data quality for each PARCC parameter. This quality assessment was based on raw reference data and the raw PE sample data.

The quality assessment was limited to an evaluation of the primary analytes. Secondary and other analytes reported by the reference laboratory were not required to meet the QC requirements specified in the QAPP. Discussion of the secondary analytes is presented in the precision, accuracy, and comparability sections for informational purposes only.

Precision

Precision for the reference laboratory data was assessed through an evaluation of the RPD produced from the analysis of predigestion laboratory duplicate samples and postdigestion laboratory duplicate samples. Predigestion laboratory duplicate samples provide an indication of the method precision, while postdigestion laboratory duplicate samples provide an indication of instrument performance. Figure 3-1 provides a graphical summary of the reference method precision data.

The predigestion duplicate RPDs for the primary and secondary analytes fell within the 20 percent control limit, specified in the QAPP, for 17 out of 23 batches of demonstration samples. The six results that exceeded the control limit involved only 11 of the 230 samples evaluated for predigestion duplicate precision (Figure 3-1). This equates to 95 percent of the predigestion duplicate data meeting the QAPP control limits. Six of the analytes exceeding control limits had RPDs less than 30 percent. Three of the analytes exceeding control limits had RPDs between 30 and 40 percent. Two of the analytes exceeding control limits had RPDs greater than 60 percent. These data points are not shown in Figure 3-1. Those instances where the control limits were exceeded are possibly due to nonhomogeneity of the sample or simply to chance, as would be expected with a normal distribution of precision analyses.

The postdigestion duplicate RPDs for the primary and secondary analytes fell within the 10 percent control limit, specified in the QAPP, for 21 out of 23 batches of demonstration samples. The two results that exceeded the control limit involved only 3 of the 230 samples evaluated for postdigestion duplicate precision in the 23 sample batches (Figure 3-1). This equates to 99 percent of the postdigestion duplicate data meeting the QAPP control limits. The RPDs for the three results that exceeded the control limit ranged from 11 to 14 percent.

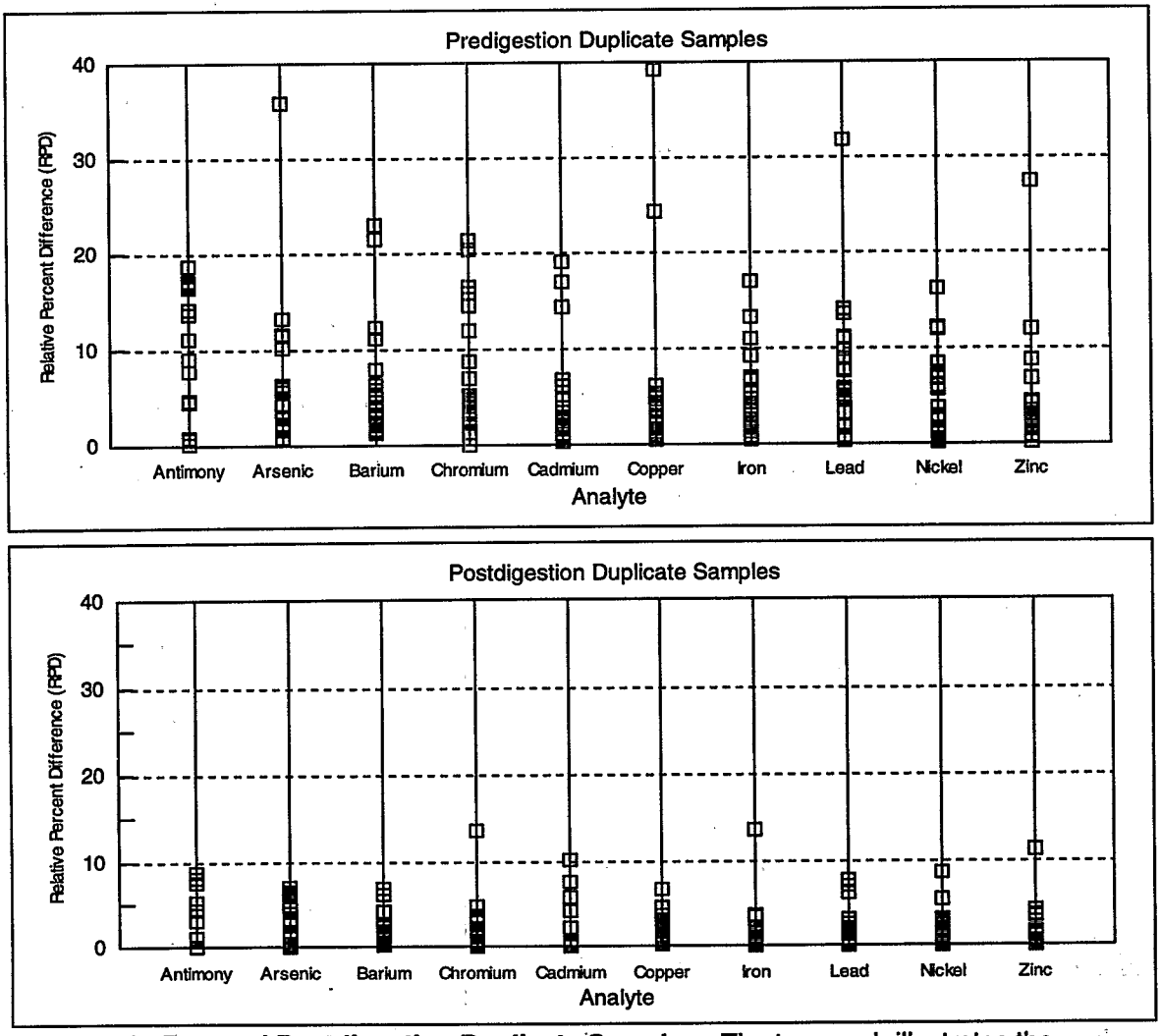


Figure 3-1. Pre- and Postdigestion Duplicate Samples: The top graph illustrates the reference laboratory's performance on analyzing predigestion duplicate samples. Twenty percent RPD represents the predigestion duplicate control limits defined in the demonstration QAPP. Two points were deleted from this top graph: barium at 65 percent RPD and copper at 138 percent RPD. The bottom graph illustrates the reference laboratory's performance on analyzing postdigestion duplicate samples. Ten percent RPD represents the postdigestion duplicate control limits defined in the demonstration QAPP.

Accuracy

Accuracy for the reference laboratory data was assessed through evaluations of the PE samples (including the CRMs), LCSs, method blank sample results, and pre- and postdigestion matrix spike samples. PE samples were used to assess the absolute accuracy of the reference laboratory method as a whole, while LCSs, method blanks, and pre- and postdigestion matrix spike samples were used to assess the accuracy of each batch of demonstration samples.

A total of eight PE and CRM samples was analyzed by the reference laboratory. These included four ERA PE samples and four RTC CRM samples. One of the ERA PE samples was submitted to the reference laboratory in duplicate, thereby producing nine results to validate accuracy. The accuracy data for all primary and secondary analytes are presented in Table 3-3 and displayed in Figure 3-2. Accuracy was assessed over a wide-concentration range for all 10 analytes with concentrations for most analytes spanning one or more orders of magnitude.

Reference laboratory results for all target analytes in the ERA PE samples fell within the PALs. In the case of the RTC CRM PE samples, reference laboratory results for copper in one CRM and zinc in two CRMs fell outside the published acceptance limits. One of the two out-of-range zinc results was only slightly above the upper acceptance limit (811 versus 774 mg/kg). The other out-of-range zinc result and the out-of-range copper result were about three times higher than the certified value and occurred in the same CRM. These two high results skewed the mean percent recovery for copper and zinc shown in Table 3-3. Figure 3-2 shows that the remaining percent recoveries for copper and zinc were all near 100 percent.

Table 3-3 shows that a total of 83 results was obtained for the 10 target analytes. Eighty of the 83 results or 96.4 percent fell within the PALs. Only 3 out of 83 times did the reference method results fall outside PALs. This occurred once for copper and twice for zinc. Based on this high percentage of acceptable results for the ERA and CRM PE samples, the accuracy of the reference methods was considered acceptable.

Table 3-3. Reference Laboratory Accuracy Data for Target Analytes

Analyte	n	Percent Within Acceptance Range	Mean Percent Recovery	Range of Percent Recovery	SD of Percent Recovery	Concentration Range (mg/kg)
Antimony	6	100	104	83 - 125	15	50 - 4,955
Arsenic	8	100	106	90 - 160	22	25 - 397
Barium	9	100	105	83 - 139	21	19 - 586
Cadmium	9	100	84	63 - 93	10	1.2 - 432
Chromium	9	100	91	77 - 101	8	11 - 187
Copper	9	89	123	90 - 332	79	144 - 4,792
Iron	7	100	98	79 - 113	12	6,481 - 28,664
Lead	8	87.5	86	35 - 108	22	52 - 5,194
Nickel	9	100	95	79 - 107	10	13 - 13,279
Zinc	9	78	120	79 - 309	72	76 - 3,021

Notes: n Number of samples with detectable analyte concentrations.
SD Standard deviation.
mg/kg Milligrams per kilogram.

LCS percent recoveries for all the primary analytes were acceptable in 21 of the 23 sample batches. Lead recovery was unacceptable in one sample batch and lead results for each sample in that batch were rejected.

Copper recovery was unacceptable in another sample batch, and copper results for each sample in this batch also were rejected. Percent recoveries of the remaining primary analytes in each of these two batches were acceptable. In all, 136 of 138 LCS results or 98.5 percent fell within the control limits.

Method blank samples for all 23 batches of demonstration samples provided results of less than 2 times the LRL for all primary analytes. This method blank control limit was a deviation from the QAPP, which had originally set the control limit at no target analytes at concentrations greater than the LRL. This control limit was widened at the request of the reference laboratory. A number of batches were providing method blank results for target analytes at concentrations greater than the LRL, but less than 2 times the LRL. This alteration was allowed because even at 2 times the LRL, positive results for the method blank samples were still significantly lower than the MDLs for each of the FPXRF analyzers. The results from the method blank samples did not affect the accuracy of the reference data as it was to be used in the demonstration statistical evaluation of FPXRF analyzers.

The percent recovery for the predigestion matrix spike samples fell outside of the 80 - 120 percent control limit specified in the QAPP in several of the 23 batches of demonstration samples. The predigestion matrix spike sample results indicate that the accuracy of specific target analytes in samples from the affected batches may be suspect. These results were qualified by the reference laboratory. These data were not excluded from use for the demonstration statistical comparison. A discussion of the use of this qualified data is included in the "Use of Qualified Data for Statistical Analysis" subsection.

The RPD for the postdigestion matrix spike samples fell within the 80 - 120 percent control limit specified in the QAPP for all 23 batches of demonstration samples.

The QA review of the reference laboratory data indicated that the absolute accuracy of the method was acceptable. Based on professional judgement, it was determined that the small percentage of outliers did not justify rejection of any demonstration sample results from the reference laboratory. The accuracy assessment also indicated that most of the batch summary data were acceptable. Two batches were affected by LCS outliers, and some data were qualified due to predigestion matrix spike recovery outliers. This data was rejected or qualified. Rejected data was not used. Qualified data were used as discussed below.

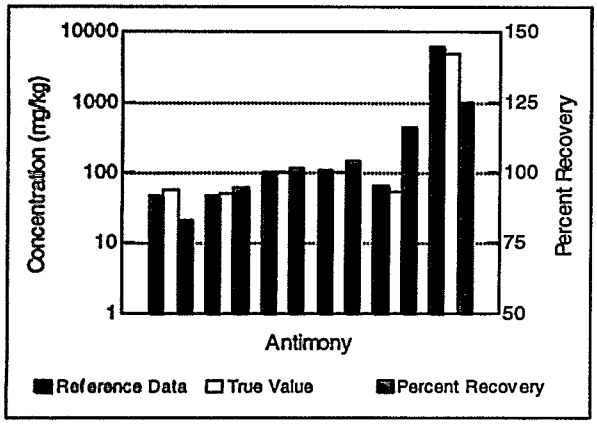
Representativeness

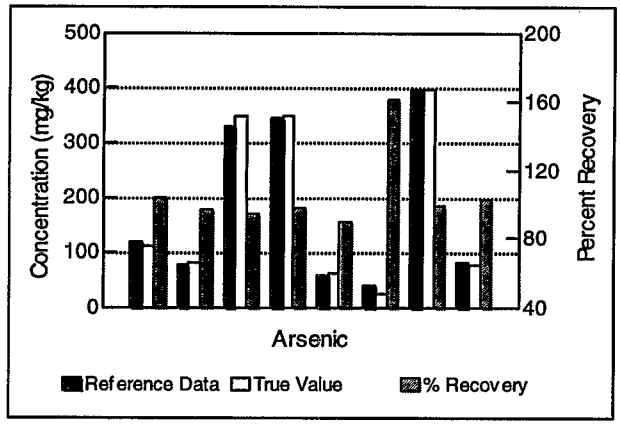
Representativeness of the analytical data was evaluated through laboratory audits performed during the course of sample analysis and by QC sample analyses, including method blank samples, laboratory duplicate samples, and CRM and PE samples. These QC samples were determined to provide acceptable results. From these evaluations, it was determined that representativeness of the reference data was acceptable.

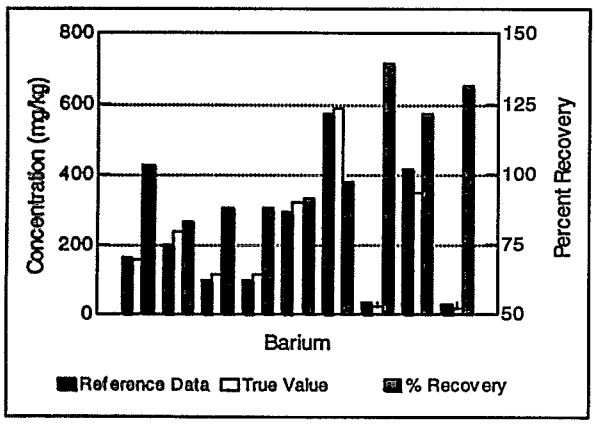
Completeness

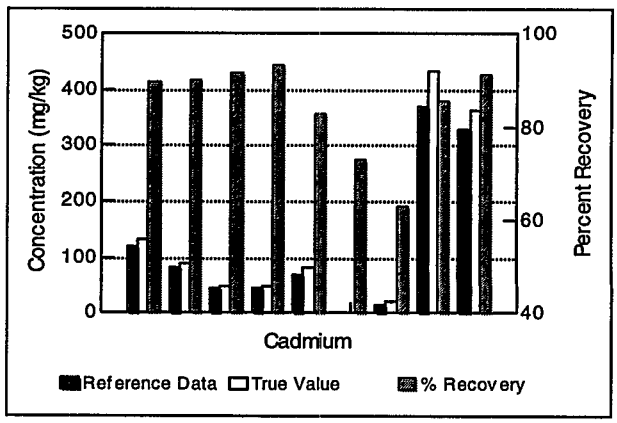
Results were obtained for all soil samples extracted and analyzed by EPA SW-846 Methods 3050A/6010A. Some results were rejected or qualified. Rejected results were deemed incomplete. Qualified results were usable for certain purposes and were deemed as complete.

To calculate completeness, the number of nonrejected results was determined. This number was divided by the total number of results expected, and then multiplied by 100 to express completeness as a percentage. A total of 385 samples was submitted for analysis. Six primary analytes were reported, resulting in an expected 2,310 results. Forty of these were rejected, resulting in 2,270 complete results. Reference laboratory completeness was determined to be 98.3 percent, which exceeded the objective for this demonstration of 95 percent. The reference laboratory's completeness was, therefore, considered acceptable.









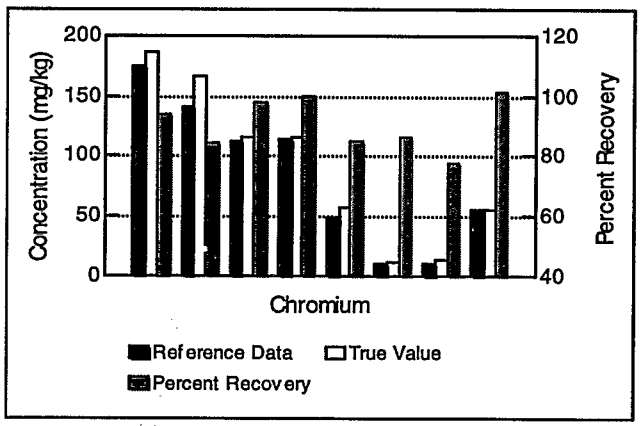
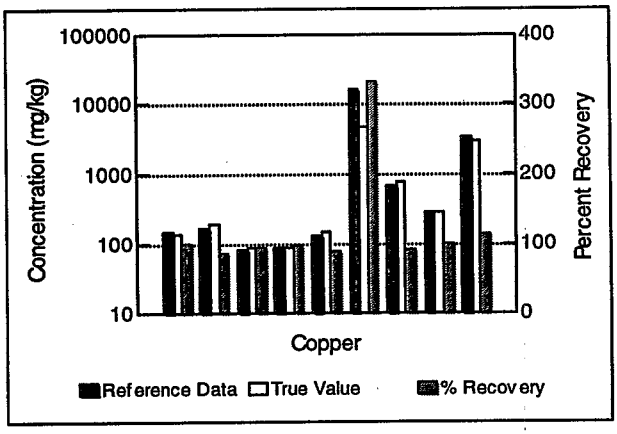
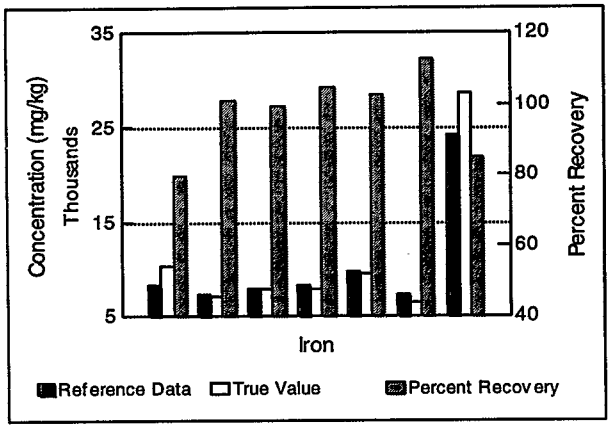
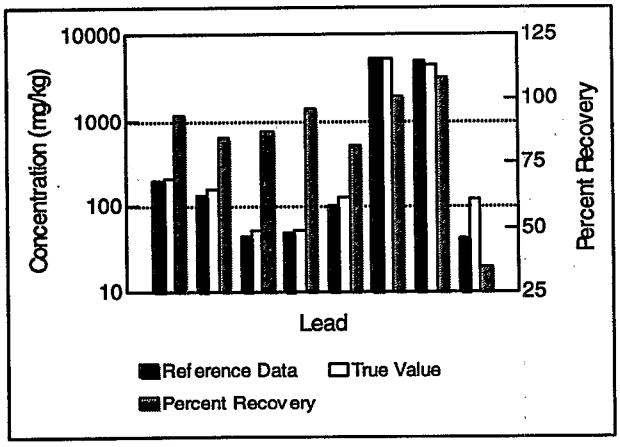
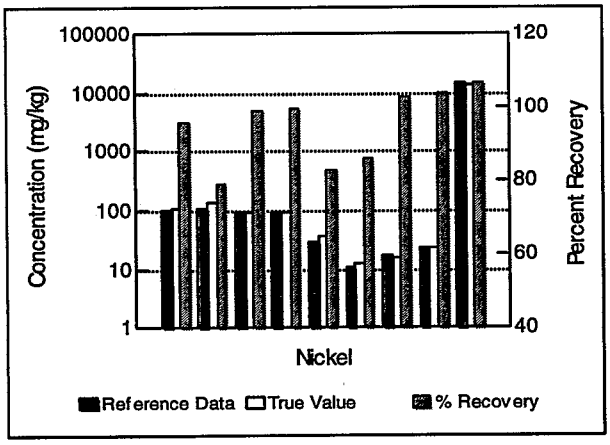


Figure 3-2. Reference Method PE and CRM Results: These graphs illustrate the relationship between the reference data and the true values for the PE or CRM samples. The gray bars represent the percent recovery for the reference data. Each set of three bars (black, white, and gray) represents a single PE or CRM sample. Based on this high percentage of acceptable results for the ERA and CRM PE samples, the accuracy of the reference laboratory method was considered acceptable.









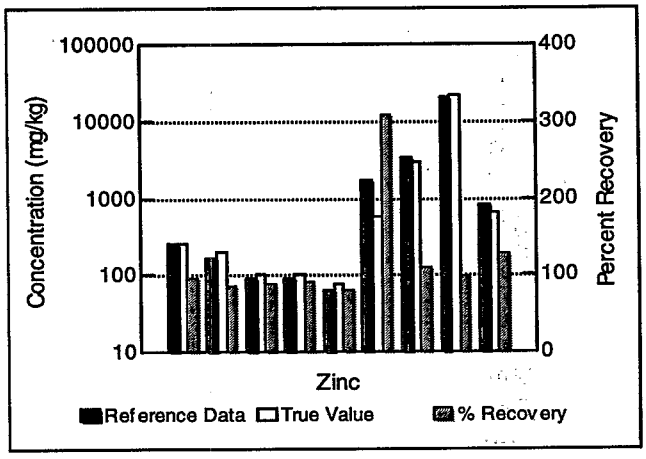


Figure 3-2 (Continued). Reference Method PE and CRM Results: These graphs illustrate the relationship between the reference data and the true values for the PE or CRM samples. The gray bars represent the percent recovery for the reference data. Each set of three bars (black, white, and gray) represents a single PE or CRM sample. Based on this high percentage of acceptable results for the ERA and CRM PE samples, the accuracy of the reference laboratory method was considered acceptable.

Comparability

Comparability of the reference data was controlled by following laboratory SOPs written for the performance of sample analysis using EPA SW-846 Methods 3050A/6010A. QC criteria defined in the SW-846 methods and the demonstration plan (PRC 1995) were followed to ensure that reference data would provide comparable results to any laboratory reporting results for the same samples.

Reference results indicated that EPA SW-846 Methods 3050A/6010A did not provide comparable results for some analytes in the SRM samples. SRM performance data for target analytes is summarized in Table 3-4 and displayed in Figure 3-3. As with the PEs, the analyte concentrations spanned up to 3 orders of magnitude in the SRMs. The percentage of acceptable (80 - 120 percent recovery) SRM results and mean percent recovery was less than 50 percent for the analytes antimony, barium, chromium, iron, and nickel. The low recoveries for these five analytes reflect the lesser tendency for them to be acid-extracted (Kane and others 1993).

Under contract to the EPA, multiple laboratories analyzed NIST SRMs 2709, 2710, and 2711 by EPA SW-846 Methods 3050A/6010A. A range, median value, and percent leach recovery based on the median value for each detectable element were then published as an addendum to the SRM certificates. These median values are not certified but provide a baseline for comparison to other laboratories analyzing these SRMs by EPA SW-846 Methods 3050A/6010A. Table 3-5 presents the published percent leach recovery for the 10 primary and secondary analytes and the reference laboratory's results for these three NIST SRMs. Table 3-5 shows that the results produced by the reference laboratory were consistent with the published results indicating good comparability to other laboratories using the same analytical methods on the same samples.

Table 3-4. SRM Performance Data for Target Analytes

Analyte	n	Percent Within Acceptance Range	Mean Percent Recovery	Range of Percent Recovery	SD of Percent Recovery	Concentration Range (mg/kg)
Antimony	5	0	22	15 - 37	9	3.8 - 171
Arsenic	11	72	84	67 - 106	10	18 - 626
Barium	8	12	41	21 - 89	21	414 - 1,300
Cadmium	10	50	80	43 - 95	15	2.4 - 72
Chromium	10	0	45	14 - 67	16	36 - 509
Copper	17	88	82	33 - 94	17	35 - 2,950
Iron	7	14	62	23 - 84	25	28,900 - 94,000
Lead	17	82	83	37 - 99	17	19 - 5,532
Nickel	16	19	67	25 - 91	17	14 - 299
Zinc	16	75	81	32 - 93	14	81 - 6,952

Notes:
n Number of SRM samples with detectable analyte concentrations.
SD Standard deviation.
mg/kg Milligrams per kilogram.

Table 3-5. Leach Percent Recoveries for Select NIST SRMs

Analyte	NIST SRM 2709		NIST SRM 2710		NIST SRM 2711	
	Published Result[a]	Reference Laboratory Result	Published Result[a]	Reference Laboratory Result	Published Result[a]	Reference Laboratory Result
Antimony	–	–	21	–	–	20
Arsenic	–	106	94	87	86	91
Barium	41	37	51	45	28	25
Cadmium	–	–	92	84	96	87
Chromium	61	–	49	–	43	49
Copper	92	85	92	92	88	90
Iron	86	84	80	78	76	66
Lead	69	87	92	96	95	90
Nickel	89	76	71	69	78	70
Zinc	94	78	85	88	89	85

Notes:
[a] Published results found in an addendum to SRM certificates for NIST SRMs 2709, 2710, and 2711.
NIST National Institute of Standards and Technology.
SRM Standard reference materials.
– Analyte not present above the method LRL.

The inability of EPA SW-846 Methods 3050A/6010A to achieve the predetermined 80 - 120 percent recovery requirement indicated that the methods used to determine the certified values for the SRM samples were not comparable to EPA SW-846 Methods 3050A/6010A. Differences in the sample extraction methods and the use of different analytical instruments and techniques for each method were the major factors of this noncomparability. Because of these differences, it was not surprising that the mean percent recovery was less than 100 percent for the target analytes. The lack of comparability of EPA SW-846 Methods 3050A/6010A to the total metals content in the SRMs did not affect the quality of the data generated by the reference laboratory.

The assessment of comparability for the reference data revealed that it should be comparable to other laboratories performing analysis of the same samples using the same extraction and analytical methods, but it may not be comparable to laboratories performing analysis of the same samples using different extraction and analytical methods or by methods producing total analyte concentration data.

Use of Qualified Data for Statistical Analysis

As noted above, the reference laboratory results were reported and validated, qualified, or rejected by approved QC procedures. Data were qualified for predigestion matrix spike recovery and pre- and postdigestion laboratory duplicate RPD control limit outliers. None of the problems were considered sufficiently serious to preclude the use of coded data. Appropriate corrective action identified in the demonstration plan (PRC 1995) was instituted. The result of the corrective action indicated that the poor percent recovery and RPD results were due to matrix effects. Since eliminating the matrix effects would require additional analysis using a different determination method such as atomic absorption spectrometry, or the method of standard addition, the matrix effects were noted and were not corrected.

PARCC parameters for the reference laboratory data were determined to be acceptable. It was expected that any laboratory performing analysis of these samples using EPA SW-846 Methods 3050A/6010A would experience comparable matrix effects. A primary objective of this demonstration was to compare sample results from the FPXRF analyzers to EPA SW-846 Methods 3050A/6010A, the most widely used approved methods for determining metal concentrations in soil samples. The comparison of FPXRF and the reference methods had to take into account certain limitations of both methods, including matrix effects. For these reasons, qualified reference data were used for statistical analysis.

The QC review and QA audit of the reference data indicated more than 98 percent of the data either met the demonstration QAPP objectives or was QC coded for reasons not limiting its use in the data evaluation. Less than 2 percent of the data were rejected based on QAPP criteria. Rejected data were not used for statistical analysis. The reference data were considered as good as or better than other laboratory analyses of samples performed using the same extraction and analytical methods. The reference data met the definitive data quality criteria and was of sufficient quality to support regulatory activities. The reference data were found to be acceptable for comparative purposes with the FPXRF data.

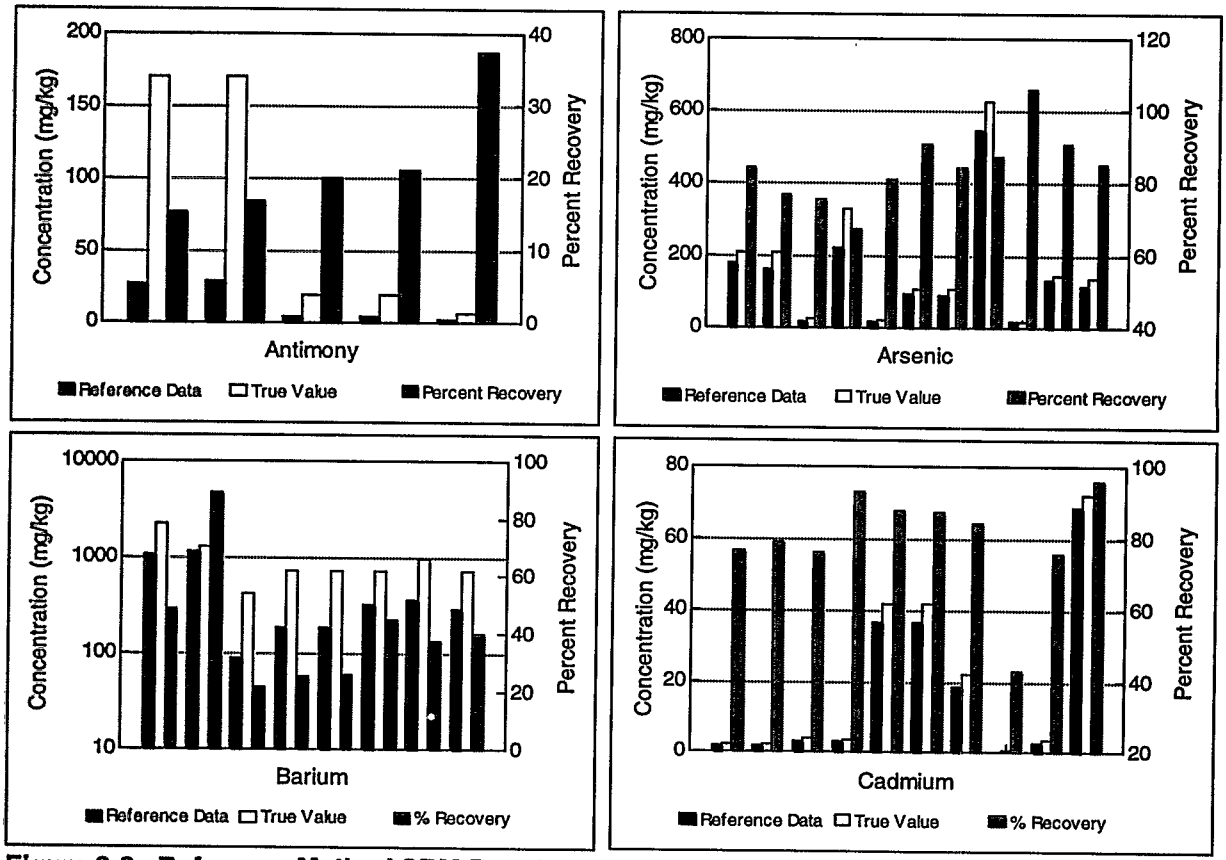


Figure 3-3. Reference Method SRM Results: These graphs illustrate the relationship between the reference data and the true values for the SRM samples. The gray bars represent the percent recovery for the reference data. Each set of three bars (black, white, and gray) represents a single SRM sample.

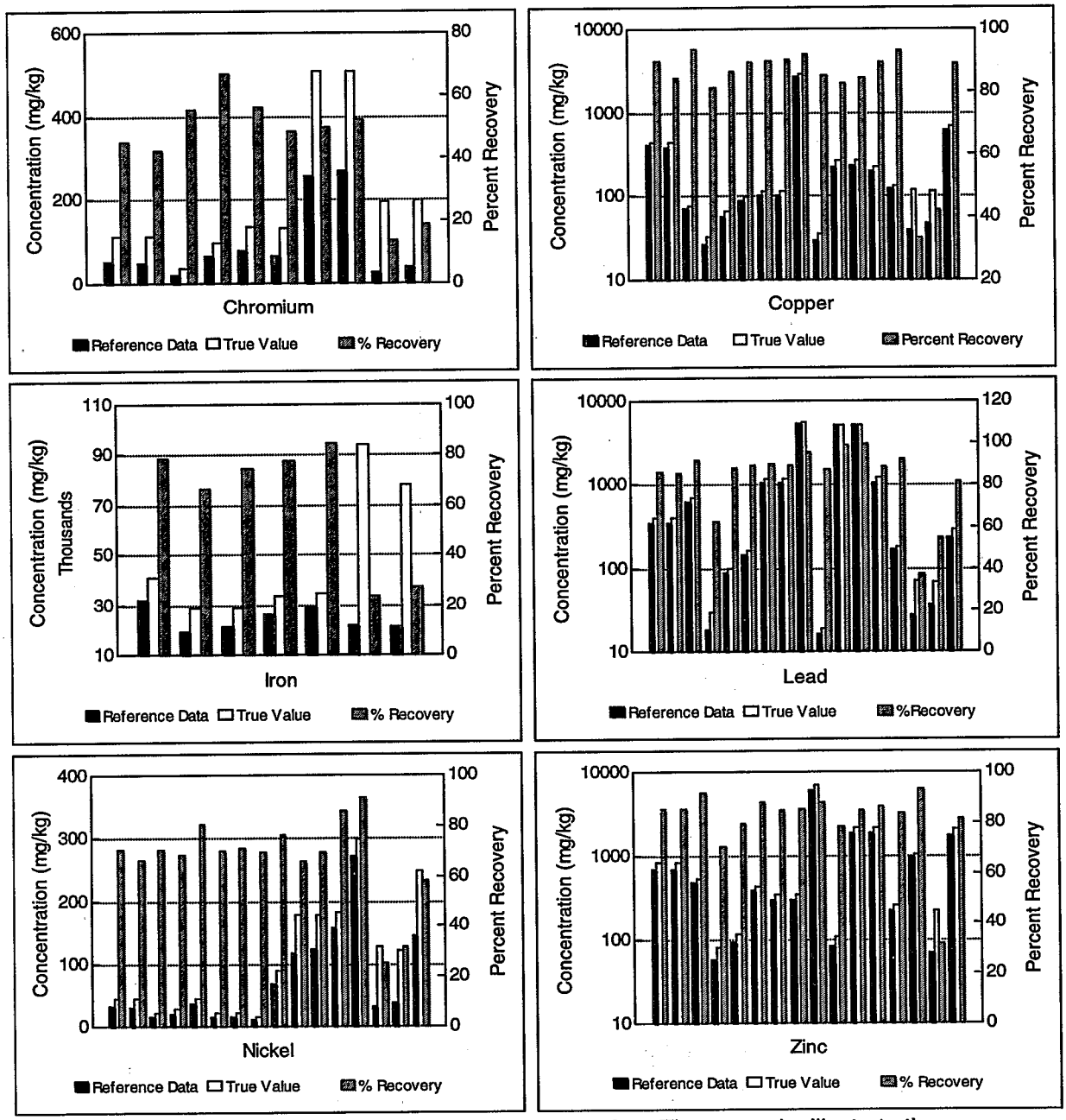


Figure 3-3 (Continued). Reference Method SRM Results: These graphs illustrate the relationship between the reference data and the true values for the SRM samples. The gray bars represent the percent recovery for the reference data. Each set of three bars (black, white, and gray) represents a single SRM sample.

Section 4
SEFA-P Analyzer

This section provides information on HNU's SEFA-P Analyzer including theory of FPXRF analysis, operational characteristics, performance factors, a data quality assessment, and a comparison of results with those of the reference laboratory.

Theory of FPXRF Analysis

FPXRF analyzers operate on the principle of energy dispersive XRF spectrometry. This is a nondestructive qualitative and quantitative analytical technique that can be used to determine the metals composition in a test sample. By exposing a sample to an X-ray source having an excitation energy close to, but greater than, the binding energy of the inner shell electrons of the target elements, electrons are displaced. The electron vacancies that result are filled by electrons cascading in from outer shells. Electrons in outer shells have higher potential energy states than inner shell electrons, and to fill the vacancies, the outer shell electrons give off energy as they cascade into the inner shell (Figure 4-1). This release of energy results in an emission of X-rays that is characteristic of each element. This emission of X-rays is termed XRF.

Because each element has a unique electron shell configuration, each will emit unique X-rays at fixed wavelengths called "characteristic" X-rays. The energy of the X-ray is measured in electron volts (eV). By measuring the peak energies of X-rays emitted by a sample, it is possible to identify and quantify the elemental composition of a sample. A qualitative analysis of the sample can be made by identifying the characteristic X-rays produced by the sample. The intensity of each characteristic X-ray is proportional to the concentration of the source and can be used to quantitate each element.

Three electron shells are generally involved in the emission of characteristic X-rays during FPXRF analysis: the K, L, and M shells. A typical emission pattern, also called an emission spectrum, for a given element has multiple peaks generated from the emission X-rays by the K, L, or M shell electrons. The most commonly measured X-ray emissions are from the K and L shells; only elements with an atomic number of 58 (cerium) or greater have measurable M shell emissions.

Each characteristic X-ray peak or line is defined with the letter K, L, or M, which signifies which shell had the original vacancy and by a subscript alpha (α) or beta (ß), which indicates the next outermost shell from which electrons fell to fill the vacancy and produce the X-ray. For example, a K_α-line is produced by a vacancy in the K shell filled by an L shell electron, whereas a $K_ß$-line is produced by a vacancy in the K shell filled by an M shell electron. The K_α transition is between 7 and 10 times more probable than the $K_ß$ transition. The K_α-line is approximately 10 times more intense than the $K_ß$-line for a given element, making the K_α-line analysis the preferred choice for quantitation purposes. Unlike the K-

lines, the L-lines (L_α and L_β) for an analyte are of nearly equal intensity. The choice of which one to use for analysis depends on the presence of interfering lines from other analytes.

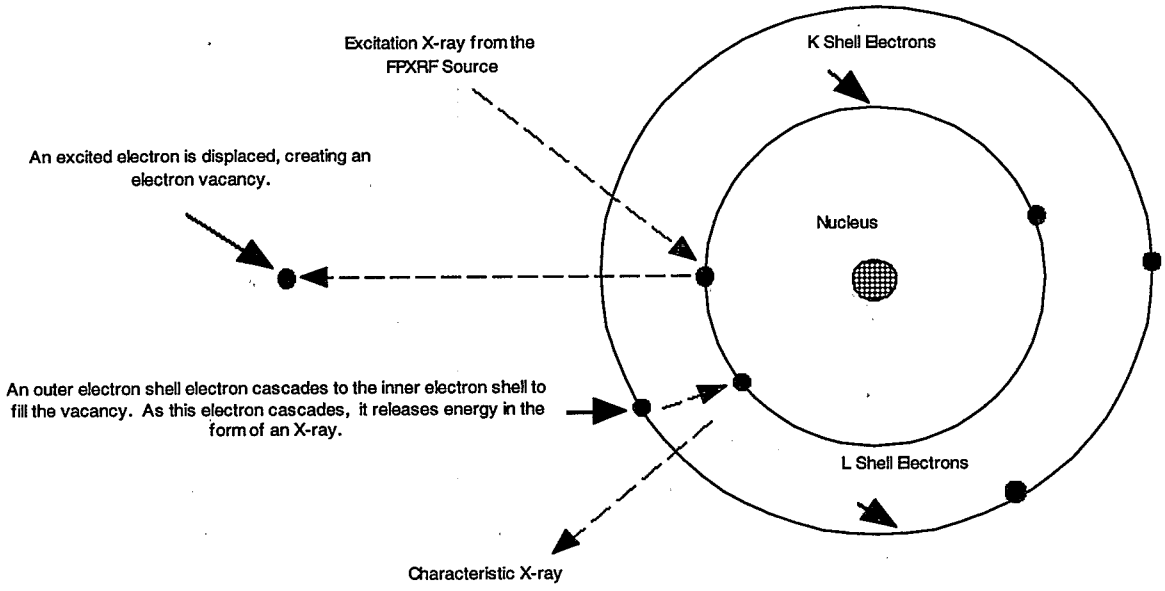


Figure 4-1. Principle of Source Excited X-ray Fluorescence: This figure illustrates the dynamics of source excited X-ray fluorescence.

An X-ray source can excite characteristic X-rays from an analyte only if its energy is greater than the electron binding energies of the target analyte. The electron binding energy, also known as the absorption edge energy, represents the amount of energy an electron has to absorb before it is displaced. The absorption edge energy is somewhat greater than the corresponding line energy. Actually, the K-absorption edge energy is approximately the sum of the K-, L-, and M-line energies of the particular element, and the L- absorption edge energy is approximately the sum of the L- and M-line energies. FPXRF analytical methods are more sensitive to analytes with absorption edge energies close to, but less than, the excitation energy of the source. For example, when using a Cd^{109} source, which has an excitation energy of 22.1 kiloelectron volts (keV), an FPXRF analyzer would be more sensitive to zirconium, which has a K-line absorption edge energy of 15.7 keV, than to chromium, which has a K-line absorption edge energy of 5.41 keV.

Background

HNU is an international supplier of environmental monitoring instrumentation. Besides the SEFA-P Analyzer, HNU also manufactures portable gas chromatographs, photoionization detectors, microanalysis systems, sensitive balances for scientific gas blending, and other ultra precise weighing instruments.

The SEFA-P Analyzer has been on the market for more than 6 years and is designed for use as a field screening instrument. It is a transportable analyzer that operates in the intrusive mode. HNU states that the SEFA-P has been proven effective in determining heavy metal concentrations in water and soil samples; metals in mixed waste at radiation contaminated sites; and paint, soil, and water for lead content and contamination.

HNU developed the SEFA-P Analyzer as a stand-alone system to determine metal concentrations in a variety of matrices. The analyzer contains a four-position sample tray to increase sample throughput and can be equipped with up to three radioactive sources, iron-55 (Fe^{55}), cadmium-109 (Cd^{109}), and americium-241 (Am^{241}), for the identification and quantification of a wide range of metal analytes. The SEFA-P is equipped with a high resolution Si(Li) detector and a 4,096-channel multichannel analyzer (MCA) that offers rapid interpretation of data from the detector. The SEFA-P Analyzer comes with an on-board operating system that can control the instrument, analyze samples, produce sample spectra, and store sample data. It contains an onboard cathode ray tube (CRT) that displays real-time spectra for qualitative analysis. An optional personal computer (PC) software program can also perform these same tasks while providing enhanced graphical imagery of sample spectra and expanded data storage capabilities.

Analysis of demonstration samples with the SEFA-P Analyzer was performed after the instrument supplied by the developer experienced a failure when the source holder locked into place, not allowing further sample exposure and analysis. This problem occurred at the start of the demonstration and HNU could not supply a replacement unit to complete the demonstration. NERL-ESD provided an EPA-owned SEFA-P Analyzer to analyze a subset of the demonstration samples. One hundred demonstration samples were randomly selected to be analyzed. These samples represented all three demonstration soil textures and a wide range of concentrations from each of the demonstration sites and included all of the PE and SRM samples. The PRC operator traveled to the NERL-ESD facility and performed this sample analysis with the substitute SEFA-P Analyzer from July 18 to July 20, 1995.

Operational Characteristics

This section discusses equipment and accessories, operation of the analyzer in the field, background of the operator, training, reliability of the analyzer, health and safety concerns, and representative operating costs.

Equipment and Accessories

The SEFA-P Analyzer comes with all of the equipment necessary for intrusive operation with the exception of an optional PC that can be used with the SEFA-PC software. A padded wooden storage box is provided for shipping and storage. Specifications for the SEFA-P Analyzer are provided in Table 4-1.

Major components of the instrument include: the main cabinet, a battery and charger, a liquid nitrogen dewar, the sample chamber, radioactive sources, a liquid nitrogen cooled Si(Li) X-ray detector, a preamplifier, a multi-channel analyzer, an RS-232 serial interface, and a PC software program.

Environmental requirements for the SEFA-P Analyzer include: a relative humidity of between 20 and 95 percent (noncondensing) and an ambient temperature range from 0 to 40 °C. The instrument can be used indoors or outdoors as long as these environmental conditions are met.

The instrument can be operated in four different electrical configurations: (1) with 110-volts AC, (2) with 220-volts AC, (3) with a built-in 12-volt direct current (DC) battery and charger, and (4) with a cable that can be plugged into an automobile cigarette lighter. The built-in battery provided with the instrument can provide approximately 8 hours of operational power before requiring a recharge.

The SEFA-P Analyzer requires liquid nitrogen to stabilize the detector crystal and to allow the preamplifier to function efficiently. The instrument contains an internal dewar capable of holding 0.85

liters of liquid nitrogen. This volume allows for approximately 24 hours of instrument operation. The analyzer contains a four-position sample tray that is loaded with sample cups prior to analysis. The sample tray can accept both 32- and 40-mm sample cups. An insert is removed from the sample tray to allow the use of the 40-mm sample cups.

Table 4-1. Analyzer Instrument Specifications

Characteristic	Specification
Sources	10 millicurie (mCi) Cd^{109} and 25 mCi Am^{241} (50 mCi Fe^{55} also available)
Detector	Si(Li)-Liquid nitrogen cooled
Resolution	170 eV (Mn-K_α)
Analyzer Size	16 in x 21 in x 12 in
Analyzer Weight	132 kg
Internal Liquid Nitrogen Dewar Size	0.85 liters
Optional PC Minimum Requirements	Standard 386, 20 MB hard drive, 1 MB RAM (A laptop increases portability)
Power Source	120V or 240V (AC), 12V DC, or internal battery
Operational Checks	Pure copper check sample
Contact: Mr. John Moore HNU Systems, Inc. Charlemont Street Newton Highlands, MA 02161-9987 Phone: (617) 964-6690 Toll Free: (800) 724-5600 Fax: (617) 558-0056	

Three sources can be installed in the SEFA-P instrument: the Am^{241}, the Cd^{109}, and the Fe^{55}. The instrument used for the analysis of demonstration samples only contained the Am^{241} and Cd^{109} sources. The SEFA-P Analyzer uses a Si(Li) detector that provides high resolution and low background noise.

The multichannel analyzer is equipped with an RS-232 serial interface to enable data acquisition and analysis using a software system developed by HNU and called the SEFA-PC program. Minimum requirements for the PC used with the SEFA-PC software are a standard 386 computer with a minimum of a 20-megabyte hard drive, and 1 megabyte of random access memory (RAM). The software offers increased data storage capabilities and enhanced graphics.

Operation of the Analyzer

Prior to operating the SEFA-P Analyzer, samples must be prepared for analysis. The analyzer is designed to be used in an intrusive mode. That is, the sample must be removed from the ground and placed in a sample cup to be analyzed. HNU recommends that soil and sediment samples be dried, thoroughly mixed, and passed through a No. 60 mesh sieve to achieve the most accurate results. All demonstration samples analyzed with the analyzer were intrusive-prepared, which is the most thorough preparation step (Section 2). Prepared samples were placed in sample cups, ensuring that the Mylar™ film was flat and smooth. Each sample cup was filled with approximately a 10-mm-thick layer of sample.

The operation of the SEFA-P Analyzer can be described in four steps: (1) setting up the instrument, (2) performing the calibration, (3) analyzing samples, and (4) managing data. The following paragraphs provide details concerning each of these four steps of the operation of the analyzer.

The SEFA-P Analyzer requires liquid nitrogen to cool the detector. The detector requires approximately 30 minutes of cool-down time before it reaches operating temperature. The 0.85-liter internal dewar holds enough liquid nitrogen to operate the instrument for up to 24 hours. During the analysis of demonstration samples, the internal dewar was filled each morning before sample analysis. The detector operated for 10 to 12 hours each day without the need to refill the dewar.

An energy calibration was performed each morning. The energy calibration is a two-point check that determines the intercept (keV) at channel zero and the slope (keV per channel) to make a straight line fit. The calibration requires the analysis of a sample that contains two known energy emissions. HNU recommends the analysis of a copper foil sample with either the Cd^{109} or Am^{241} source. The intercept and slope of the energy calibration changed very little from day to day, illustrating a consistent linearity of an energy calibration. This consistency is important to maintain the integrity of the qualitative identification of target analytes being analyzed with the instrument.

Instrument setup also includes setting the MCA or PC software to match the radioactive sources installed in the instrument. Two radioactive sources were used during the analysis of demonstration samples, the Cd^{109} and the Am^{241}. During the analysis of demonstration samples with the SEFA-P Analyzer, 10 target analytes were monitored: arsenic, lead, chromium, copper, iron, nickel, and zinc which were monitored with the Cd^{109} source; and barium, cadmium, and antimony, which were monitored with the Am^{241} source. The K_α line of arsenic and the L_α line of lead overlap, so the operator chose to monitor for the K_β line for arsenic and the L_β line for lead.

Prior to the analysis of samples, it is recommended that an instrument calibration be performed. However, sample analysis can be initiated before performing this calibration. An instrument calibration can be performed using an empirical calibration or a Compton calibration technique. Both calibration techniques can be performed using either site-specific soil samples containing known concentrations of the target analytes or synthetic standards prepared with metal oxides mixed in an inert matrix, such as lithium carbonate or silicon dioxide. The empirical calibration involves the analysis of multiple calibration standards containing varying concentrations of each target analyte, while the Compton calibration uses a single standard containing one concentration level of each target analyte. Both the empirical and Compton calibration techniques were used for the analysis of demonstration samples.

Synthetic calibration standards used for the analysis of demonstration samples consisted of metal oxide standards mixed in lithium carbonate. Synthetic calibration standards included antimony, arsenic, barium, cadmium, chromium, copper, iron, lead, nickel, and zinc. A standard curve for iron is shown in Figure 4-2. This standard curve is similar to the standard curves obtained from the other target analytes.

An empirical site-specific calibration was performed using three samples collected and analyzed from each site during the predemonstration activities. Predemonstration analysis of these site-specific standards included analysis using a laboratory-grade XRF and reference method analysis. The concentration values for target analytes in the site-specific soil samples entered into the SEFA-PC software were generated using the reference method results, allowing the calibration to more closely match the reference method data. The three samples from each site contained varying concentrations of the target analytes, but the main advantage to the use of the site-specific samples was that the samples were representative of the sample matrix at each site.

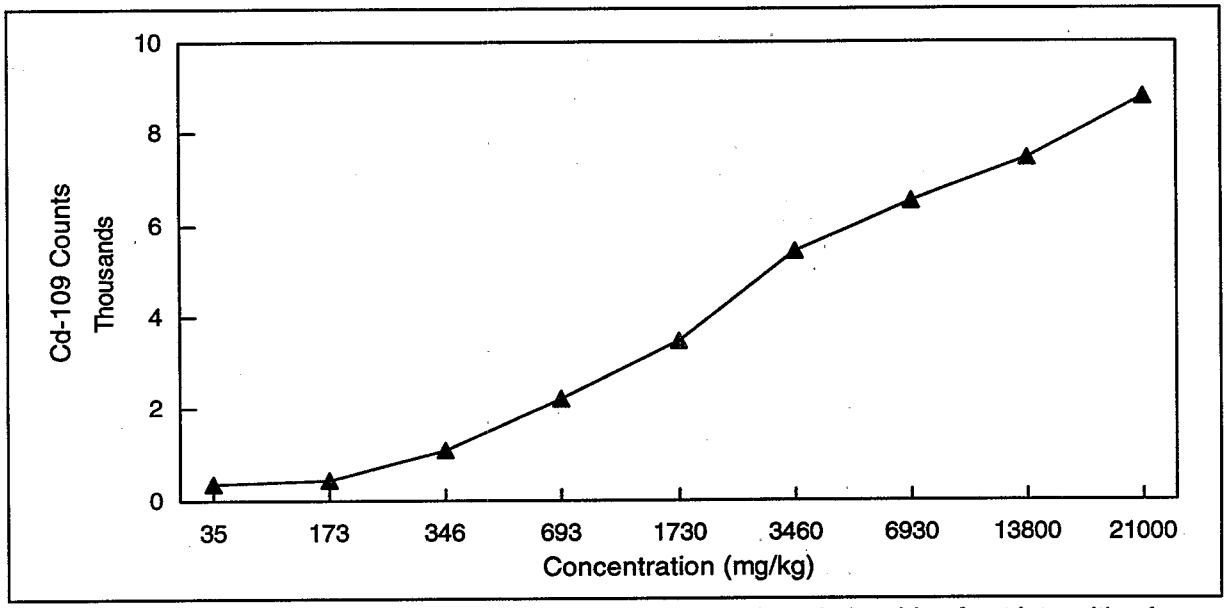


Figure 4-2. Standard Curve for Iron: This graph illustrates the relationship of net intensities for iron as derived from the HNU SEFA-P Analyzer versus increasing concentrations of iron in the synthetic calibration standards.

A synthetic standard was used for the single-point Compton calibration. This standard contained approximately 750 mg/kg of each target analyte. All demonstration samples were analyzed using this Compton calibration. This standard contains metals concentrations closer to potential action levels and thus, was appropriate for general environmental applications.

During the analysis of demonstration samples, results were reported using the Compton calibration technique. Verification of the validity of the initial calibration was performed through the use of a continuing calibration. The continuing calibration standard was the same standard used for the Compton calibration. The response of the continuing calibration was compared to the response of the same standard analyzed during the initial calibration using a percent difference (%D) calculation.

The overlap of fluorescence spectra can cause errors in the reported results. Demonstration calibration standards appeared to have no overlap of spectra for the specific K or L electron shell energies monitored for each target analyte. Any overlaps present in the samples may adversely affect the sample results generated with the SEFA-P. To ensure accurate results, the user should construct a table that includes all the alpha and beta energies of the K, L, and M electron shells of all metals that may be found in soil and sediment samples. This table could be used to determine any possible overlap of fluorescence spectra from metals in samples. This will allow the user to build suitable strategies to avoid effects from overlapping fluorescence spectra. This may include use of a different electron shell or monitoring the beta energy rather than the alpha energy of a metal.

Absorption or enhancement of certain metals from other metals is also a problem in XRF analysis. The SEFA-PC software provides an option that corrects results based upon absorption and enhancement effects. This option was not used for demonstration samples because the operator was not familiar with how to determine when absorptions and enhancements were affecting data nor how to use this option to best handle these effects. The instruction manual provided by the developer gave little detail on this option and no information concerning how and when to apply this option. Due to operator inexperience, absorption and enhancement effects may have impacted demonstration results.

Background of the Technology Operator

The PRC operator chosen for analyzing soil samples using the SEFA-P Analyzer has been a PRC employee for 4 years. He holds a bachelor's degree in geography and a minor in chemistry. While at PRC, he has worked largely on field screening projects involving on-site analysis of samples for organic contaminants. Prior to working for PRC, he spent 6 years working in an environmental laboratory performing analysis of samples for organic contaminants. He had no prior experience performing XRF sample analysis.

Training

The operator traveled to the HNU facility in Newton, Massachusetts, for a 1-day training course. One-quarter of the day was dedicated to the theory of XRF technology and one-quarter of the day was spent on safety issues associated with the operation of the SEFA-P Analyzer. Safe handling of liquid nitrogen and radiation safety issues were discussed. The remaining one-half day was a hands-on training on the operation of the SEFA-P. This course included: the mechanics of filling the unit with liquid nitrogen, instrument start-up, panel control operation, use of the SEFA-PC software, and sample analysis. Calibration techniques also were discussed. The training received by the operator was comparable to training provided to anyone who purchases this analyzer.

Reliability

More than 267 individual measurements were collected with the SEFA-P Analyzer. This included the measurement of 100 soil samples prepared using the intrusive-prepared technique, 10 replicate measurements on 12 samples for a precision assessment, and the measurement of QC samples such as blanks, PE samples, SRMs, and continuing check standards. While collecting these measurements over a period of 3 working days, no problems were encountered that affected the reliability of the analyzer.

Health and Safety

The potential for exposure to radiation from the excitation source is the largest health and safety consideration while using the analyzer. External exposure to radiation from the instrument is reported by HNU to be less than 0.25 millirems per hour at a distance of 5 centimeters from the instrument with the sample access door open or closed. Before beginning the operation of the SEFA-P, radiation levels were monitored in and around the instrument. Background radiation was approximately 0.008 millirems per hour. The highest radiation levels obtained in and around the SEFA-P were 0.035 millirems per hour near the upper left-hand side of the front of the main cabinet. These readings were obtained approximately 1 centimeter from the instrument. These readings verified a claim made by HNU that the external exposure to radiation from the SEFA-P is less than 0.25 millirems per hour 5 centimeters from the instrument. Radiation levels from the SEFA-P also are less than the permissible occupational exposure value in Kansas of 5,000 millirems per year, which equates to approximately 2 to 3 millirems per hour assuming constant exposure for an entire work year. Radiation wipe tests must be performed at 6-month intervals to verify the integrity of the sealed sources. This is a requirement of the Nuclear Regulatory Commission general license granted to HNU.

Transferring liquid nitrogen from an external dewar to the internal dewar of the SEFA-P Analyzer was another health and safety consideration. Due to the extremely low temperature of liquid nitrogen, the operator must take care to avoid contact with the liquid nitrogen during the filling operation. Safety goggles and gloves must be worn during filling operations. It is also recommended that a laboratory coat

be worn when filling the SEFA-P with liquid nitrogen. When pouring liquid nitrogen from the external dewar into the funnel attached to the internal dewar, the operator should avoid splashing the liquid nitrogen and must be aware that venting processes within the internal dewar may cause some liquid nitrogen to splash out of the funnel.

Cost

At the time of this demonstration, the cost of a new SEFA-P Analyzer was $37,000, excluding the cost of the radioactive sources. Each radioactive source costs $4,000. Therefore, a SEFA-P Analyzer equipped with all three radioactive sources would cost $49,000. This includes the SEFA-PC software and 1 day of training. HNU offers a separate 2-day, in-house XRF training course for a cost of $750 per person. Travel and accommodation costs for the training are not included. A laptop PC can be rented for use of the SEFA-PC software for approximately $300 per month. Periodic maintenance includes radioactive source leak testing which is required at 6-month intervals and is performed at HNU's facility in Newton, Massachusetts. HNU offers a wipe test and system optimization that includes an instrument cleaning, an electronic diagnostic check, and an overall system check for a cost of $500. The Cd^{109} source requires disposal and replacement every 2 years at a cost of $4,000 with an additional $500 disposal fee due to the licensing requirements for the radioactive sources. HNU does not rent the SEFA-P Analyzer and does not know of any firm that rents the unit.

The primary cost benefit of field analysis is the quick access to analytical data. This allows the process dependent on the testing to move efficiently onto the next stage. Costs associated with field analysis are dependent on the scope of the project. Since most of the mobilization costs are fixed, analyzing a large number of samples lowers the per sample cost. This is a key advantage that field analysis has over a conventional laboratory. Furthermore, more samples are usually taken for field analysis since questions raised in the preliminary findings may be resolved completely without the need to return for another sample collection event.

A representative list of the operating costs associated with the SEFA-P is presented in Table 4-2. Also included in this table is the measured throughput and the per sample charge of the reference laboratory. Given the special requirements of this demonstration, it was not considered fair to report a per sample cost for the field analysis. However, some estimate can be derived from the data provided in this table.

Performance Factors

The following paragraphs describe performance factors, including detection limits, sample throughput, and drift.

Detection Limits

HNU acknowledges that the detection limits for soil samples are highly matrix dependent, but that the SEFA-P Analyzer should be able to determine metals concentrations in soil samples down to levels of 100 mg/kg. The only exception to this rule is for chromium, which the SEFA-P claims to be able to detect at a concentration of 200 mg/kg in soil samples. Table 4-3 lists detection limits reported by the developer and those determined during this demonstration.

MDLs, using SW-846 protocols, were determined by collecting 10 replicate measurements on site-specific soil samples with metals concentrations 2 to 5 times the expected MDLs. These data were

obtained during the precision evaluation. Based on this data, a standard deviation was calculated and the MDLs were defined as 3 times the standard deviation for each target analyte. All the precision-based MDLs were calculated for soil samples that had been dried and ground and placed in a sample cup, the highest degree of sample preparation. Precision data was derived from the SEFA-P Analyzer results from the Compton calibration. The precision-based MDLs for the SEFA-P are shown in Table 4-3.

Table 4-2. Summary of Analysis Costs

Item	Amount	
SEFA-P Analyzer (3 sources)	$ 49,000	Purchase Price
Replacement Source	4,000	—
Operator Training (Vendor Provided)	750	—
Radiation Safety License (State of Kansas)	500	—
Field Operation Costs		
Supplies and Consumables (Sample cups, window film, sieves, standards)	300 - 500	(Varies, depending on sample load)
Field Chemist (Labor Charge)	100 - 150	Per day
Per diem	80 - 120	Per day
Travel	200 - 500	Per traveler
Sample Throughput	7 - 8	Samples per hour
Cost of Reference Laboratory Analysis	150	Per sample

The precision-based MDLs were obtained using the same 180 second count times for the Cd^{109} source and 60 second count times for the Am^{241} source as was used for all samples analyzed with the SEFA-P Analyzer. The counting statistics for FPXRF analysis indicate that it would take a fourfold increase in count-time to increase the precision and therefore reduce MDLs by 50 percent. In other words, it would require count times of 720 seconds for the Cd^{109} source and 240 seconds for the Am^{241} source to realize MDLs of one-half the value listed in Table 4-3.

Table 4-3. Method Detection Limits

Analyte	Developer-listed Detection Limits (mg/kg)	Precision-based MDL (mg/kg)	Field-based MDL (mg/kg)
Antimony	100	120	165
Arsenic	100	360	600
Barium	100	1,150	Not Determined
Cadmium	100	Not Determined	135
Chromium	200	Not Determined	1,250
Copper	100	225	320
Iron	100	900	Not Determined
Lead	100	120	225
Nickel	100	Not Determined	Not Determined
Zinc	100	990	1,400

Note: mg/kg Milligrams per kilogram.

Another method of determining MDLs involved the direct comparison of the FPXRF data and the reference method data. When these sets of data were plotted against each other the resultant plots were linear. As the plotted line approached zero for either method, there was a point at which the FPXRF data appeared to respond either randomly or with the same reading for decreasing concentrations of the reference data. This point was determined by observation and was somewhat subjective; however, a sensitivity analysis showed that even 25 percent errors in determining this point resulted in up to 10 percent changes in the MDL calculation. By determining the mean values of this FPXRF data and subsequently two standard deviations around this mean, it was possible to determine a field or performance-based MDL for the analyzer. For the SEFA-P Analyzer, these field-based MDLs are shown in Table 4-3.

In the sample chosen for this demonstration, iron was found mostly at tens of thousands of milligrams per kilogram, and barium was found at concentrations at thousands of milligrams per kilogram, so that reasonable detection limits could not be calculated. Nickel concentrations were too low to determine an MDL.

Throughput

A total of 237 analyses was conducted over a period of three 10-hour work days. The SEFA-P Analyzer used a total live-second count time of 240 seconds. With the additional "dead" time of the detector, the time required to print the data, and the time required to enter samples into the SEFA-PC software, the time required to analyze one soil sample was 7 to 8 minutes. This resulted in a sample throughput of 7 to 8 samples per hour.

The sample analysis time did not include the time required for sample handling and preparation, or for data downloading and documentation. Considerable time was spent preparing the intrusive soil samples for analysis. Homogenization required an average of 5 minutes per sample and grinding and sieving for the intrusive-prepared sample required an additional 10 minutes. The operator noted that it took 15 minutes to fill the dewar with liquid nitrogen each morning and 30 minutes for the detector to cool sufficiently to allow sample analysis. The PRC operator used this time to perform data calculations, data management tasks such as saving spectra and result files to disk, and segregating and preparing for the day's samples. On average, it took about 1 hour each morning to prepare the SEFA-P for sample analysis activities.

Downloading of the data from the hard drive of the PC took approximately 3 hours for the entire set of demonstration data. Transferring data from floppy disks to the hard drive of a computer for reprocessing also took 3 hours. Recalibration and reanalysis of all of the demonstration samples required approximately 1 hour.

Drift

Drift is a measurement of an analyzer's variability in quantitating a known amount of a standard over time. An evaluation of the SEFA-P Analyzer's qualitative drift was performed with the energy calibration check performed each morning before beginning sample analysis. The energy calibration provided consistent intercept and slope values for each of the energy calibrations performed during the analysis of demonstration samples. Quantitative drift was evaluated through the use of a continuing calibration check standard. The continuing calibration check standard was the same standard used to perform the Compton calibration for the analysis of the demonstration samples. Counts obtained for each continuing calibration check standard performed were compared to the original analysis of the standard, and a percent difference

calculation was performed. In all, seven continuing calibration check standards were analyzed, one at the beginning and one at the end of each work day. Percent differences ranged from 0 to 35 percent for all of the target analytes (Figure 4-3).

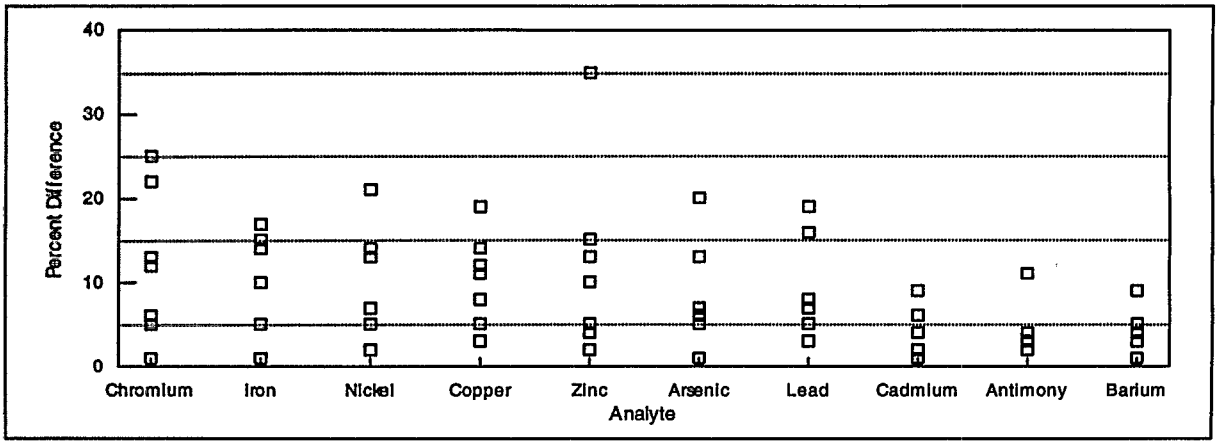


Figure 4-3. Drift Summary: This graph illustrates the analyzer's drift as determined through the use of a continuing calibration standard. The percent difference on the y axis reflects the absolute value of percent difference.

Intramethod Assessment

Intramethod assessment measures of the analyzer's performance include results on analyzer blanks, completeness of the data set, intramethod precision, intramethod accuracy, and intramethod comparability. The following paragraphs discuss these characteristics.

Blanks

Analyzer blanks for the SEFA-P Analyzer consisted of pure lithium carbonate. The blanks were placed directly in a sample cup after all four preparation steps like the dried and ground soil samples. The blanks were used to monitor for cross contamination from the sample preparation steps and to monitor for background readings produced by the analyzer. Three blanks were analyzed along with demonstration samples by the SEFA-P. A number of metals were detected in these blank samples. Because the concentration of metals in the blank samples did not exceed the precision-based or field-based MDLs, blank contamination is not believed to have impacted demonstration sample results.

Completeness

The SEFA-P produced data for 100 out of the 100 samples for a completeness of 100 percent. However, prior to the mechanical failure in the field, the SEFA-P was to have analyzed 630 samples.

Precision

Precision was expressed in terms of the percent RSD between replicate measurements. The precision data for the target analytes detected by the analyzer are shown in Table 4-4. The precision data reflected in the 5 to 10 times the MDL range reflects the precision generally referred to in analytical methods such as SW-846.

Table 4-4. Precision Summary

Analyte	Mean %RSD Values by Concentration Range			
	5 - 10 Times MDL[a] (mg/kg)	50 - 500 (mg/kg)	500 - 1,000 (mg/kg)	>1,000 (mg/kg)
Antimony	7.20 (1)	32.93 (3)	9.07 (2)	ND
Arsenic	ND	ND	25.58 (3)	11.45 (2)
Barium	6.99 (3)	ND	ND	7.11 (12)
Cadmium	ND	ND	ND	ND
Chromium	ND	ND	ND	ND
Copper	2.54 (1)	26.73 (8)	7.83 (1)	3.22 (3)
Iron	ND	ND	ND	2.44 (12)
Lead	4.86 (3)	22.82 (6)	5.90 (1)	3.11 (5)
Nickel	ND	ND	ND	ND
Zinc	ND	ND	ND	17.74 (8)

Notes:
[a] The MDLs referred to in this column are the precision-based MDLs shown in Table 4-3.
mg/kg Milligrams per kilogram.
ND No data.
( ) Number of samples, each consisting of 10 replicate analyses.

The SEFA-P Analyzer performed 10 replicate measurements on 12 soil samples that had analyte concentrations ranging from less than 50 mg/kg to tens of thousands of milligrams per kilogram. The replicate measurements were taken using the same source count times used for regular sample analysis. For each detectable analyte in each precision sample, a mean concentration, SD, and RSD were calculated.

In this demonstration, the analyzer's precision or RSD for a given analyte had to be less than or equal to 20 percent to be considered quantitative screening level data, and less than or equal to 10 percent to be considered definitive level data. The analyzer's precision data, reflected by its precision data in the 5 to 10 times MDL range, were below the 10 percent RSD required for definitive level data quality classification for antimony, barium, copper, and lead. No precision data within the range of 5 to 10 times the MDL was generated for the other six target metal analytes. Antimony, copper, and lead all provided RSD values of greater than 20 percent in the 50 - 500 mg/kg classification. No precision data within the range of 50 - 500 mg/kg was generated for the other seven target metal analytes. Antimony, copper, and lead provided RSD values less than 10 percent in the 500 - 1,000 mg/kg classification. Arsenic provided an RSD value of greater than 20 percent in this classification. No precision data within the range of 500 - 1,000 mg/kg was generated for the other six target analytes. Barium, copper, iron, and lead provided RSD values of less than 10 percent in the greater than 1,000 mg/kg classification. Arsenic and zinc provided an RSD of greater than 10 percent but less than 20 percent in this classification. No precision data greater than 1,000 mg/kg was generated for the other four target analytes.

There was a concentration effect on precision data as shown in Figure 4-4. The precision samples were purposely chosen to span a large concentration range to test the effect of analyte concentration on precision. As the concentration of the target analyte increased, the precision improved. In addition, Figure 4-4 shows an asymptotic relationship between concentration and precision. In this figure, precision

shows little improvement at concentrations greater than 1,000 mg/kg; however, at concentrations less than 1,000 mg/kg, precision is adversely affected as the concentration decreases. Although lead is shown in Figure 4-4, this trend is believed to be true for the other target analytes.

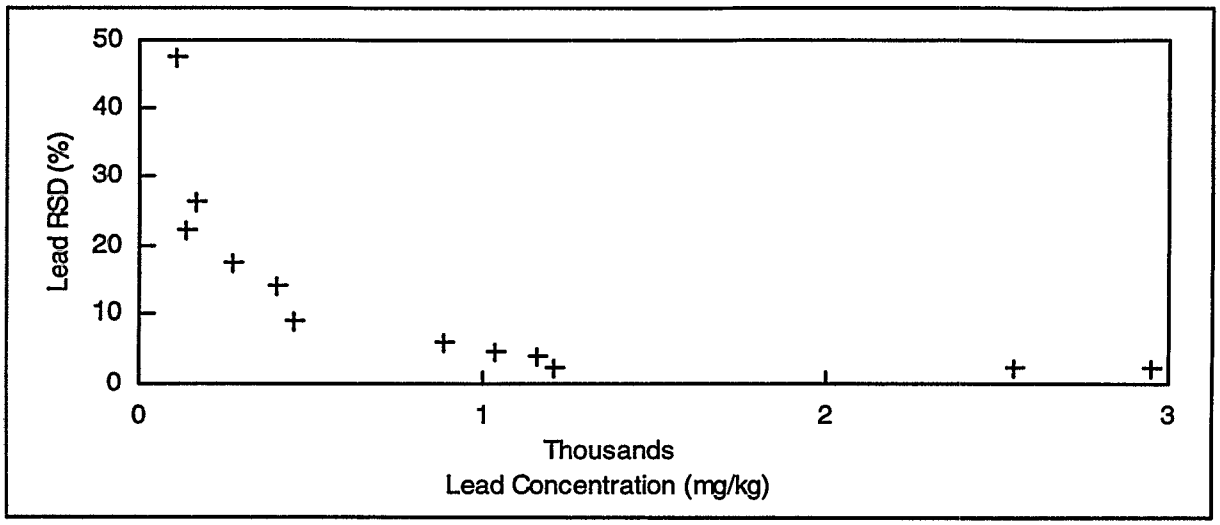


Figure 4-4. Precision vs. Concentration: This graph illustrates the analyzer's precision as a function of analyte concentration.

Accuracy

Accuracy refers to the degree to which a measured value for a sample agrees with a reference or true value for the same sample. Intramethod accuracy was assessed for the analyzer using site-specific PE samples and SRMs. Accuracy was evaluated through a comparison of percent recoveries for each target analyte reported by the analyzers. The analyzer measured six site-specific PE samples and 14 SRMs. The operator knew the samples were PE samples or SRMs, but did not know the true concentration or the acceptance range. These PE samples and SRMs were analyzed in the same way as all other samples.

The six site-specific PE samples consisted of three from each of the two demonstration sites that were collected during the predemonstration activities and sent to six independent laboratories for analysis by laboratory-grade XRF analyzers. The mean measurement for each analyte was used as the true value concentration. The 14 SRMs included seven soil, four stream or river sediment, two ash, and one sludge SRM. The SRMs were obtained from NIST, USGS, Commission of European Communities, National Research Council-Canada, and the South African Bureau of Standards. The SRMs contained known certified concentrations of certain target analytes.

These PEs and SRMs did not have published acceptance ranges. As specified in the demonstration plan, an acceptance range of 80 - 120 percent recovery of the true value was used to evaluate accuracy for the six site-specific PEs and 14 SRMs. Table 4-5 summarizes the accuracy data for the target analytes for the analyzers. Figures 4-5 and 4-6 show the true value, the measured value, and percent recovery for the individual site-specific PEs and SRMs, respectively. Analytes with two or less measured values greater than the precision-based MDLs are excluded from the figures.

Data generated with the Compton calibration technique using the 750 mg/kg concentration level standard were used to evaluate the accuracy of the HNU SEFA-P. True value results from the site-

specific PEs and SRMs with concentrations less than the precision-based MDLs listed in Table 4-3 were excluded from the accuracy assessment.

Table 4-5. Accuracy Summary for Site-Specific PE and SRM Results

Analyte	n	Percent Within Acceptance Range	Mean Percent Recovery	Range of Percent Recovery	SD of Percent Recovery	Concentration Range (mg/kg)
Site-Specific Performance Evaluation Samples						
Antimony	3	0	144	130 - 161	16	238 - 2,253
Arsenic	3	67	84	67 - 94	15	419 - 22,444
Barium	6	0	648	545 - 724	76	792 - 7,240
Cadmium	1	0	78	NA	NA	353
Chromium	1	0	54	NA	NA	3,800
Copper	5	40	78	58 - 113	21	300 - 7,132
Iron	6	17	69	55 - 89	12	27,320 - 70,500
Lead	6	33	84	59 - 126	23	292 - 14,663
Zinc	2	0	362	332 - 391	42	3,490 - 4,205
Soil Standard Reference Materials						
Arsenic	2	50	119	94 - 143	35	330 - 626
Barium	5	0	911	718 - 1,343	256	707 - 2,240
Copper	1	0	60	NA	NA	2,950
Iron	3	0	68	64 - 75	6	28,900 - 35,000
Lead	5	40	112	70 - 157	34	101 - 5,532
Zinc	2	0	426	356 - 495	98	1,055 - 6,952
Sediment Standard Reference Materials						
Antimony	1	0	133	NA	NA	171
Barium	3	0	688	554 - 770	117	335 - 414
Copper	3	33	144	90 - 173	47	219 - 452
Iron	3	0	65	64 - 67	2	41,100 - 197,100
Lead	4	25	97	64 - 152	40	161 - 5,200
Zinc	1	0	340	NA	NA	2,200
Ash and Sludge Standard Reference Materials						
Arsenic	1	0	159	NA	NA	136.2
Barium	2	0	902	829 - 974	103	709 - 1,500
Copper	1	0	76	NA	NA	696
Iron	2	0	67	65 - 68	2	77,800 - 94,000
Lead	1	100	117	NA	NA	286
Zinc	1	0	433	NA	NA	2,122

Notes:
n Number of samples with detectable analyte concentrations.
SD Standard deviation.
mg/kg Milligrams per kilogram.
NA Not applicable.

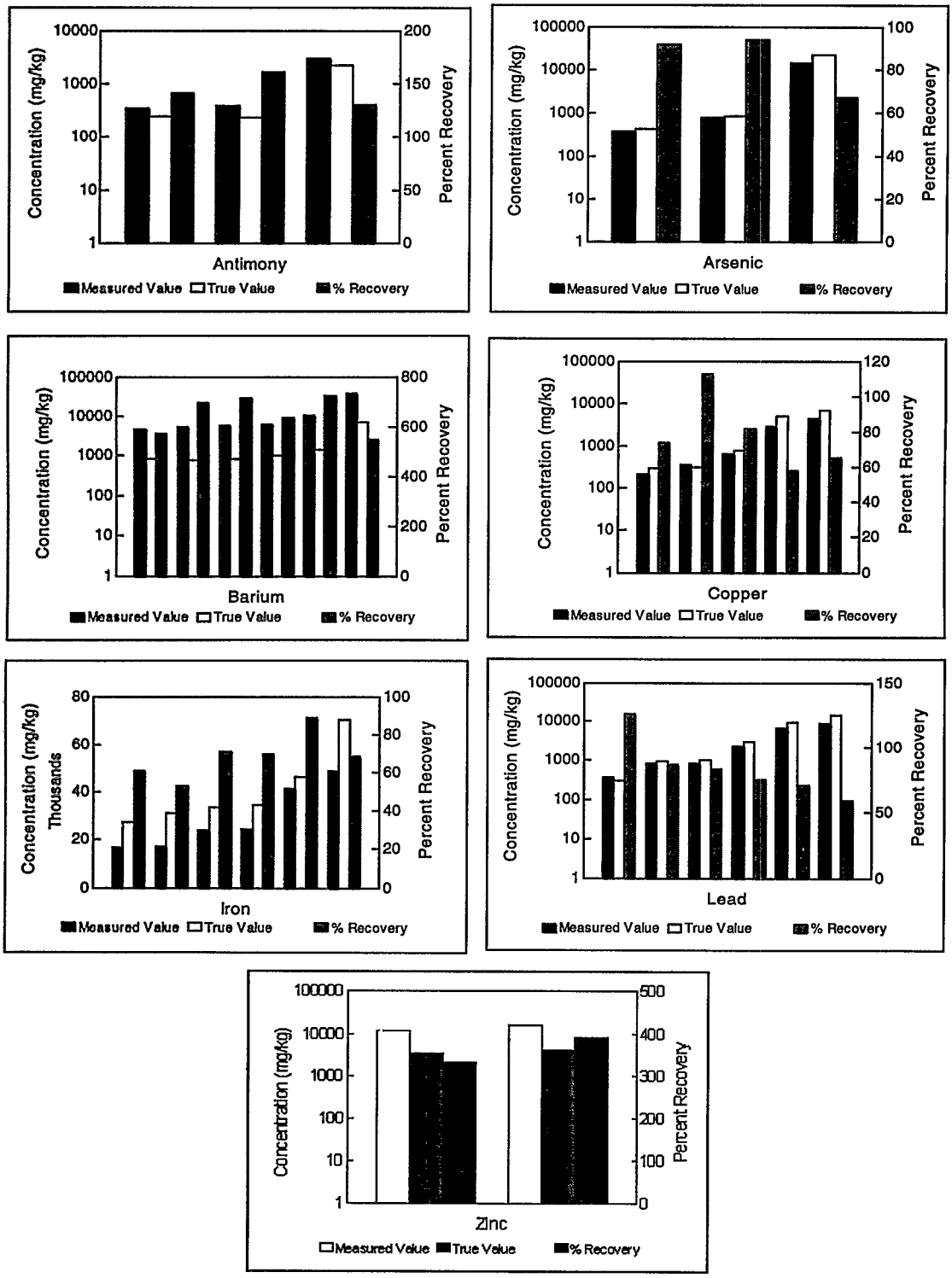


Figure 4-5. Site-specific PE Sample Results: These graphs illustrate the relationship between the analyzer's data (measured values) and the true values for the site-specific PE samples. The gray bars represent the percent recovery for the analyzer. Each set of three bars (black, white, and gray) represents a single site-specific PE sample.

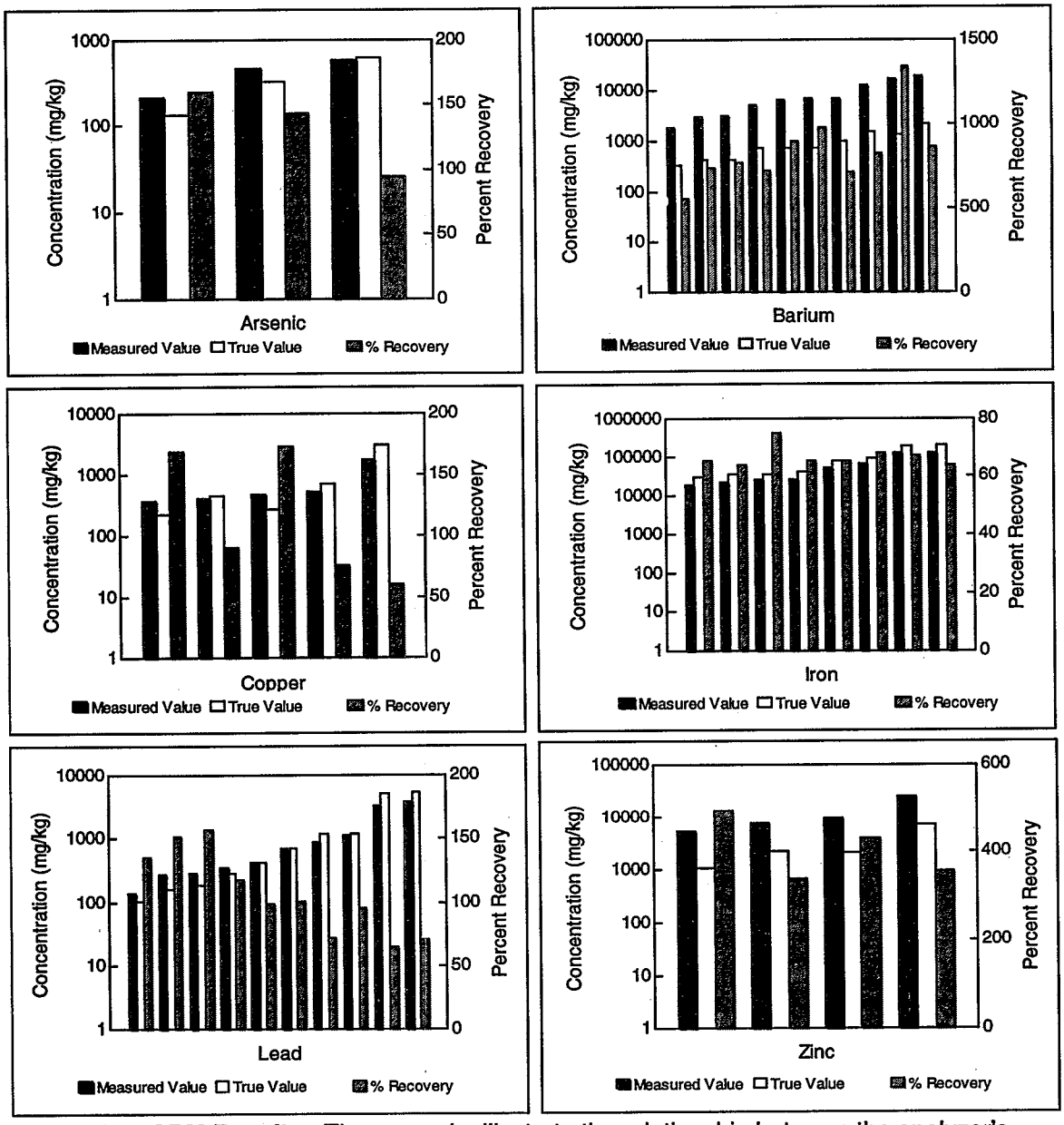


Figure 4-6. SRM Results: These graphs illustrate the relationship between the analyzer's data (measured values) and the true values for the SRMs. The gray bars represent the percent recovery for the analyzer. Each set of three bars (black, white, and gray) represents a single SRM sample.

A total of 33 site-specific PE results was obtained with concentrations greater than the precision-based MDLs. Seven of these results (21 percent) fell within the 80 - 120 percent recovery acceptance range. The analyzer overestimated the concentration of antimony in the samples with percent recoveries ranging from 130 to 161 percent. Arsenic was underestimated by the analyzer, but two of three arsenic results fell within the acceptance range. Barium was grossly overestimated by the analyzer with recoveries ranging from 545 to 724 percent.

Cadmium was reported low by the analyzer, but with a 78 percent recovery, this result was just outside of the acceptance range. Chromium was reported low by the analyzer as was iron. One of six iron results fell within the acceptance range. Copper and lead each provided two results within the acceptance range. Zinc was biased high, with recoveries ranging from 332 to 391 percent. It is believed that the high recoveries for zinc were an enhancement or spectral overlap effect from another element present in the samples, possibly copper. With the exception of barium and zinc, the accuracy of the analyzer was acceptable as evaluated with the site-specific PEs. Excluding barium and zinc, recoveries for the other eight analytes with concentration greater than the precision-based MDLs ranged from 54 to 161 percent. The analyzer provided good accuracy when compared to the site-specific PEs.

A total of 41 SRM results was obtained with concentrations greater than the precision-based MDLs. Six of these results (15 percent) fell within the 80 - 120 percent recovery acceptance range. Again, barium and zinc were grossly overestimated by the analyzer. Arsenic produced 1 of 4 results within the acceptance range. Lead provided 4 of 10 results within the acceptance range, while copper provided 1 of 5 results within the acceptance range. Excluding barium and zinc, recoveries ranged from 60 to 173 percent.

Overall, the analyzer produced 13 of 74 (18 percent) results that fell within the 80 - 120 percent recovery acceptance range. Given the intended uses of the SEFA-P for field screening, the accuracy of the analyzer, compared to site-specific PE samples analyzed with laboratory-grade XRFs and SRMs, appears to be adequate.

Comparability

Intramethod comparability for the analyzer was assessed through the analysis of four ERA PEs and four CRM PEs. This was done to present users additional information on data comparability relative to different commercially available QC samples. The eight PEs were analyzed in the same way as all other samples. As described in Section 3, these eight PEs had certified analyte values determined by EPA SW-846 Methods 3050A/6010A. Therefore, since these methods do not necessarily determine total metals concentrations in a soil, it was expected that the analyzer would overestimate analyte concentrations relative to PALs. The ability of the analyzer to produce results within the PALs and the percent recovery for each of the analytes was used to evaluate intramethod comparability. As with the site-specific PEs and SRMs, the Compton calibration technique using the 750 mg/kg standard was used for the comparability assessment. The SEFA-P Analyzer grossly overestimated barium concentrations relative to the true or reference values. This overestimation was magnified further in the comparability assessment with percent recoveries of barium exceeding 1,000 percent in some samples.

The analyzer's performance data for target analytes with concentrations greater than the precision-based MDLs listed in Table 4-3 are shown in Table 4-6. The measured values, true values, and percent recoveries for these analytes are shown in Figure 4-7. Analytes with two or less measured values greater than the precision-based MDLs are excluded from this figure.

For the ERA PEs, the SEFA-P produced four out of eight or 50 percent within the acceptance range. The ERA PEs contained low concentrations of metal analytes and resulted in a low number of data points to evaluate. Arsenic was present in one ERA PE sample at a concentration of 349 mg/kg which is near the precision-based MDL of 360 mg/kg. Lead values in these ERA PEs ranged from 128 to 208 mg/kg, near the precision-based MDLs of 120 mg/kg. The arsenic result fell within the acceptance range. Both of the lead values which fell outside of the acceptance range were overestimated by the analyzer. Three out of

four iron results were within the acceptance range. All iron results were overestimated compared to the reference values provided with the ERA PEs.

Table 4-6. PE and CRM Results

Analyte	n	Percent Within Acceptance Range	Mean Percent Recovery	Range of Percent Recovery	SD of Percent Recovery	Concentration Range (mg/kg)
ERA Performance Evaluation Samples						
Arsenic	1	100	98	NA	NA	349
Iron	4	75	131	117 - 157	19	7,130 - 10,400
Lead	3	33	161	96 - 207	58	128 - 208
Certified Reference Materials						
Antimony	1	100	119	NA	NA	4,955
Arsenic	1	100	85	NA	NA	397
Cadmium	2	50	153	122 - 184	43	362 - 432
Chromium	1	0	58	NA	NA	161,518
Copper	3	33	47	41 - 56	8	753 - 4,792
Iron	3	33	89	49 - 149	53	6,481 - 191,645
Lead	4	75	83	55 - 132	33	120 - 144,742
Nickel	1	100	118	NA	NA	13,279
Zinc	3	0	367	166 - 482	175	635 - 22,217

Notes:
n Number of samples with detectable analyte concentrations.
SD Standard deviation.
mg/kg Milligrams per kilogram.
NA Not applicable, analyte not present above the LRL.

For the CRM PEs, the analyzer produced nine out of 19 or 47 percent within the acceptance range. The CRM PEs generally contained higher concentrations of the analytes and resulted in more data points to evaluate. Concentrations of analytes in the CRMs ranged from 120 to nearly 200,000 mg/kg. Antimony, arsenic, and nickel each provided one result greater than the precision-based MDLs and the analyzer's result for each was within the acceptance range. Cadmium provided two results near, but above, the precision-based MDL. One of these results fell within the acceptance range. Cadmium results for the CRM PEs were biased high. Chromium provided one result within the acceptance range. Copper and iron each provided three results for the CRM PEs and each provided one result within the acceptance range.

Copper was biased low for all three results with recoveries ranging from 41 to 56 percent. Iron showed no bias trend with recoveries ranging from 49 to 149 percent. Lead provided four results for the CRM PEs with three falling within the acceptance range. Zinc was overestimated by the analyzer for the CRM PEs in much the same fashion as in the SRMs. Again, this is believed to be caused by an enhancement or spectral overlap effect from one of the other target analytes, most probably iron or copper. Excluding zinc, percent recoveries for the CRM PEs ranged from 41 to 184 percent.

Overall, the analyzer produced 13 out of 27 results or 48 percent within the acceptance ranges for the ERA and CRM PEs. Barium and zinc were grossly overestimated by the analyzer. Percent recoveries for

the other analytes ranged from 41 to 207 percent. The analyzer did overestimate the concentrations of some analytes, such as antimony, cadmium, and nickel. However, the analyzer did not overestimate the concentrations of iron or lead and underestimated the concentrations of chromium and copper. Although the analyzer's results for the CRM and ERA PEs fell within the acceptance ranges only 48 percent of the time, the analyzer can provide results that are comparable to EPA SW-846 Methods 3050A/6010A.

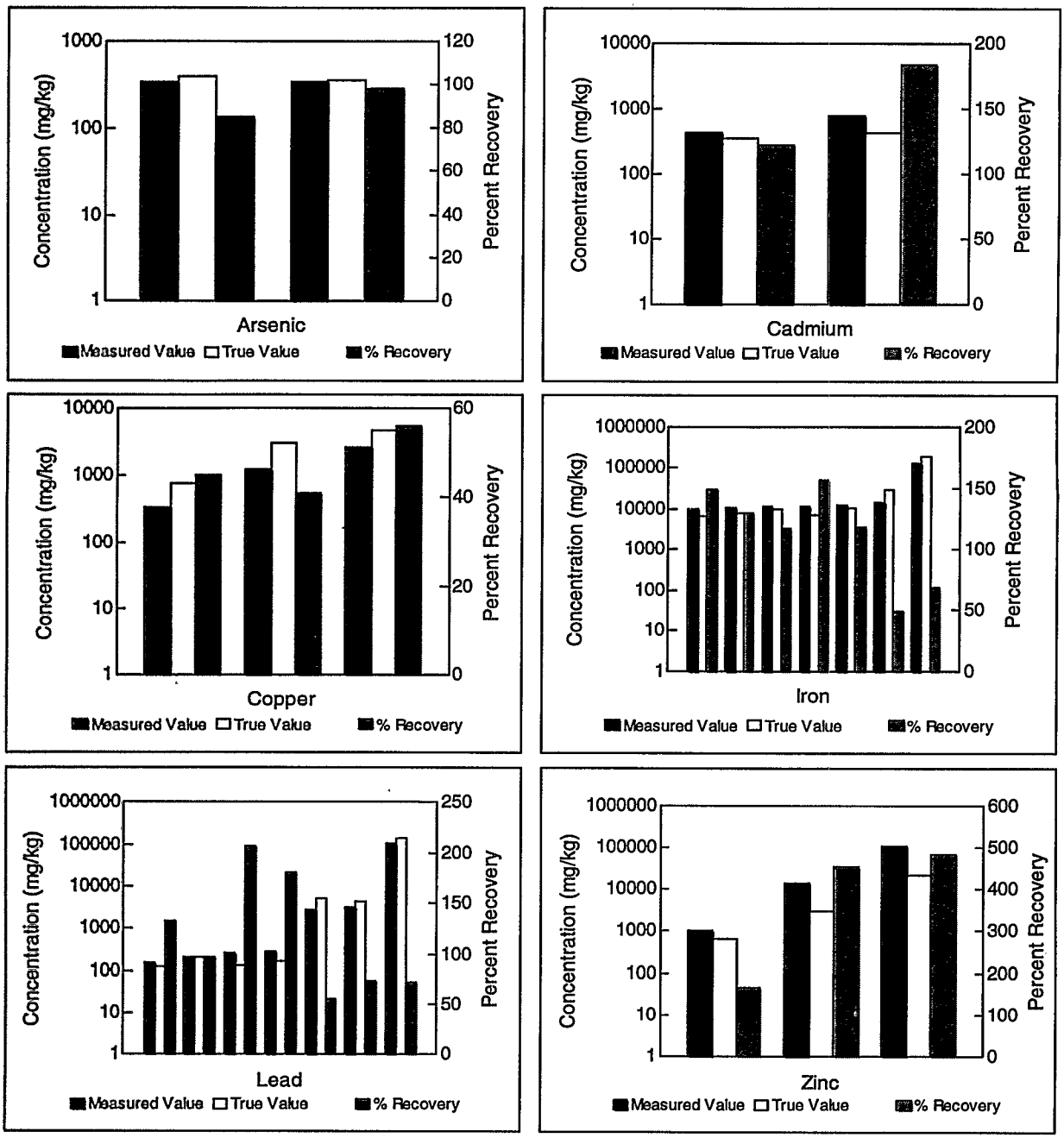


Figure 4-7. PE and CRM Results: These graphs illustrate the relationship between the analyzer's data (measured values) and the true values for the PE and CRM samples. The gray bars represent the percent recovery for the analyzer. Each set of three bars (black, white, and gray) represent a single PE or CRM sample.

Intermethod Assessment

The comparison of the SEFA-P's results to those of the reference method was performed using the statistical methods described in Section 2. The purpose of this evaluation was to determine the degree of comparability between data produced by the analyzer and that produced by the reference laboratory. If the $\log_{10}$ transformed FPXRF data were statistically equivalent to the $\log_{10}$ transformed reference data and had acceptable precision (10 percent RSD), the data met the criteria for the definitive level. If the data did not meet the definitive level criteria, but could be mathematically corrected to be equivalent to the reference data and met the other criteria described in Table 2-2, it would be classified as quantitative screening data. If the analyzer did not meet the definitive level criteria, and the statistical evaluation could not identify a predictable bias in the data, but the analyzer identified the presence or absence of contamination, the data was classified as qualitative screening level.

The SEFA-P Analyzer was configured to report concentrations for all of the target analytes. This analyzer produced two sets of data for the analysis of the demonstration samples. One data set was based on a Compton calibration, using a mid-level standard, while the second data set was based on an empirical calibration that used the reference method results for three site-specific calibration standards from each site. This intermethod data assessment will focus on the preferred Compton calibration; however, the empirical-based will be briefly discussed. The Compton calibration-based data generally exhibited better comparability relative to the empirical-based data, as expressed by the r^2 (Figure 4-8).

The regression analysis of the entire $\log_{10}$ transformed data set (Compton) showed that arsenic, copper, lead, zinc, and antimony had r^2 values at or above 0.85 (Table 4-7). Not enough nickel concentrations were reported above the instrument detection level under the Compton calibration to allow an evaluation of FPXRF nickel data. Because the empirical-based calibration was site-specific, an evaluation of the entire data set (both sites combined) produced only one r^2 value above 0.70 (Table 4-8).

The next step in the data evaluation of the Compton calibrated data involved the assessment of the potential impact of the variables: site and soil type on the regression analysis (Table 4-7). The examination of the site variables effect was restricted to lead and zinc, the two most evenly distributed analytes across both sites. Based on this evaluation, there was no apparent impact of ether the site or soil variables on the regression. Although the data in Table 4-7 for zinc indicate that the ASARCO site data had a lesser comparability relative to the RV Hopkins site data, this was most likely an artifact of zinc concentration distribution. The ASARCO data had a disproportionate amount of zinc data at and below the analyzer's precision-based MDL.

The empirical calibration-based data exhibited the same data trends as discussed above, and therefore, exhibited no site or soil texture effect on comparability.

Within the soil texture variables, the effect of contaminant concentration was also examined. The data sets for the primary analytes were sorted into the following concentration ranges: 0 - 100 mg/kg, 100 - 1,000 mg/kg, and greater than 1,000 mg/kg. The regression analysis for each target analyte for each sample preparation step was rerun on these concentration-sorted data sets. A review of these results showed general improvement in the r^2 and standard error for each target analyte with increasing concentration. The 0 - 100 mg/kg concentration range showed the poorest comparability, this range generally occurs below the analyzer's MDLs. The analyzer's precision and accuracy are lowest in this concentration range. Generally, the r^2 values improved between the 100 and 1,000 mg/kg and greater than 1,000 mg/kg ranges. This data indicated that there was a concentration effect on comparability. This effect appears to be linked to the general proximity of a measurement to its associated MDL. The further

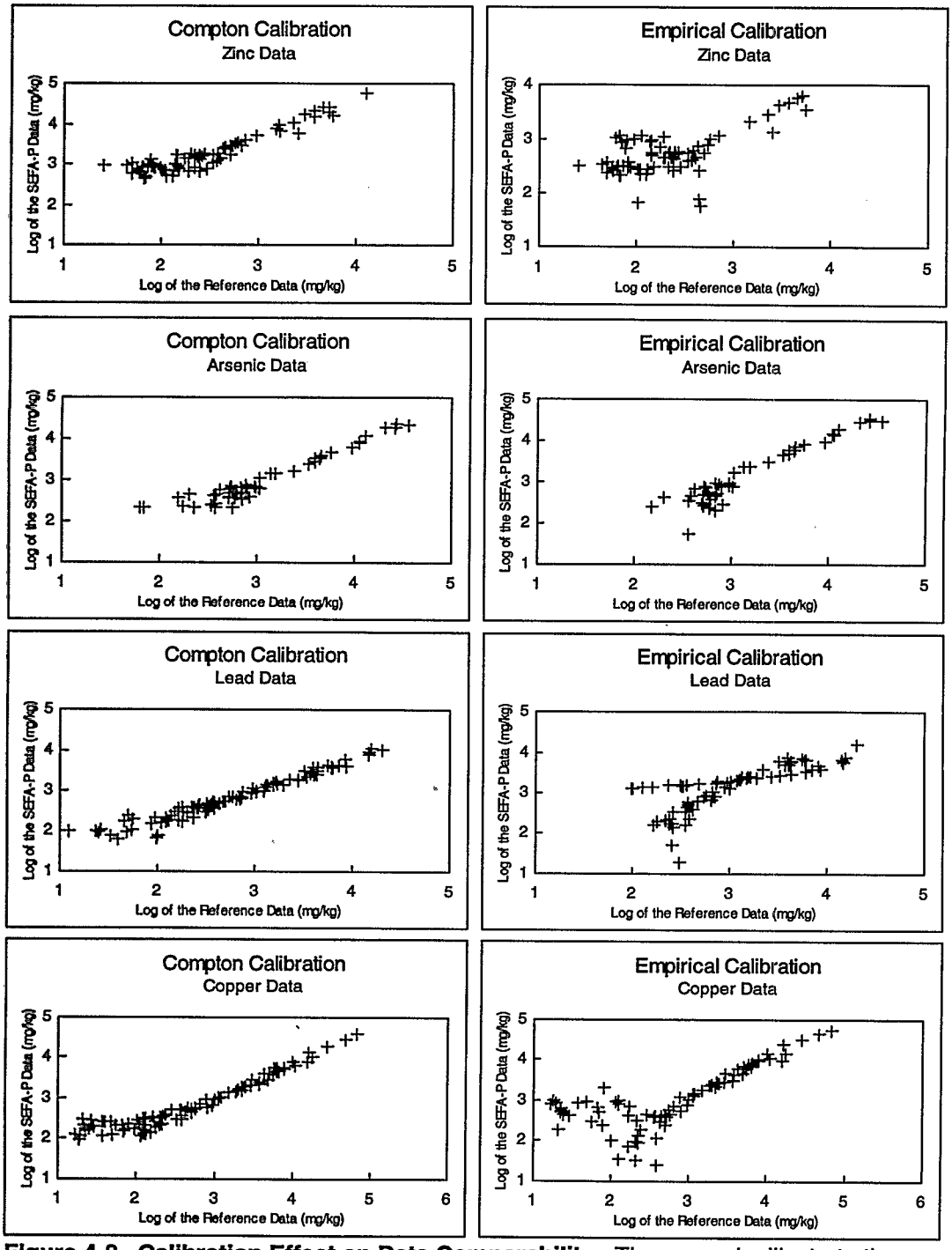


Figure 4-8. Calibration Effect on Data Comparability: These graphs illustrate the effect of calibration technique on the comparability between the SEFA-P data and the reference data. The reduced scatter at low concentrations seen in the Compton data may be an artifact of the quantitation equation uniform response to the inherent matrix background. The increased scatter at low concentrations for the empirical calibration based data may be the result of these measurements falling outside the empirical calibration range.

Table 4-7. Regression Parameters[a] by Primary Variable — Compton Calibration — SEFA-P

Antimony					Variable	Arsenic				
n	r^2	Std. Err.	Y-Int	Slope		n	r^2	Std. Err.	Y-Int.	Slope
66	0.894	0.18	1.15	0.82	All Data	49	0.946	0.14	0.14	0.93
51	0.904	0.18	1.18	0.81	ASARCO Site	49	0.946	0.14	0.14	0.93
15	0.727	0.18	1.17	0.73	RV Hopkins Site	ND	ND	ND	ND	ND
22	0.953	0.15	1.22	0.81	Sand Soil	21	0.972	0.13	-0.12	1.00
29	0.853	0.21	1.15	0.81	Loam Soil	29	0.919	0.15	0.63	0.79
15	0.727	0.18	1.17	0.73	Clay Soil	ND	ND	ND	ND	ND

Barium					Variable	Cadmium				
n	r^2	Std. Err.	Y-Int	Slope		n	r^2	Std. Err.	Y-Int.	Slope
94	0.730	0.12	2.86	0.42	All Data	19	0.636	0.24	1.62	0.37
66	0.196	0.16	3.10	0.15	ASARCO Site	16	0.667	0.25	1.54	0.41
30	0.925	0.09	2.58	0.53	RV Hopkins Site	ND	ND	ND	ND	ND
29	0.064	0.07	3.86	-0.09	Sand Soil	9	0.387	0.24	1.77	0.22
36	0.462	0.17	2.49	0.57	Loam Soil	7	0.972	0.10	1.19	0.63
30	0.925	0.09	2.58	0.53	Clay Soil	ND	ND	ND	ND	ND

Chromium					Variable	Copper				
n	r^2	Std. Err.	Y-Int	Slope		n	r^2	Std. Err.	Y-Int.	Slope
24	0.351	0.19	2.46	0.17	All Data	94	0.923	0.18	1.04	0.67
13	0.105	0.16	1.93	0.55	ASARCO Site	67	0.972	0.10	0.61	0.80
11	0.530	0.21	2.18	0.26	RV Hopkins Site	27	0.067	0.14	2.09	0.10
8	0.364	0.12	1.40	0.87	Sand Soil	30	0.953	0.12	0.59	0.81
ND	ND	ND	ND	ND	Loam Soil	37	0.977	0.09	0.67	0.79
11	0.530	0.21	2.18	0.26	Clay Soil	27	0.067	0.14	2.09	0.10

Iron					Variable	Lead				
n	r^2	Std. Err.	Y-Int	Slope		n	r^2	Std. Err.	Y-Int.	Slope
95	0.758	0.07	1.02	0.76	All Data	92	0.973	0.09	0.81	0.74
67	0.775	0.07	0.57	0.87	ASARCO Site	64	0.967	0.09	0.83	0.73
27	0.944	0.04	1.01	0.75	RV Hopkins Site	29	0.974	0.09	0.52	0.82
30	0.728	0.07	0.86	0.81	Sand Soil	27	0.976	0.08	0.97	0.69
36	0.859	0.06	-0.03	1.01	Loam Soil	36	0.975	0.08	0.72	0.76
27	0.944	0.04	1.01	0.75	Clay Soil	29	0.974	0.09	0.52	0.82

Zinc					Variable
n	r^2	Std. Err.	Y-Int.	Slope	
75	0.892	0.16	1.36	0.78	All Data
46	0.844	0.20	1.46	0.74	ASARCO Site
29	0.950	0.11	1.37	0.79	RV Hopkins Site
24	0.849	0.22	1.42	0.76	Sand Soil
22	0.843	0.17	1.54	0.69	Loam Soil
29	0.950	0.11	1.37	0.79	Clay Soil

Notes:

[a] Regression parameters based on $\log_{10}$ transformed data. Since these parameters were based on the FPXRF data being the dependent variable, they cannot be used to correct FPXRF data. See Section 5.

Y-Int. Y-Intercept.

Std. Err. Standard error.

n Number of data points.

ND Analytes not present in significant quantities to provide meaningful regression. This data does not include copper data from ASARCO samples 102 to 201 for the intrusive-prepared analyses.

Table 4-8. Regression Parameters[a] by Primary Variable — Empirical Calibration — SEFA-P

Antimony					Variable	Arsenic				
n	r^2	Std. Err.	Y-Int.	Slope		n	r^2	Std. Err.	Y-Int.	Slope
49	0.006	0.92	1.79	0.10	All Data	54	0.413	0.62	1.69	0.48
37	0.796	0.32	-0.09	1.02	ASARCO Site	43	0.906	0.22	-0.32	1.10
10	0.003	0.20	3.10	0.04	RV Hopkins Site	ND	ND	ND	ND	ND
13	0.945	0.19	0.06	0.97	Sand Soil	16	0.913	0.27	-0.54	1.15
24	0.682	0.41	-0.31	1.12	Loam Soil	26	0.932	0.16	-0.13	1.05
10	0.003	0.20	3.10	0.04	Clay Soil	ND	ND	ND	ND	ND

Barium					Variable	Cadmium				
n	r^2	Std. Err.	Y-Int.	Slope		n	r^2	Std. Err.	Y-Int.	Slope
80	0.005	0.45	2.01	0.13	All Data	42	0.015	0.85	1.85	0.15
62	0.387	0.14	1.46	0.42	ASARCO Site	32	0.410	0.50	1.08	0.56
16	0.011	0.36	2.88	-0.21	RV Hopkins Site	10	0.028	0.43	2.95	0.23
28	0.005	0.11	2.36	-0.06	Sand Soil	15	0.366	0.48	1.28	0.47
34	0.063	0.56	1.06	0.56	Loam Soil	17	0.509	0.51	0.83	0.70
16	0.011	0.36	2.88	-0.21	Clay Soil	10	0.028	0.43	2.95	0.23

Chromium					Variable	Copper				
n	r^2	Std. Err.	Y-Int.	Slope		n	r^2	Std. Err.	Y-Int.	Slope
56	0.553	0.65	0.15	1.12	All Data	81	0.587	0.47	1.34	0.61
37	0.065	0.28	2.32	-0.49	ASARCO Site	56	0.961	0.15	-0.41	1.12
17	0.242	0.37	2.82	0.27	RV Hopkins Site	22	0.001	0.25	-2.73	0.03
20	0.121	0.30	2.60	-0.74	Sand Soil	21	0.944	0.19	-0.88	1.25
17	0.052	0.23	2.19	-0.35	Loam Soil	35	0.970	0.12	-0.05	1.02
17	0.242	0.37	2.82	0.27	Clay Soil	22	0.001	0.25	2.73	0.03

Iron					Variable	Lead				
n	r^2	Std. Err.	Y-Int.	Slope		n	r^2	Std. Err.	Y-Int.	Slope
96	0.711	0.05	2.41	0.46	All Data	74	0.596	0.40	0.71	0.80
66	0.808	0.03	2.51	0.44	ASARCO Site	42	0.923	0.18	-0.29	1.12
27	0.852	0.05	1.89	0.57	RV Hopkins Site	29	0.925	0.05	2.50	0.28
30	0.761	0.03	2.64	0.41	Sand Soil	18	0.934	0.18	-0.30	1.10
36	0.847	0.03	2.40	0.46	Loam Soil	24	0.917	0.17	-0.27	1.12
27	0.852	0.05	1.89	0.57	Clay Soil	29	0.925	0.05	2.50	0.28

Nickel					Variable	Zinc				
n	r^2	Std. Err.	Y-Int.	Slope		n	r^2	Std. Err.	Y-Int.	Slope
90	0.176	0.40	0.83	0.66	All Data	68	0.279	0.40	1.66	0.45
65	0.136	0.06	1.46	0.14	ASARCO Site	46	0.840	0.17	1.26	0.62
22	0.034	0.34	2.97	-0.18	RV Hopkins Site	20	0.593	0.24	4.93	-0.98
29	0.078	0.05	1.54	0.08	Sand Soil	24	0.846	0.19	1.21	0.65
35	0.169	0.07	1.99	-0.19	Loam Soil	22	0.842	0.14	1.36	0.57
22	0.034	0.34	2.97	-0.18	Clay Soil	20	0.593	0.24	4.93	-0.98

Notes:

[a] Regression parameters based on $\log_{10}$ transformed data. Since these parameters were based on the FPXRF data being the dependent variable, they cannot be used to correct FPXRF data. See Section 5.

Y-Int. Y-Intercept.

Std. Err. Standard error.

n Number of data points.

ND Analytes not present in significant quantities to provide meaningful regression. This data does not include copper data from ASARCO samples 102 to 201 for the intrusive-prepared analyses.

away from the MDL, the less effect concentration will have on quantitation and comparability. This relationship held for both the Compton and empirical calibrations.

Another way to examine the comparability between the two methods involves measuring the average relative bias and accuracy between the FPXRF data and reference. These measurements were made using the raw FPXRF and reference data. The average relative bias indicates the average factor by which the two data sets differ. Concentration effects can affect bias. For example, it is possible for an analyzer to greatly underestimate low concentrations but greatly overestimate high concentrations and have a relative bias of zero. To eliminate this concentration effect, the data can be corrected by a regression approach (see Section 5), or only narrow concentration ranges can be analyzed, or average relative accuracy can be examined. The average relative accuracy is the average factor by which each individual analyzer measurement differs from the corresponding reference measurement.

A final decision regarding the assignment of data quality levels derived from this demonstration involves an assessment of both r^2 and one precision RSD. Using the criteria presented in Table 2-2, a summary of the HNU SEFA-P's data quality performance in this demonstration is provided in Table 4-9.

Table 4-9. Summary of Data Quality Level Parameters

Target Analytes	SEFA-P Analytes	Precision (mg/kg) Mean % RSD 5 - 10 X MDL	Method Detection Limit (mg/kg) (Precision-based)	Coefficient of Determination (r^2 All Data)	Data Quality Level
Arsenic	Arsenic	Not determined	360	0.946	Not determined
Barium	Barium	6.99	1150	0.730	Quantitative
Chromium	Chromium	Not determined	Not determined	0.351	Not determined
Copper	Copper	2.54	225	0.923	Definitive
Lead	Lead	4.86	120	0.973	Definitive
Zinc	Zinc	Not determined	990	0.892	Not determined
Nickel	Nickel	Not determined	Not determined	Not determined	Not determined
Iron	Iron	Not determined	900	0.758	Not determined
Cadmium	Cadmium	Not determined	Not determined	0.636	Not determined
Antimony	Antimony	7.20	120	0.894	Definitive

Section 5
Applications Assessment and Considerations

The SEFA-P Analyzer is designed to produce quantitative data for metals in soils, sludges, and other solids. HNU-developed software can be used for calibration and quantitation to maximize instrument performance and account for common soil-related matrix interferences. Both Compton-based and empirical calibrations were applied during this demonstration. This analyzer is designed for field use in an intrusive mode. The EPA-owned instrument did not experience a failure resulting in down time while analyzing the subset of 100 samples. The developer offers a training class in the use of the analyzer and this training, coupled with on-line technical support, was sufficient to allow uninterrupted operation and no data loss throughout the demonstration. A summary of the operational features of this instrument is presented in Table 5-1.

Comparison of SEFA-P and reference $\log_{10}$ transformed data indicated that the analyzer generally produced quantitative screening level quality data. This data quality level is applicable to most field applications. The data produced by this analyzer was $\log_{10}$-$\log_{10}$ linearly related to the reference data. The linear relationship between the analyzer and the reference methods would indicate that if 10 - 20 percent of the samples analyzed were submitted for reference method analysis, SEFA-P data could be corrected to more closely match the reference data. In the case of copper, lead, and antimony, the analyzer's data was statistically equivalent to the reference data. This analyzer also exhibited analyzer precision similar to the reference methods, indicating a high degree of reproducibility.

The SEFA-P can use up to three radioactive sources allowing analysis of a large number of metals in soils. This analyzer generally uses longer count times (greater than 60 live-seconds per source). The longer count times and multiple sources generally increase accuracy and lower the detection limits but decrease sample throughput. In a 10-hour day, 70 - 80 samples were analyzed during the demonstration.

For the analyzer, there was no apparent effect of site or soil texture on performance. The Compton calibration generally produced more comparable data relative to the reference data. The empirical calibration, however, produced more comparable results for select analytes.

Based on this demonstration, this analyzer is well suited for the rapid real-time assessment of metals contamination in soil samples. Although in several cases, the analyzer produced data statistically equivalent to the reference data, generally confirmatory analysis will be required or requested for FPXRF analysis. This holds for either the Compton or empirical calibration. If 10 - 20 percent of the samples analyzed by using either calibration are submitted for reference method analysis, instrument bias, relative to standard methods such as EPA SW-846 Methods 3050A/6010A, can be corrected. This will only hold true if the analyzer and the laboratory analyze nearly identical samples. This was accomplished in this demonstration by thorough sample homogenization. Bias correction allows analyzer data to be corrected so that it approximates EPA SW-846 Methods 3050A/6010A data. The demonstration showed that this analyzer exhibits a strong linear relationship with the reference data over a range of 5 orders of

magnitude. For optimum correlation, samples with high, medium, and low concentration ranges from a project must be submitted for reference method analysis.

Table 5-1. Summary of Test Results and Operational Features

Throughput averaged 7 to 8 samples per hour
Software handles both empirical and Compton calibrations
Three excitation sources
Requires liquid nitrogen to cool the detector
Uses an auxiliary computer for data storage and quantitation
Definitive data for lead, copper, and antimony
Quantitative screening level data for barium

During this demonstration, three calibration techniques were evaluated for the SEFA-P: empirical with synthetic standards, empirical with site-specific standards, and Compton calibration. The empirical calibration with synthetic standards provided a number of negative values for analytes. The operator believed that this may have been related to the use of a synthetic matrix. Both the empirical site-specific and Compton calibrations were used to evaluate demonstration data. The comparability of the Compton calibration was better than the empirical site-specific calibration.

This analyzer can provide rapid assessment of the distribution of metals contamination in soil at a hazardous waste site. This data can be used to characterize general site contamination, guide critical conventional sampling and analysis, and monitor removal actions. This demonstration suggested that in some applications and for some analytes, the data may be statistically similar to the reference data. The approval of SW-846 Method 6200 "Field Portable X-Ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment" will help in the acceptance of this data. The analyzer data can be produced and interpreted in the field on a daily or per sample basis. This real-time analysis allows the use of contingency-based sampling for any application and greatly increases the potential for meeting project objectives on a single mobilization. This analyzer is an effective tool for site characterization and remediation. It provides a faster and less expensive means of analyzing metals contamination in soil.

FPXRF data can be corrected using the regression approach presented below; usually this procedure results in an improvement of both the average relative bias and accuracy. The average relative bias numbers will no longer be strongly influenced by a concentration effect since the regression approach used to correct the data used $\log_{10}$ transformed data. The average relative bias and accuracy for the corrected data are similar to the acceptable average relative bias between the reference data and PE samples (true values), as shown by the last column in Table 5-2.

The steps to correct the FPXRF measurements to more closely match the reference data are as follows:

1. Conduct sampling and FPXRF analysis.
2. Select 10 - 20 percent of the sampling locations for resampling. These resampling locations can be evenly distributed over the range of concentrations measured or they can focus on an action level concentration range.
3. Resample the selected locations. Thoroughly homogenize the samples and have each sample analyzed by FPXRF and a reference method.

Table 5-2. Effects of Data Correction on FPXRF Comparability to Reference Data for All In Situ-Prepared Samples

Target Analyte	Average Relative Bias on Raw Data[a]	Average Relative Bias on Corrected Data[b]	Average Relative Accuracy on Raw Data[c]	Average Relative Accuracy on Corrected Data[d]	Acceptable Relative Accuracy Based on PE Samples[e]
Antimony	8.36	1.10	9.35	1.51	2.94
Arsenic	8.89	1.06	1.39	1.44	1.76
Barium	47.99	1.17	53.94	1.63	1.36
Chromium	9.68	1.46	12.45	1.40	1.55
Copper	1.85	1.16	3.06	1.75	1.18
Iron	1.02	1.02	1.19	1.19	1.54
Lead	1.33	1.04	1.85	1.33	1.63
Zinc	7.10	1.03	9.66	1.68	1.64

Notes:

[a] A measurement of average relative bias, measured as a factor by which the FPXRF, on average, over- or underestimates results relative to the reference methods. This measurement of bias is based on raw (not $\log_{10}$ transformed) data. This average relative bias does not account for any concentration effect on analyzer performance.

[b] A measurement of average relative bias on the FPXRF data after it has been corrected using the eight-step regression approach.

[c] A measurement of average relative accuracy at the 95 percent confidence interval, measured as a factor by which the raw FPXRF, on average, over- or underestimates individual results relative to the reference methods. This measurement of accuracy is based on raw (not $\log_{10}$ transformed) data. This average relative accuracy is independent of concentration effects.

[d] A measurement of average relative accuracy at the 95 percent confidence interval, of the corrected FPXRF data obtained using the eight-step regression approach.

[e] A measurement of accuracy represents a factor and 95 percent confidence interval that define the acceptable range of differences allowed between the reference method reported concentrations and the true value concentrations in the PE samples. This bias is included only as a general reference for assessing the improvement on comparability of FPXRF data and reference data after FPXRF data correction.

The average relative bias is calculated as follows:

$$\text{Average relative bias} = ((\sum_i[\text{FPXRF}_i/\text{Reference}_i])/\text{number of paired samples})-1$$

This value represents the percentage that the FPXRF over- or underestimates the reference data, on average, for the entire data set. To convert this calculated value to a factor, 1.0 is added to the calculated average relative bias. The above table presents the average relative bias as a factor.

The average relative accuracy is calculated as follows:

$$\text{Average relative accuracy} = \text{SQRT}\ (\sum_i([\text{FPXRF}_i/\text{Reference}_i]-1)^2/\text{number of paired sample})$$

This value represents the percentage that an individual FPXRF measurement over- or underestimates the reference data. The relative accuracy numbers in the table are calculated at the 95 percent confidence interval. This is accomplished by adding two standard deviations to the above formula before the square root is taken. To convert this calculated value to a factor, 1.0 is added to the calculated average relative accuracy. The above table presents the average relative bias as a factor.

4. Tabulate the resulting data with reference data in the x-axis column (dependent variable) and the FPXRF data in the y-axis column (independent variable). Transform this data to the equivalent $\log_{10}$ value for each concentration.

5. Conduct a linear regression analysis and determine the r^2, y-intercept and slope of the relationship. The r^2 should be greater than 0.70 to proceed.
6. Place the regression parameters into Equation 5-1:

$$Y(\log_{10}\ corrected\ FPXRF\ data) = slope * (\log_{10}\ FPXRF\ data) + Y\text{-}intercept \quad \textbf{(5-1)}$$

7. Use the above equation with the $\log_{10}$ transformed FPXRF results from Step 4 above and calculate the equivalent $\log_{10}$ corrected FPXRF data.
8. Take the anti-$\log_{10}$ (10 [$\log_{10}$ transformed corrected FPXRF data]) of the equivalent $\log_{10}$ corrected FPXRF data calculated in Step 7. These resulting values (in milligrams per kilogram) represent the corrected FPXRF data.

To show the effect of correcting the FPXRF data, the change in average relative bias and accuracy can be examined. The average relative bias between the FPXRF data and the reference data is a measure of the degree to which the FPXRF over- or underestimates concentrations relative to the reference methods. The relative bias is an average number for the entire data set and may not be representative of an individual measurement. An example of this can be seen in an analyzer's data where measurements are underestimated at low concentrations but overestimated at high concentrations. On average, the relative bias for this analyzer is zero; however, this bias is not representative of high or low concentration measurements. To avoid this dilemma, three approaches can be taken: (1) the evaluation of average relative bias can be focused on a narrow concentration range, (2) the analyzer's data can be corrected using the regression approach described above, or (3) the average relative accuracy can be calculated. Average relative accuracy represents the percentage that an individual measurement is over- or underestimated relative to a reference measurement. Table 5-2 shows the average relative bias and accuracy exhibited by the analyzer, for the *in situ*-prepared data set, before and after data correction using the eight-step approach previously discussed.

The average relative bias and accuracy for the analytes which fall into the definitive level data quality category were generally small. The analytes falling into the quantitative screening level data quality categories had generally larger average relative bias and accuracy and often showed a greater change when corrected by this procedure.

General Operational Guidance

The following paragraphs describe general operating considerations for FPXRF analysis. This information is derived from SW-846 Method 6200 for FPXRF analysis.

General operation of FPXRF instruments will vary according to specific developer protocols. For all environmental applications, confirmatory or reference sampling should be conducted so that FPXRF data can be corrected. Before operating any FPXRF instrument, the developer's manual should be consulted. Most developers recommend that their instruments be allowed to warm up for 15 - 30 minutes before analysis of samples. This will help alleviate drift or energy calibration problems.

Each FPXRF instrument should be operated according to the developer's recommendations. There are two modes in which FPXRF instruments can be operated: *in situ* and intrusive. The *in situ* mode involves analysis of an undisturbed soil or sediment sample. Intrusive analysis involves collecting and preparing a soil or sediment sample before analysis. Some FPXRF instruments can operate in both modes of analysis, while others are designed to operate in only one mode. The two modes of analysis are discussed below.

For *in situ* analysis, one requirement is that any large or nonrepresentative debris be removed from the soil surface before analysis. This debris includes rocks, pebbles, leaves, vegetation, roots, and concrete. Another requirement is that the soil surface be as smooth as possible so that the probe window will have good contact with the surface. This may require some leveling of the surface with a stainless-steel trowel. Most developers recommend that the soil be tamped down to increase soil density and compactness. This step reduces the influence of soil density variability on the results. During the demonstration, this modest amount of sample preparation was found to take less than 5 minutes per sample location. The last requirement is that the soil or sediment not be saturated with water. Developers state that their FPXRF instruments will perform adequately for soils with moisture contents of 5 - 20 percent, but will not perform well for saturated soils, especially if ponded water exists on the surface. Data from this demonstration did not see an effect on data quality from soil moisture content. Source count times for *in situ* analysis usually range from 30 to 120 seconds, but source count times will vary between instruments depending on required detection limits.

For intrusive analysis of surface soil or sediment, it is recommended that a sample be collected from a 4- by 4-inch square that is 1 inch deep. This will produce a soil sample of approximately 375 grams or 250 cm^3, which is enough soil to fill an 8-ounce jar. The sample should be homogenized, dried, and ground before analysis. The data from this demonstration indicated that sample preparation, beyond homogenization, does not greatly improve data quality. Sample homogenization can be conducted by kneading a soil sample in a plastic bag. One way to monitor homogenization when the sample is kneaded in a plastic bag is to add sodium fluorescein dye to the sample. After the moist sample has been homogenized, it is examined under an ultraviolet light to assess the distribution of sodium fluorescein throughout the sample. If the fluorescent dye is evenly distributed, homogenization is considered complete; if the dye is not evenly distributed, mixing should continue until the sample has been thoroughly homogenized. During the demonstration, the homogenization procedure using the fluorescein dye required 3 to 5 minutes per sample.

Once the soil or sediment sample has been homogenized, it can be dried. This can be accomplished with a toaster oven or convection oven. A small portion of the sample (20 - 50 grams) is placed in a suitable container for drying. The sample should be dried for 2 to 4 hours in the convection or toaster oven at a temperature not greater than 150 °C. Microwave drying is not recommended. Field studies have shown that microwave drying can increase variability between the FPXRF data and reference data. High levels of metals in a sample can cause arcing in the microwave oven, and sometimes slag will form in the sample.

The homogenized, dried sample material can also be ground with a mortar and pestle and passed through a 60-mesh sieve to achieve a uniform particle size. Sample grinding should continue until at least 90 percent of the original sample passes through the sieve. The grinding step normally averages 10 minutes per sample.

After a sample is prepared, a portion should be placed in a 31-mm polyethylene sample cup (or equivalent) for analysis. The sample cup should be completely filled and covered with a 2.5-micrometer Mylar™ (or equivalent) film for analysis. The rest of the soil sample should be placed in a jar, labeled, and archived. All equipment, including the mortar, pestle, and sieves, must be thoroughly cleaned so that the method blanks are below the MDLs of the procedure.

Section 6
References

Havlick, Larry L., and Ronald D. Crain. 1988. *Practical Statistics for the Physical Sciences.* American Chemical Society. Washington, D.C.

Kane, J. S., S. A. Wilson, J. Lipinski, and L. Butler. 1993. "Leaching Procedures: A Brief Review of Their Varied Uses and Their Application to Selected Standard Reference Materials." *American Environmental Laboratory*. June. Pages 14-15.

Kleinbaum, D. G., and L. L. Kupper. 1978. "Applied Regression Analysis and Other Multivariable Methods." Wadsworth Publishing Company, Inc., Belmont, California.

Morgan, Lewis, & Bockius. 1993. RODScan®.

PRC Environmental Management, Inc. 1995. "Final Demonstration Plan for Field Portable X-ray Fluorescence Analyzers."

U.S. Environmental Protection Agency. 1993. "Data Quality Objectives Process for Superfund-Interim Final Guidance." Office of Solid Waste and Emergency Response. Washington, D.C. EPA/540/R-93/071.