United States Environmental Protection Agency

Office of Water Regulations and Standards (WH-552) Washington, D.C. 20460

Office of Solid Waste and Emergency Response Washington, D.C. 20460

EPA 530-SW-87-005 B

EPA

TECHNICAL REPORT: APPENDIX B

EXPLORATION, DEVELOPMENT, AND PRODUCTION OF CRUDE OIL AND NATURAL GAS

SAMPLING STRATEGY

U.S. Environmental Protection Agency
Region 5, Library (PL-12J)
77 West Jackson Boulevard, 12th Floor
Chicago, IL 60604-3590

OIL AND GAS EXPLORATION, DEVELOPMENT, AND PRODUCTION SAMPLING STRATEGY - DRAFT

JANUARY 31, 1987

PREPARED FOR
U. S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF WATER - INDUSTRIAL TECHNOLOGY DIVISION
401 M STREET, SW (WH-552)
WASHINGTON, D. C. 20460

PREPARED BY
CENTEC CORPORATION
11260 ROGER BACON DRIVE
RESTON, VIRGINIA 22090

U.S. Environmental Protection Agency
Region 5, Library (PL-12J)
77 West Jackson Boulevard, 12th Floor
Chicago, IL 60604-3590

LIST OF ABBREVIATIONS

DOE	--	Department of Energy
EPA	--	Environmental Protection Agency
ITD	--	Industrial Technology Division
OSW	--	Office of Solid Waste
OWRS	--	Office of Water Regulations and Standards
QA	--	Quality Assurance
QC	--	Quality Control
RCRA	--	Resource Conservation and Recovery Act

TABLE OF CONTENTS

INTRODUCTION 1

BACKGROUND 3

SAMPLING DESIGN 7
- Objective 7
- Target Populations 8
- Selection of Analytes 16
- Sample Size 17
- Site Selection Process 20
- Implementation of the Sampling Design 27

STATISTICAL ANALYSIS 31
- Sampling Design 31
- Precision and Accuracy 33

CONCLUSION 35

REFERENCES 37

LIST OF FIGURES AND TABLES

Figure 1	Site Selection Zones	10
Figure 2	Example of Process Used for Selection of States Within Zones	24
Table 1	1984 Oil and Gas Production and Drilling Activity	12
Table 2	Summary of Analytes	18
Table 3	Distribution of Types of Sample Sites . . .	19
Table 4	Percentage (by Zone) of 1984 Oil and Gas Production and Drilling Activity	23
Table 5	Planned Distribution of Probability Sites And Specifically-Selected Sites Among Zones	28
Table 6	Distribution of Probability Sites and Specifically-Selected Sites Among Zones . .	29

INTRODUCTION

This document is a revision of the May, 1986, EPA Oil and Gas Exploration, Development, and Production, Sampling Strategy - Draft (EPA, 1986). It reflects actual implementation of the sampling strategy and incorporates comments regarding the May 1986 draft document. The EPA strategy for sampling wastes from oil and gas exploration, development, and production is presented herein. This strategy is a part of the Office of Solid Waste (OSW)/ Office of Water Regulations and Standards (OWRS) Exploration, Development, and Production of Crude Oil and Natural Gas Field Sampling and Analysis Project, which in turn is part of the OSW Oil, Gas, and Geothermal Exempt Waste Study. This document presents the sampling project objectives, the sampling design, the development of a concept of geographical "zones" with common characteristics, and the details of selection of sample sites.

EPA sampled wastes from the oil and gas extraction industry (see the "Regulatory Background" section of the "Technical Report on the Exploration, Development, and Production of Crude Oil and Natural Gas" (EPA-530-SW-87-005)). In the process of oil and gas extraction, liquid and solid wastes are generated. These wastes are disposed in pits, injected underground, and/or discharged to the environment in other ways (discharge water, for example). It is estimated that there are 885,077 oil or gas wells producing in the United States (IPAA, 1986). Prior to 1986, approximately

70,000 wells were drilled annually. For 1986, an additional 40,000 drill sites were created (Oil and Gas Journal, 1986). Each of these sources generates waste.

The first priority of the sampling project was to develop data nationwide for the range of constituents in wastes from drilling and production operations. This was achieved in the design of the project by creating clusters of states, termed "zones," that reflect geological formation, geographical distribution, drilling and production activity, climate, and drilling methods. Zone characteristics are described in detail in the "Sampling Design" section of this document.

The sampling design also included identification of a target population, selection of analytes, sample size, and the process used for site selection. For example, after development of the concept of zones, EPA addressed the characteristics of the sites to be sampled in each zone. To select sites within a zone, EPA chose a random approach. This approach used randomization techniques to select 69.4 percent of the samples and a directed approach to select the remaining sites. The directed approach assured that sites with certain characteristics would be sampled. The "Statistical Analysis" section of this document gives the technical details that support the site selection.

BACKGROUND

EPA regulates or makes determinations as to whether to regulate the oil and gas extraction industry under several major environmental statutes. These statutes include the Clean Water Act (i.e., appropriate effluent limitations guidelines), the Safe Drinking Water Act (i.e., the Underground Injection Control program) and the Resource Conservation and Recovery Act (i.e., the regulatory determination under RCRA Section 3001(b)(2)(B), which uses information from the study under Section 8002(m)) (see the "Regulatory Background" section of the "Technical Report on the Exploration, Development, and Production of Crude Oil and Natural Gas" (EPA-530-SW-87-005)). Therefore, this project was designed as a screening sampling project to obtain information fulfilling some requirements mandated by RCRA and the Clean Water Act.

EPA conducted a comprehensive literature search (of both published and unpublished information) regarding the oil and gas industry. No nationwide waste characterization projects had been conducted to test for organic analyses and for RCRA characteristics (i.e., ignitability, reactivity, corrosivity, and extraction-procedure toxicity). Therefore, the Agency decided to undertake a sampling project to develop information for the study required under RCRA Section 8002(m), which would be used in making the regulatory determination under RCRA Section 3001(b)(2)(B). The sampling project was also intended to gather

more information for the Agency's review of the existing effluent limitations guidelines for the oil and gas industry and, if appropriate, to develop future effluent limitations guidelines.

The purpose of this document is to describe the design and rationale of the field sampling project. It includes the criteria for site selection for the Environmental Protection Agency (EPA) oil and gas sampling project. The objectives are also described.

The sampling project produced site specific data on the sources, volumes, and constituents of wastes from exploration, development, and production of crude oil or natural gas. The results are included in "Technical Report on the Exploration, Development, and Production of Crude Oil and Natural Gas" (EPA-530-SW-87-005). The sampling project gathered information about the wastes from this industry and the control and disposal practices being utilized by the industry.

The project did not include sampling of wastes from geothermal, tar sands, or sampling of surface water, groundwater, or air. For this sampling project, the Agency decided to concentrate its available resources on characterizing drilling wastes and produced water.

Potential sample sites were selected by working with state and regional authorities to identify candidate sites. These sites were then used in the randomized site selection process.

Oil and gas wastes were sampled, preserved, and shipped according to the EPA Sampling Plan and Sampling Quality Assurance/Quality Control Plan (Appendix G of the EPA Technical Report, (EPA-530-SW-87-005)). This QA/QC plan was specifically designed for this sampling project. Analytes are listed in the List of Analytes (Appendix F of the EPA Technical Report (EPA-530-SW-87-005)).

In addition to the sampling and analysis, reports describing operations at each site were prepared. These reports contain detailed site-specific data, including descriptions, sampling information, engineering data, waste volumes, disposal practices utilized, and cost information (Appendix C of the EPA Technical Report (EPA-530-SW-87-005)).

SAMPLING DESIGN

OBJECTIVE

The objectives of the sampling project are:

- To provide data to be included in the report to Congress on wastes associated with oil and gas exploration, development, and production activities as required by RCRA
- To provide nationwide data on sources and volumes of oil and gas wastes
- To provide information on the complexity and diversity of the wastes generated by the industry, current disposal practices and ultimate treatment
- To identify characteristics and constituents of the wastes streams and to estimate the variability of these waste streams
- To provide data that can be used in the design of a larger, more comprehensive sample survey of the industry

The first priority of the sampling project was to develop data nationwide for the range of constituents in wastes from drilling and production operations. This design priority addressed four of the objectives listed above. It was achieved through execution of the sampling design described in this section. The sampling design included identification of a target population, selection of analytes, sample size, and the process used for site selection.

Another goal of this project was to increase EPA's knowledge of industrial processes such as drilling and production by gathering technical information at each of the sample sites. This objective was achieved by collecting relevant state and federal

permits issued for sample sites, collecting geological maps of the areas, and by preparing reports describing the operations at each sample site.

Another goal was to gain information about centralized treatment facilities and centralized disposal pits handling oil and gas wastes. The sampling design limited sampling to those facilities designed solely for treatment of oil and gas wastes.

TARGET POPULATIONS

Drill Sites and Production Sites

The sampling design was based on identification of the target populations for drill sites and production sites. In the oil and gas industry, total annual drilling activity was normally about 70,000 new drill sites per year, but slumped to about 40,000 in 1986 (Oil and Gas Journal, 1986). Total annual U.S. production was estimated at 2.8 billion barrels of oil from 646,735 oil wells, and 12,586,200 MMCF of gas from 238,342 gas wells in 1985 (IPAA, 1986). Many difficulties were apparent in attempting to sample a large and heterogeneous population with the limited sample size available in this program.

For these reasons drill site and production site populations were divided into eleven zones, which were composed of groups of individual states (refer to Site Selection Process for more information regarding the importance of this decision).

Development of zones. Clusters of states were defined according to geological formation, geographic diversity, drilling and production activity, climate, topography, drilling methods used, and groundwater aquifers in certain areas. The U.S. was subdivided into 11 zones, each representing a geologic and geographic region of influence (Figure 1). State borders were used as zone boundaries to conform with state record keeping and regulations. The design of the sampling project dictated that only the most current drilling and production records be utilized. The most current records are kept largely by state agencies. State agencies generally have site information available on a day-to-day basis.

Geological distribution was the primary criterion for development of zones. Geological formations have characteristics which require adaptation of drilling techniques and/or drilling media. Thus, geologic distribution directly affects drilling waste composition. For instance, in tight hard-rock formations air drilling (using air, other gas, mist, foam, or aerated mud) can be used; in porous sandy formations rotary drilling (using water-base or oil-base muds) may be necessary. Combinations of drilling methods/media may be used as different geologic zones are encountered.

The second criterion for development of zones was geography. This sampling project was designed for nation-wide coverage of the oil and gas extraction industry.

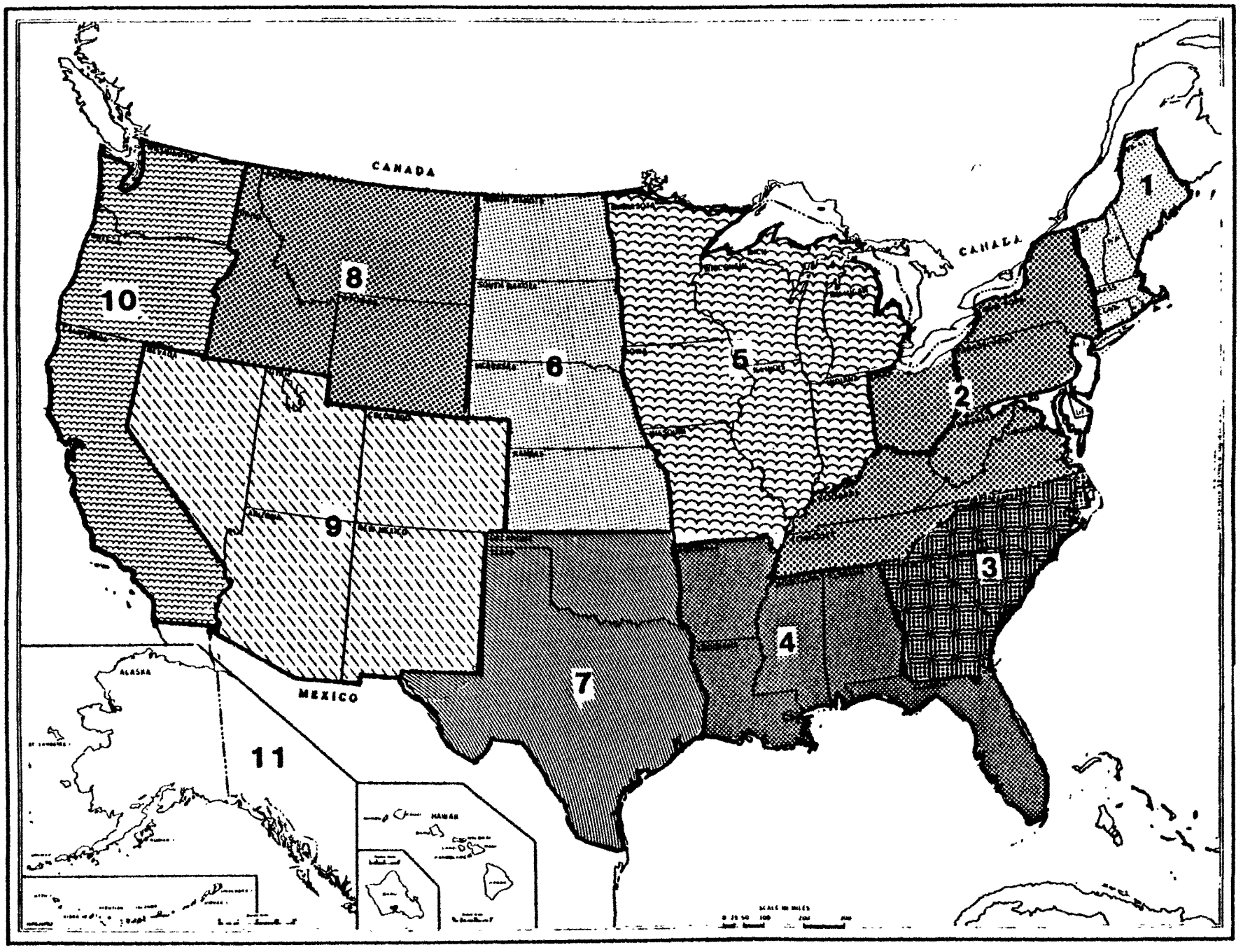

Figure 1. Site Selection Zones

Finally, consideration was given to climate, topography, drilling methods used, and groundwater aquifers in certain areas.

Description of zones. Two test areas, Zone 1 and Zone 3 on Figure 1, have essentially no drilling or production. Other states with little or no production are necessarily included in zones that were linked by similar characteristics.

A summary of drilling and production figures for 1984 is shown in Table 1. A description of the distinctive features of each zone is presented below.

It should be noted that Ohio was placed in Zone 5 when the zones were created and for site selection in this project. However, Ohio will be considered a member of Zone 2 in the future. This change is due to comments received by EPA at a meeting in July 1986 with industry representatives and state officials. This change occurred after site selection was complete and sampling was underway. Thus, Ohio is grouped with Zone 5 states in following discussions of sampling activities.

Zone 1. On Figure 1, Zone 1 is a non-producing zone. No sampling was planned for the states in Zone 1.

Zone 2. Zone 2 includes New York, Pennsylvania, West Virginia, Virginia, Kentucky, and Tennessee. Zone 2 encompasses the northern part of the Appalachian Basin, where mountainous terrain, hard-rock geology, and net precipitation combine to create a difficult environment for waste disposal. Air drilling is a predominant exploratory and developmental practice. Produced water injection is practiced primarily in large fields.

Table 1. 1984 Oil and Gas Production and Drilling Activity

	Total Annual Oil And Gas Liquids Production[1] (mbbl)	Natural Gas Production[2] (mbbl/COE)	Total Production (mbbl)	Total Wells[4] Drilled
Zone 2				
Kentucky[3]	11,635	8.064	11,643	3,546
New York	952	4.660	957	517
Pennsylvania[3]	4,491	20.436	4,511	3,261
Tennessee	968	0.863	969	307
Virginia	0	1.432	1	66
West Virginia	3,303	24.808	3,327	1,821
TOTAL	16,858	60.263	21,408	9,518
Zone 4				
Alabama	19,878	23.646	19,902	365
Arkansas[3]	21,519	23.284	21,542	866
Florida[3]	23,401	4.101	23,405	20
Louisiana	187,371	357.267	187,728	5,775
Mississippi[3]	31,613	34.848	31,648	606
TOTAL	283,782	443.146	284,225	7,632
Zone 5				
Illinois[3]	28,873	0.178	28,873	2,577
Indiana[3]	5,320	0.023	5,320	918
Ohio	14,971	26.080	14,997	4,677
Michigan[3]	30,888	24.899	30,913	1,077
TOTAL	80,052	51.180	80,103	9,249
Zone 6				
Kansas	75,723	81.111	75,804	6,169
Nebraska	6,470	0.397	6,470	361
North Dakota	52,654	13.566	52,668	692
South Dakota	1,343	0.431	1,343	43
TOTAL	136,190	95.505	136,285	7,265
Zone 7				
Oklahoma	168,604	360.912	168,965	9,241
Texas	1,165,757	1,012.337	1,166,769	26,103
TOTAL	1,334,361	1,373.249	1,335,734	35,344
Zone 8				
Montana[3]	30,063	9.774	30,073	804
Wyoming	127,116	103.524	127,220	1,680
TOTAL	157,179	113.298	157,293	2,484
Zone 9				
Colorado	29,102	36.988	29,139	1,778
New Mexico	79,335	163.280	79,498	1,609
Utah	38,054	31.597	38,086	486
TOTAL	146,491	231.875	146,723	3,873
Zone 10				
California	412,400	76.220	412,476	3,261
Zone 11				
Alaska	631,079	52.216	631,131	155

[1]IOCC, 1985

[2]DOE, 1985

[3]IOCC, 1984

[4]API, 1986

States in Zone 2 are permitting publicly-owned treatment works (POTWs) and commercial facilities to treat wastes associated with oil and gas production. As of June 1986, four or five commercial centralized treatment facilities were dedicated to oil and gas waste treatment in Zone 2. There are approximately six additional centralized treatment facilities in the planning stage.

Total zone production in 1984 was 21,408 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 9,518 wells drilled (API, 1986).

Zone 3. On Figure 1, Zone 3 is considered a non-producing zone. No sampling was planned for the states in Zone 3.

Zone 4. Zone 4 is formed by Arkansas, Louisiana, Mississippi, and Alabama. It includes the southern tip of the Appalachian Basin, the Black Warrior Basin, the Mississippi Salt Dome Basin, the eastern portion of the Gulf Coast Salt Dome Basin, and several smaller basins and uplifts. It is a net precipitation area, with major oil and gas production from both onshore and coastal regions. Drilling is predominately conducted with mud systems. Mud recycling and mud injection are practiced. Muds from coastal and offshore oil and gas operations which are unacceptable for offshore disposal are brought to Zone 4 for treatment, recovery, or disposal. Zone 4 also includes natural gas production from coal deposits found in the Black Warrior Basin. Injection is a common practice for enhanced oil recovery, for produced water disposal, and for drilling mud disposal. Construction, reclamation, and permanent disposal of drilling muds are handled differently from state to state in Zone 4.

Total zone production in 1984 was 284,225 thousand barrels (including natural gas as crude oil equivalent). Total 1984 drilling activity was 7,632 wells drilled.

Zone 5. Zone 5 includes Michigan, Indiana, Illinois, Wisconsin, Minnesota, Iowa, and Missouri. Three major basins in this zone are the Illinois Basin, Michigan Basin, and the Appalachian Basin. Mud drilling and air drilling are both practiced. One unique situation in Michigan is that there are areas where drilling must pass through substantial salt formations to reach oil and gas formations. Wastes from these drilling operations are typically salt-saturated, and may contain organics and several tons of salt.

Total zone production in 1984 was 80,052 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 9,239 wells drilled (API, 1986).

Zone 6. Zone 6 includes North Dakota, South Dakota, Nebraska, and Kansas. In the Dakotas, the Williston Basin is the major producing formation. Nebraska production is from the

Denver Basin. Kansas produces from the Forest City Basin in the east, and several smaller reservoirs in the west. Water-base mud drilling predominates. Drilling pit contents are land-spread after the supernatant has evaporated. Most reserve pits must be reclaimed after land-spreading. Much of the production in Zone 6 is from stripper wells. Hydrogen sulfide and produced water salinity is extremely high in the Dakotas, discouraging wildlife and agricultural use of the wastes. Salinity and hydrogen sulfide are low in Nebraska and Kansas.

Total zone production in 1984 was 136,285 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1986). Total 1984 drilling activity was 7,265 wells drilled (API, 1986).

Zone 7. Texas and Oklahoma combine to form Zone 7. Although water-base drilling operations predominate in Texas and Oklahoma, this zones uses oil-base muds as well. Production is from a dozen major basins, dominated by the Gulf Coast Salt Dome Basin, Permian Basin, and Anadarko Basin.

In Texas, reserve pit wastes may be land-applied if the chloride content is less than 500 mg/L (inland). Oklahoma allows disposal of drilling wastes by annular injection, evaporation of reserve pits, or in centralized offsite pits. Injection of produced water is practiced in most areas of Texas, except in coastal areas where discharges are allowed. Produced water disposal pits are banned in Texas. In Oklahoma, centralized offsite pits are commonly used for disposal of produced water and drilling muds. Oklahoma is grouped with Texas because they share climate, geology, and terrain.

Total zone production in 1984 was 1,335,734 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 35,344 wells drilled (API, 1986).

Zone 8. Zone 8 contains Montana, Wyoming, and Idaho. Montana and Wyoming share the Big Horn Basin and Powder River Basin. Idaho, a non-producing state, is also included. Lined and unlined reserve pits are permitted by regulatory authorities in Zone 8. Drilling waste pit reclamation may include treatment and land application of the supernatant. This land application usually takes the form of trenching in a radius around the pit, or land-spreading (sometimes by spraying). Produced water is injected or discharged under the provisions of the Agricultural and Wildlife Water Use Subcategory of the onshore segment of the oil and gas extraction industry effluent limitations guidelines (EPA, 1979).

Total zone production in 1984 was 157,293 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 2,484 wells drilled (API, 1986).

Zone 9. Zone 9, commonly known as the "four corners" area, includes Colorado, Utah, Nevada, Arizona, and New Mexico. Production is primarily from the San Juan Basin, Paradox Basin, and Denver Basin. Drilling wastes are disposed onsite by evaporation and tilling into the soil. Colorado and New Mexico also permit centralized offsite pits for disposal of drilling wastes and produced water. Produced water is almost entirely injected. Surface methods are used in Utah for produced water disposal. The Agricultural and Wildlife Water Use subcategory of the onshore segment of the oil and gas extraction industry effluent limitations guidelines is also applicable to this zone (EPA, 1979).

Total zone production in 1984 was 146,723 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 3,873 wells drilled (API, 1986).

Zone 10. Zone 10 includes California, Oregon, and Washington. California's San Joaquin Basin, Santa Maria Basin, and Ventura Basin are the major producing zones for oil and gas. Secondary and tertiary recovery production technologies are common in Zone 10. Oregon is a newcomer to oil and gas production, producing from several basins. Washington is a non-producing state. Drilling wastes may be evaporated or land-farmed except in California. Produced water is injected for enhanced recovery, or may be disposed in evaporation/percolation ponds. Use of produced water for irrigation is also allowed. All three states have some geothermal exploratory activity, and some active utilization of geothermal energy.

Total zone production in 1984 was 412,476 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 3,261 wells drilled (API, 1986).

Zone 11. Alaska and Hawaii form Zone 11. Hawaii is a non-producing state. In Alaska, South Central and North Slope reserves produce oil and gas from wetlands. The South Central area includes the Kenai Peninsula and Cook Inlet. Cook Inlet drilling wastes and produced water are disposed offshore. Kenai reserve pits are generally unlined. On the North Slope, drillers are required to hold reserve pit wastes for one freeze/thaw cycle, but are then permitted to discharge the supernatant to the tundra. Annular injection is sometimes used for drilling wastes and stimulation fluids. Lined reserve pits are required on federal land.

Total zone production in 1984 was 631,131 thousand barrels (including natural gas as crude oil equivalent) (IOCC, 1985). Total 1984 drilling activity was 155 wells drilled (API, 1986).

Thus, target populations for drill sites and production sites were compartmentalized into nine zones for consideration in the selection of sample sites (refer to Site Selection Process).

Centralized Pits and Centralized Treatment Facilities. EPA decided to sample specifically identified sites to develop preliminary information regarding these types of sites. The Agency is interested in gathering information about the types of processes, characteristics of commingled wastes, and the characteristics of treated commingled wastes.

SELECTION OF ANALYTES

The analytes selected for testing in this program were those on various regulatory lists or those specific to characterization of wastes from this industry. The regulatory lists from which the pollutants were taken were the:

Priority Pollutant List [NRDC vs Train, 8 ERC 2120 (DDC1976)]

Priority Pollutant Appendix C List [ibid.]

RCRA Appendix VIII List [50 FR 1999]

Michigan List [49 FR 49793]

Superfund Hazardous Substances List

Analytes to be added to the RCRA Appendix VIII List (EPA memo from Robert April to Marcia Williams, 20 Dec 85)

Paragraph 4(c) List (NRDC vs Train, op. cit.; high priority compounds detected in wastewaters)

List of Analytes (Appendix C of the EPA Technical Report (EPA-530-SW-87-005). These are analytes specific to EPA - Industrial Technology Division programs)

This list includes all classical analytes and metals but does not list every individual organic compound.

A few specific organic compounds not on the list were tested for in wastes from centralized treatment facilities, from centralized disposal pits, and from certain other sources. These compounds were all of the tetra- through octa-chlorodibenzo-p-dioxins and furans (136 compounds in all). Pesticides and herbicides also were tested for in wastes from centralized treatment facilities, from centralized disposal pits, and from certain other sources.

The analytes are summarized in Table 2.

SAMPLE SIZE

The total number of sample sites was planned to be 48; 49 sites were actually sampled in this project. The Agency distributed the number of sample sites among the types of sites needed as shown in Table 3.

Eighty-six percent of the sites sampled were active drill or production sites. This allocation reflected the Agency's interest in developing information regarding drill wastes and produced water. Fourteen percent of the sites sampled were centralized waste storage or treatment facilities. This allocation addressed the need to develop information regarding long term storage of wastes, commingling of wastes, and/or

Table 2. Summary of Analytes

	Number of Analytes	
Organics		
Volatiles by GCMS		
Isotope dilution	32	
Reverse search	23	
Semi-volatiles by GCMS		
Isotope dilution	82	
Reverse search	94	
Dioxins and furans by GCMS	136	
Pesticides by GC		
Electron capture detector (ECD)	39	
Flame photometric detector (FPD)	35	
Herbicides by GC/ECD	3	
Total organics		444
Metals		
Atomic absorption	6	
Calibrated Inductively Coupled Plasma (ICP)	21	
ICP screening	41	
Total metals		68
Conventionals by wet chemistry		19
RCRA (Corrosivity, Ignitability, Reactivity)		3
Total number of analytes		534

Table 3. Distribution of Types of Sample Sites

Type of Site	Number of Sites Sampled
Drill sites	19
Production sites	
Produced water	21
Tank bottom sludges	2
Centralized pits	4
Centralized treatment	3
Total	49

treating wastes from multiple oil and/or gas exploration, development, or production sources.

The sampling design, which specifies types of sites, numbers of sites, and the numbers and types of samples to be collected, was restricted by:

- Funding
- Sampling time span (June-September, 1986)
- Availability of qualified sites
- Logistics

SITE SELECTION PROCESS

The site selection process was designed to handle two types of sites: (1) sites randomly selected, and (2) sites specifically selected on the basis of judgement. Known populations were available for use in the site selection process for drill sites and for production sites. These populations were utilized for development of a protocol that produced randomized site selection. A description of that protocol follows in "Random Site Selection Process." Sixteen of nineteen total drill sites were selected randomly. Seventeen of twenty-one total production sites were selected randomly. The remaining sites were reserved for specific site selection.

Centralized pits and centralized treatment facilities were selected on the basis of judgement. About 15 percent of all drill sites and 19 percent of all production sites were also specifically selected. These sites were used to supplement equal distribution of sites, to examine particular practices of

interest, or to replace randomly selected sites which could not be sampled. The site selection rationale for specifically selected sites is described in "Specific Site Selection Process."

Random Site Selection Process

Random site selection was conducted in three tiers. The first tier was based on the zones described in the section entitled "Target Populations" and in Figure 1. Randomly selected sites were distributed uniformly across the zones to ensure coverage. Eighteen potential drill sites were assigned evenly across the U.S., two to each active zone (see Implementation of Site Selection for final account of sites sampled). A similar process was used to identify zones for production sample sites. The first nine production sites were distributed equally through the zones, one site to each active zone. The remaining samples were randomly distributed through the zones (see Figure 2 for example of random selection of states within zones; a similar process was used to select which zones had two production sites.)

The second tier identified the state in which each site was to be located. This was essential to the sampling design since state agencies provided current population data. Populations were weighted within each zone according to statewide drilling or production activity at the time of site selection. Drill sites were weighted on the basis of total wells drilled in each state. Production sites were weighted on the basis of total annual hydrocarbon production. The basis for drill site and production

site weighting is presented in Table 4. It should be noted that 1984 data were the most recently published data at the time of site selection.

As shown in Table 4, probabilities were assigned as a function of percent of zone drilling activity for drill sites and as a percent of zone hydrocarbon production for production sites. Using these probabilities and a random number table (Sorcim/IUS, 1984), two states were selected for drill site sampling in each zone. (In four cases the same state was randomly selected for both drill sites.) A similar process was used to identify states for production samples. An example of the selection process is presented in Figure 2.

In the third tier, EPA requested assistance from state agencies to develop lists containing potential sites for the scheduled date of each sample. Each list needed at least five potential sites unless fewer than five appropriate sites were available. (The actual size of the lists ranged from one to several thousand.) For drill sites, EPA requested a list of sites in a specific basin, with the well at or near completion. For production sites, EPA requested a list of active production sites in a specific basin. EPA assigned sequential integers to the members of each list. (In some cases, unique state permit numbers were used for site selection.) Using a random number table (Sorcim/IUS, 1984), each site was selected from a list of potential sites. A second site was also randomly selected from

Table 4. Percentage (by Zone) of 1984 Oil and Gas Production And Drilling Activity

	Percent of Zone Production	Percent of Zone Drilling
Zone 2		
Kentucky	54	37
New York	4	6
Pennsylvania	21	34
Tennessee	5	3
Virginia	0	1
West Virginia	16	19
TOTAL	100	100
Zone 4		
Alabama	7	5
Arkansas	8	11
Florida	8	0
Louisiana	66	76
Mississippi	11	8
TOTAL	100	100
Zone 5		
Illinois	36	28
Indiana	7	10
Michigan	38	12
Ohio	19	50
TOTAL	100	100
Zone 6		
Kansas	55	85
Nebraska	5	5
North Dakota	39	9
South Dakota	1	1
TOTAL	100	100
Zone 7		
Oklahoma	13	26
Texas	87	74
TOTAL	100	100
Zone 8		
Montana	19	32
Wyoming	81	68
TOTAL	100	100
Zone 9		
Colorado	20	46
New Mexico	54	41
Utah	26	13
TOTAL	100	100
Zone 10		
California	100	100
Zone 11		
Alaska	100	100

STEP 1: Choose a zone for selection of states within zones. In this example, Zone 7 was chosen.[1]

STEP 2: Using the most recent published reports, assign sequential integers to each state within Zone 7 (our example), in proportion to the type of activity and sample. In this example, 26 integers (sequentially 001 through 026) were assigned to Oklahoma. This assignment corresponds to Oklahoma's 26 percent of the drilling activity in Zone 7. The next 74 integers (027 through 101) were assigned to Texas (corresponding to Texas' 74 percent of the zone's drilling activity).

NOTE: For selection of production sites, integers were assigned proportional to percent of zone production.

STEP 3: Use a random number table to select the appropriate number of sites (see below). In this example, two drill sites were to be selected. As can be seen on the random table below, the first two integers encountered between 001 and 101 were 039 and 098.[2] Both integers fell between 027 and 101; thus, both sites were selected to be sampled in Texas.

.4562287	.0285848	.5731043	.6380392	.7548944	.3618006	.2250784	.1348046
.2155620	.9452450	.1556293	.4536573	.7071668	.3958532	.4816542	.2067786
.4938779	.7347435	.8429384	.3673052	.9131544	.1442967	.0370935	.6366895
.1168452	.6056545	.7068156	.3519544	.9943023	.2877842	.9730231	.6278825
.0566039	.0754877	.4359655	.4956858	.9607196	.0899488	.2435953	.4494166
.7653243	.6655329	.1916084	.9510508	.8813455	.1681906	.0238223	.9777881
.3698315	.2289397	.6174680	.1835056	.9382041	.2755072	.4383974	.7996694
.3141360	.2669989	.3748569	.8571130	.1391276	.3909494	.8686801	.5850169
.1252473	.6559180	.9897554	.7194242	.9280202	.0025309	.3163665	.5458144
.8458867	.7358407	.9800887	.5110891	.8861431	.7678877	.9859674	.2459249
.1100185	.7523098	.0387286	.8410738	.1342213	.7776624	.2078029	.9753584
.8162183	.0272849	.4106072	.3258962	.7370280	.1285043	.0630326	.8790774
.1424900	.8112515	.4064430	.8053689	.6711154	.8894297	.1787127	.3390880
.3318185	.4773176	.6647042	.0880304	.0038059	.4757344	.4668019	.3502432
.6895318	.1914800	.9350022	.8752769	.4096173	.2021570	.2696246	.7030698
.2841714	.5214285	.1785654	.3206702	.0837761	.4720133	.0016623	.2077818
.7635107	.4388415	.8551911	.8988911	.3613879	.1734906	.6863279	.7909816

[1]Zone 7 is chosen as an example. The process was followed for selection of states within each zone.

[2]The random numbers presented here represent an excerpt from a table generated using Supercalc®[3] software from the Sorcim/IUS Corporation (Sorcim/IUS, 1984).

Figure 2. Example of Process Used for Selection of States Within Zones.[1]

the same list. The second site served as a back-up site when the primary site proved inaccessible or inappropriate. For example, flooding prevented access to one site. At another site the sampling team arrived to find a four month old producing well at the "drill site" which was selected. In both cases, the back-up site was sampled.

Details of the site selection process for each site is presented in trip reports prepared for each site. These reports are presented in "Technical Report on the Exploration, Development, and Production of Crude Oil and Natural Gas" (31 January 1987).

Specific Site Selection Process

A total of 30.6 percent of all sites in the sampling project were specifically selected on the basis of judgement. All centralized pits and centralized treatment facilities were specifically chosen in addition to several drilling, production, and tank bottom sites.

Other specifically selected sites were used to supplement equal distribution of sites, to examine particular practices of interest, or to replace random selected sites which could not be sampled. Specifically selected sites included three drilling sites, four production sites, one tank bottom site, four centralized pits, and three centralized treatment facilities.

One drill site was specifically selected in each of Zones 7, 10, and 11. In Zone 7, a drill site was specifically selected because the random site selection process did not choose Oklahoma, a state with considerable drilling activity (see Table 1). In Zones 10 and 11, circumstances of site availability, scheduling, and accessibility compelled specific selection of one drill site in each zone.

One production site was specifically selected in Zone 7 and three sites were specifically selected in Zone 4. In Zone 7, a production site was specifically chosen because the random site selection process did not choose Oklahoma, a state with considerable production (see Table 1). Three sites in Zone 4 were specifically selected to further characterize the Gulf Coast Salt Dome Basin.

A specifically-selected tank bottom site was chosen in Zone 10. The absence of a tank bottom sample at the random selected site compelled selection of a specific site nearby.

All centralized pits and centralized treatment facilities were specifically selected. One centralized pit was chosen in Zone 7 and one in Zone 8; two were selected in Zone 9. One centralized treatment facility was chosen in each of Zones 2, 4, and 11. Locations and numbers of centralized pits and treatment facilities were identified through discussions with state officials through the EPA Eastern Workshop (EPA, 1985a) and EPA

Western Workshop (EPA, 1985b) and through discussions with state officials at the time the state officials were generating lists of possible drill sites and production sites. No particular state or zone was chosen in advance of selecting a centralized pit or centralized treatment facility.

Summary of Sampling Design

Table 5 shows the detailed sampling design for this sampling project as of May 5, 1986. (Actual implementation of the plan is discussed in the next section.) EPA targeted 70 percent of the planned 48 samples to be randomly selected, to be distributed among the zones as shown in Table 5. Zone distribution of the specifically-selected sites was considered tentative in recognition that these sites might have to be utilized on a quick-response basis as replacements for randomly-selected sites.

IMPLEMENTATION OF THE SAMPLING DESIGN

Table 6 shows the actual implementation of the sampling design. A comparison between Table 5 and Table 6 shows the degree of implementation of the original sampling plan. Overall implementation of the sampling strategy was high. For example, EPA had targeted 70 percent of the sites for random selection; 69.4 percent of the sites were actually randomly selected.

EPA considered comments on the May, 1986, EPA Sampling Strategy - Draft (EPA, 1986) in making some revisions to the sampling design prior to its initiation. For instance, EPA recognized that some

Table 5. Planned Distribution of Probability Sites and Specifically-Selected Sites Among Zones.

Zone	NO. PROBABILITY SITES Drill	Prod.	T.B.	NO. SPECIFIC SITES Drill	Prod.	T.B.	Cent. Pit	Cent. Trtmt.	TOTAL
2	2	2	-	-	-	1	-	1	6
4	2	2	-	-	-	-	-	1	5
5	2	2	-	-	-	-	-	-	4
6	2	1	-	-	-	1	-	-	4
7	2	2	-	1	-	1	1	-	7
8	2	1	-	-	-	1	-	-	4
9	2	2	-	1	-	-	2	-	7
10	2	1	-	-	1	1	-	-	5
11	2	1	-	-	1	1	-	1	6
TOTAL	18	14	-	2	2	6	3	3	46

Abbreviations: Drill - drill site
Prod. - production site
T.B. - tank bottom sludge site
Cent. Pit - centralized pit site
Cent. Trtmt. - centralized treatment site

Table 6. Distribution of Probability Sites and Specifically-Selected Sites Among Zones.

Zone	NO. PROBABILITY SITES Drill	Prod.	T.B.	NO. SPECIFIC SITES Drill	Prod.	T.B.	Cent. Pit	Cent. Trtmt.	TOTAL
2	2	2	1	-	-	-	-	1	6
4	2	2	-	-	3	-	-	1	8
5	2	1	-	-	-	-	-	-	3
6	2	1	-	-	-	-	-	-	3
7	2	2	-	1	1	-	1	-	7
8	2	2	-	-	-	-	1	-	5
9	2	3	-	-	-	-	2	-	7
10	1	2	-	1	-	1	-	-	5
11	1	2	-	1	-	-	-	1	5
TOTAL	16	17	1	3	4	1	4	3	49

Abbreviations: Drill - drill site
Prod. - production site
T.B. - tank bottom sludge site
Cent. Pit - centralized pit site
Cent. Trtmt. - centralized treatment site

tank bottom sludge samples might be available at randomly-selected production sites. As randomly-selected sites were preferred, Table 6 shows the realization of that expectation.

Of eighteen planned random drill sites, sampling was actually conducted at sixteen. EPA was unable to locate a drill site at or near completion during the sampling time span in the Cook Inlet Basin of Alaska (Zone 11), or in southwest Wyoming (Zone 8). The planned Cook Inlet Basin site was replaced by a northern Alaska site (this one specifically selected). The Wyoming site was not replaceable because of the absence of any available site within the schedule.

EPA conducted sampling at seventeen randomly selected production sites. Only one of two planned production sites was sampled in Zone 5 (planned for Michigan) due to scheduling difficulties. To compensate for this loss, EPA added one randomly-selected production site to Zone 10 (in California) and one to Zone 11 (in Alaska) later in the program.

EPA conducted tank bottom sludge sampling at two random sites and one specifically-selected site, one less than originally planned. Four centralized pits were sampled, one more than originally planned. Centralized treatment facilities were sampled as planned.

STATISTICAL ANALYSIS

SAMPLING DESIGN

One objective of the sample collection was to form estimates of the national average pollutant concentration in each waste unit. Statistical sampling methods were used to construct a sampling design which would allow construction of such estimates, while still allowing for examination of particular sites of special interest.

In order to gain the maximum amount of information from the numbers of samples and analyses planned, and to improve the precision of the estimates derivable from this study, a multi-stage stratified sampling design was employed. The effectiveness of stratification depends on the association of the stratification attributes with the quantities of interest. The 11 zones or strata in this study were geographically defined to group common geological formations and operating practices together, as these were deemed to be likely to be related to the waste characteristics. Optimal allocation of samples to strata depends on the relative numbers of sampling units in the strata and the relative sizes of the intrastratum variances of the quantities of interest. In this case, optimal sample size allocation across strata was not possible, both because of the limited prior information on the strata sizes and variances, and because of the limited number of samples to be taken. The samples were distributed uniformly across the strata to ensure

maximal coverage--about two samples for each of the none zones in the study (two zones have been excluded because there is essentially no known oil and gas production, exploration, or drilling in these zones).

The second stage of sampling was to select sites within zones. Because no complete frame of sites was available, the site selection was performed by selecting a state according to specified probability from the region, and then selecting a site from a list of sites meeting the design attributes (type of site, etc.) known to the state regulatory agencies. For each site assigned to a region, the state was selected with probability proportional to the total oil and gas production in the region. The use of such size measures provides sample estimates that are more efficient for the key quantities of interest. In calculating the estimated national average concentrations of pollutants, the measured pollutant concentrations will be weighted by the actual waste volume at the site to produce volume-weighted average concentrations, with appropriate accounting for the probabilities of selection of each site. The key factor to obtaining valid national estimates is to obtain samples according to a protocol that will produce a national random sample.

As described in the Site Selection Process section, the sampling design consisted of two sample site types: those selected as a random sample of sites and those selected on the basis of

judgment. The analytical results from the samples collected at these sites were combined to form national estimates as follows. Each directed sample collected at a specifically selected site represents that site alone in any national estimate. The random samples represent the population from which they are drawn. The known probability of selection is used to weight the results of samples collected at randomly-selected sites, and are combined with the results of the directed samples in the proportion of the waste volume at the directed sites to the waste volume of the sampling population in producing the zone estimates that go into the national average.

PRECISION AND ACCURACY

Clearly, about 20 randomly selected samples for an industry segment is a very small sample size for a national estimate. To get a sense of the precision, suppose we were to base a national estimate of a proportion of sites with a characteristic such as significant level of a pollutant, a simple random sample would provide an estimate $p = x/20$.

This value can at best estimate a value within +/- 5% just because the denominator is 20. That is, the possible estimates for 100p% would be 0%, 5%, 10%, 15%, 20%, etc. Furthermore, the estimate p is subject to sampling error of 2*SQR(p*(1-p)/20), this being two times the standard error of the estimate. For an estimate of 100p% = 50%, the sampling error would be on the order of 22%.

These calculations demonstrate the difficulty of using this small sample size to provide discriminating national estimates. Furthermore, regional estimates would be based on an even smaller sample size and subject to a larger sampling error. Consequently, the study is truly a screening study designed to detect on the most blatant or extreme results reliably. As in any sampling effort, the unexpected or important special case discovered by this process can be the basis for further study by EPA, the Region or State Agencies, if warranted.

Though precision calculations for the multi-stage sampling plan for this study would be more difficult, calculations based in simple probability sampling give a general idea of the expected precision for a national survey of this size. A larger sample size was not used because of the cost constraints. The actual sample design uses techniques typically employed in large national surveys to improve the efficiency of sampling that is subject to major constraints such as travel time to reach a sampling site. Accuracy depends on the survey design and our ability to follow the protocol. With the small sample size, adherence to the sample design and protocol is especially important.

CONCLUSION

Based on the above considerations, EPA believes that the strategy described in this document meets its needs for screening for pollutants in wastes from the oil and gas extraction industry. The information gathered in this project may be used to develop more detailed strategies for further definition of various parts of this industry, for greater coverage of strata, site types, or pollutants specific to a given process within the industry.

REFERENCES

American Petroleum Institute. 1986. Quarterly Completion Report, Fourth Quarter, Vol. 1, No. 4 (March).

"First Quarter Crude Price, Drilling Plunges Wrack U.S. Oil Industry." Oil and Gas Journal. 1986. Vol. 84, No. 13 (March 31).

Independent Petroleum Associates of America. 1986. The Oil and Gas Producing Industry in Your State, 1986-1987. (September).

Interstate Oil Compact Commission. 1984. The Oil and Gas Compact Bulletin, Vol. XLIII, No. 2 (December).

Interstate Oil Compact Commission. 1985. The Oil and Gas Compact Bulletin. Vol. XLIV, No. 1 (June).

U.S. DOE. 1985. Monthly Energy Review, DOE/EIA-0035 (85/10).

U.S. EPA. 1979. Effluent Guidelines Division. Office of Water and Hazardous Materials. Development Document for Interim final Effluent Limitations Guidelines and Proposed New Source Performance Standards for the Oil and Gas Extraction Point Source Category. Prepared by Russell E. Train, Andrew W. Breidenbach, Eckardt C. Beck, Robert B. Schaffer, Martin Halper (September).

U.S. EPA. 1985. Industrial Technology Division. Onshore Oil and Gas Workshop Meeting Report (July).

U.S. EPA. 1985. Industrial Technology Division. Proceedings Onshore Oil and Gas State/Federal Western Workshop (December).

U.S. EPA. 1986. Industrial Technology Division. Oil and Gas Exploration, Development, and Production - Sampling Strategy - Draft (May).

Federal Register. 1979. Vol. 44, No. 73 (April 13). Effluent Guidelines and Standards, Oil and Gas Extraction Point Source Category, 40 CFR Part 435.

Sorcim/IUS Corporation. 1984. SuperCalc®3, Sorcim/IUS Micro Software, Release 2 (August).