CIFE CENTER FOR INTEGRATED FACILITY ENGINEERING

UTILIZATION OF SARA DATABASE TO IDENTIFY DESIGN STRATEGIES FOR MINIMIZING EMISSIONS OF HAZARDOUS CHEMICALS

by

Royal J. Kopperud
Lynn M. Hildemann
Paul V. Roberts

TECHNICAL REPORT
Number 74

September, 1992

Stanford University

If you would like to contact the authors please write to:

c/o CIFE, Civil Engineering,
Stanford University,
Terman Engineering Center
Mail Code: 4020
Stanford, CA 94305-4020

CIFE SUMMARY: Technical Report No. 74 **Date:** September, 1992

Title: Utilization of SARA (Superfund Amendments and Reauthorization Act) Database to Identify Design Strategies for Minimizing Emissions of Hazardous Chemicals

Authors: Royal J. Kopperud (graduate student, Civil Engineering)
Lynn M. Hildemann (Assistant Professor, Civil Engineering)
and Paul V. Roberts (Professor, Civil Engineering)

Funding: CIFE Seed Grant, plus partial student support from an NSF PYI award.

Subject: This report describes research examining the potential utility to the construction industry of information contained in the SARA Database, also known as the toxic release inventory (TRI). This nationwide computerized database, which quantifies releases of hazardous chemicals from industries, was considered a possible source of information regarding successful strategies for designing industrial facilities to minimize the emission to the atmosphere of hazardous chemicals.

Objectives: A spate of new regulations requiring reductions in environmental releases of hazardous chemicals is forcing many companies to undergo costly retrofitting of their existing facilities. In addition, little information is available on how best to design new facilities to meet current and pending regulations. The goal of this project was to identify strategies for utilizing the TRI in the planning process to minimize the release of toxic emissions to the atmosphere from new industrial facilities.

Methodology: The project consisted of three phases:

- Closely examine the items and level of detail available in the database, to determine how much and what kind of information is potentially available in the database.
- Evaluate the accuracy of the reported quantities, to determine how reliable the database would be for identifying promising design strategies.
- Interpret the trends observed in toxic air emissions reported in the TRI for three categories of industry, to investigate the extent to which a few chemicals may be responsible for the majority of industrial emissions, and to evaluate the ease with which emission reduction strategies might be implemented.

Results: A computer program was developed to access the data contained on the CD-ROM and format it for export into 4th Dimension; however, this was considered quite straightforward. The main contributions of this research project were related to evaluating the TRI as an environmental database.

Based on the entries it contains, the TRI might allow identification of "problem areas" for a given type of industrial facility. However, due to uncertainties regarding the accuracy of the data, extreme caution must be utilized in interpreting results reported by one or a few companies. In addition, the level of detail available in the TRI is insufficient to provide much insight on the relationship between specific industrial processes and their associated emissions. Finally, due to chronic delays in releasing the database on CD-ROM, the TRI cannot be considered as an up-to-date source of information for construction planning. Nonetheless, in the future, the TRI may become useful for identifying retrofitting strategies, through information available in the waste minimization section.

Status: At the end of this one-year seed project, a fairly complete evaluation has been made of the 1987 TRI database. Evaluation of more recent years and examination of trends in the data from one year to the next awaits the release of more recent data on CD-ROM. Given the results reported above, the next logical step would be to examine the 1991 database, the first year of data where reporting of waste minimization strategies is required rather than optional. This would allow an evaluation of the utility to industries of information describing waste minimization strategies.

CIFE Seed Research Project

Final Report

UTILIZATION OF SARA DATABASE TO IDENTIFY DESIGN STRATEGIES FOR MINIMIZING EMISSIONS OF HAZARDOUS CHEMICALS

Royal J. Kopperud
Lynn M. Hildemann
Paul V. Roberts

Environmental Engineering Group
Department of Civil Engineering

I. Introduction

Title III of the Superfund Amendments and Reauthorization Act (SARA) of 1986 requires every manufacturing facility in the United States to annually submit toxic chemical release inventory forms to the U.S. Environmental Protection Agency (EPA) if they: (i) fall within Standard Industrial Classification (SIC) Codes 20–39; (ii) have 10 or more employees; and (iii) manufacture, process, or otherwise use any one of over 300 listed chemicals in excess of specified threshhold amounts (Sarokin, 1988; USEPA, 1989a, 1989b).

The information required on the inventory forms includes the amounts of each of the hazardous chemicals present at the facility, the types of processes utilizing each chemical, the amounts released to different parts of the environment (*e.g.*, air, water, land), the estimation technique used for each of the release estimates, and the waste treatment methods used and their efficiencies (USEPA, 1989b; Krieger, 1989). As an option, information on the strategies used to reduce or minimize the hazardous wastes generated can also be documented.

SARA Title III also requires the EPA to generate an annual inventory of toxic chemical emissions, based on the information received from the manufacturing facilities. This inventory, which must be made accessible to the public, is intended to inform both governmental agencies and the public about routine and accidental releases of toxic chemicals to the environment. Besides being available to the public on a state-by-state basis (via floppy discs), the entire, nationwide inventory is intended to be available both on magnetic tapes and on a CD-ROM.

Escalating concerns about atmospheric toxins are resulting in a spate of new regulations regarding hazardous chemical emissions. These regulations are having a substantial impact on industries: many companies are faced with costly retrofitting of their existing facilities, and little information is available on how best to design new facilities to meet current and pending regulations. The SARA database could be promising as a source of data for identifying successful strategies currently used by various industries to handle, treat and manage hazardous chemicals in the workplace.

The goals of this project were:

- To examine the potential utility to the construction industry of information in the SARA database pertaining to treatment and management practices for minimizing hazardous emissions to the atmosphere, focusing on three industry groups.

- To identify strategies for utilizing the database to obtain information useful to the construction industry.

II. Description of Database Parameters

The database contains a substantial amount of information. Each reporting facility must submit a separate report for each toxic compound which is manufactured, processed, or otherwise used in amounts exceeding a threshhold value. Table I (next page) shows an example of the entries typically found. Besides a variety of information regarding the facility location and contact person, the database also includes estimates of the amounts of the chemical released to the air, the water, and the land; the amounts transported offsite or treated onsite; and (as an optional section) the methods employed to reduce the amount of waste generated.

This seed project focused on the information available in the database regarding the release of toxic chemicals to the atmosphere. As shown in Figure 1, for 1987, one-quarter of the toxic chemical releases listed in the inventory were released to the atmosphere, indicating that this is a significant mode of environmental contamination. Out of the range of industrial groups filing TRI reports (shown in Table II), this project focused on three industries expected to have substantial toxic chemical releases to the atmosphere. Thus, the database was prescreened to select just those industries having 4-digit SIC codes of 2811–2899 (Chemicals and Allied Products), 2911–2999 (Petroleum Refining and Related Industries), and 3611–3699 (Electrical and Electronic Machinery, Equipment, and Supplies).

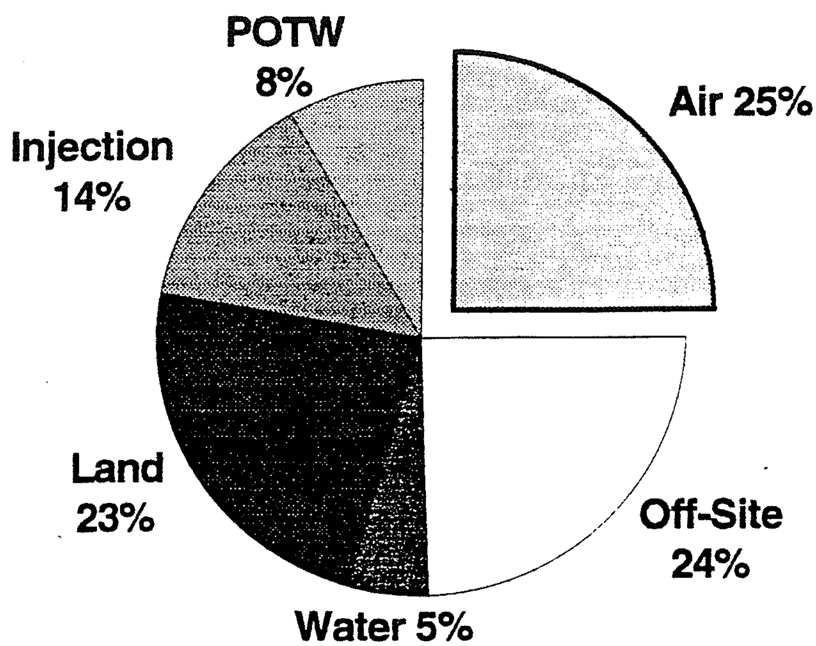


Figure 1. Fraction of Releases to each Medium

Table I. TRI Data Structure

EPA Submission No.	1387010179707SD
Facility Name	Smith Equipment
Facility Address	2601 Lockheed Ave. Codington County Watertown, SD 57201
EPA Region	8
Public Contact/Phone No.	James Goble, Lynn Hedin 605-882-3200
Latitude	044 Deg, 56 Min, 00 Sec
Longitude	097 Deg, 09 Min, 00 Sec
Facility Dun & Bradstreet Num.	00-625-0880
EPA ID Number	SDD980635783
NPDES Permit No.	23-4567-0093
Parent Co. Name	Tescom Corporation
Parent Co. D & B Number	00-625-0880
Substance Name	Trichloroethylene
CAS Registry Number	79-01-6
Manufacturing Uses	As a manufacturing aid
SIC Code(s)	3400 (Fabricated Metal Products) 3451 (Screw Machine Products) 3471 (Plating and Polishing)
Max. Amount On-Site	0 – 99
Non-Pt. Air Release (lbs/yr)	14798
Basis of Estimate	C (Mass Balance Calculations)
Point Air Release (lbs/yr)	
Basis of Estimate	
Water Release (lbs/yr)	
Stream Name	
Storm Water %	
Undergrd. Injection (lbs/yr)	
UIC IC Number	
Land Release (lbs/yr)	
Off-Site Transfer (lbs/yr)	5000
Off-Site EPA ID	WID990829475
Off-Site Name	Waste Research & Reclamation Co. Route 7 and Hwy 93 South Eau Claire, WI 54701
General Wastestream Treatment	
Treatment Method used on Wastestream	NA (not applicable)
Waste Minimization/ Modification	M7 (Improved Housekeeping, Training, Inventory Control)
Current Yr Quantity (lbs)	5050
Prior Yr Quantity (lbs)	6600
Percent Change	
Ratio Production	
Reason for Action	R2 (Reduction of Treatment/Disposal Costs)

Table II. Industrial Groups Required to File TRI Reports

2-Digit Code	Industry Group
20	Food
21	Tobacco
22	Textiles
23	Apparel
24	Lumber and Wood
25	Furniture
26	Paper
27	Printing and Publishing
28	Chemicals
29	Petroleum and Coal
30	Rubber and Plastics
31	Leather
32	Stone, Clay, and Glass
33	Primary Metals
34	Fabricated Metals
35	Machinery (excluding electrical)
36	Electrical and Electronic Equipment
37	Transportation Equipment
38	Instruments
39	Miscellaneous Manufacturing

The data entries selected from the complete data set for further analysis were those relating to the estimated air releases, along with any information available regarding waste minimization strategies. Table III lists the data subset which was selected for input into a data management software package. The EPA submission number was used to uniquely identify each record.

Table III. Filtered Data Subset

TRI Abbrev.	Data Entry
SN	EPA Submission Number
NS	Name of Substance
SC	SIC Code
NP	Non-Point Air Release (lbs/yr)
BN	Basis of Non-Point Estimate
PA	Point Air Release (lbs/yr)
BP	Basis of Point Estimate
WM	Waste Minimization/Modification
WY	Waste Min.–Current Yr Qty (lbs)
WP	Waste Min.–Prior Yr Qty (lbs)
WC	Waste Min.–Percent Change
WX	Waste Min.–Ratio Production
WA	Waste Min.–Reason for Action

The 4-digit SIC code was included to provide more detailed information on the activities of the industry. Many industries have multiple SIC codes, and the TRI provides space for up to three SIC codes to be listed. As shown in Table IV (next page), the 4-digit SIC codes utilized in the three industrial categories examined in this study give a more specific idea of the scope of industrial activities taking place.

The ability of the database to distinguish between point and non-point air releases was considered important. Point sources (like tall stacks) are relatively easy to quantify, control, and/or treat; in contrast, emissions from non-point sources (like open containers, valve leaks, spills) are much more difficult to mitigate. The filtered data subset contained the mass emission rate to the atmosphere estimated on an annual basis for both point and non-point sources.

For both the point and non-point source emission rates, a basis for determining the estimate was indicated. As shown in Table V, estimates could be based on actual measurements, mass balance calculations, published emission factors, or "best engineering judgement".

Table V. Categories for Basis of Estimate

M – Estimate based on monitor data or measurements for the toxic chemical as released to the environment and/or off-site facility.

C – Estimate based on mass balance calculations, such as calculation of the amount of the toxic chemical in streams entering and leaving process equipment.

E – Estimate based on published emission factors, such as those relating release quantity to through-put or equipment type (e.g., air emission factors).

O – Estimate based on other approaches such as engineering calculations (e.g., estimating volatilization using published mathematical formulas) or best engineering judgment. Includes applying an estimated removal efficiency to a waste stream, even if the composition of the stream before treatment was fully identified through monitoring data.

The Waste Minimization Section, shown in Figure 2, has been optional, although this section will become mandatory in the future. For the entry regarding the type of waste minimization modification, the choices of codes, examples of which are listed in Table VI, range from recycling to better housekeeping. However, for the 1987 data available, very few of the firms filled in this section, as we discovered on closer examination of the data. Accordingly, no further evaluation regarding the possible utility of this information was possible.

Figure 2.

8. OPTIONAL INFORMATION ON WASTE MINIMIZATION

(Indicate actions taken to reduce the amount of the chemical being released from the facility. See the instructions for coded items and an explanation of what information to include.)

A. Type of modification (enter code)	B. Quantity of the chemical in the wastestream prior to treatment/disposal			C. Index	D. Reason for action (enter code)
	Current reporting year (lbs/yr)	Prior year (lbs/yr)	Or percent change		
☐☐	______	______	______ %	☐.☐	☐☐

Table IV. Four-Digit SIC Codes

28 Chemicals and Allied Products

2812 Alkalies and chlorine
2813 Industrial gases
2816 Inorganic pigments
2819 Industrial inorganic chemicals, n.e.c.
2821 Plastics materials, synthetic resins, and nonvulcanizable elastomers
2822 Synthetic rubber (vulcanizable elastomers)
2823 Cellulosic manmade fibers
2824 Manmade organic fibers, except cellulosic
2833 Medicinal chemicals and botanical products
2834 Pharmaceutical preparations
2835 In vitro and in vivo diagnostic substances
2836 Biological products, except diagnostic substances
2841 Soap and other detergents, except specialty cleaners
2842 Specialty cleaning, polishing, and sanitation preparations
2843 Surface active agents, finishing agents, sulfonated oils, and assistants
2844 Perfumes, cosmetics, and other preps.
2851 Paints, varnishes, lacquers, enamels, and allied products
2861 Gum and wood chemicals
2865 Cyclic organic crudes and intermediates, and organic dyes and pigments
2869 Industrial organic chemicals, n.e.c.
2873 Nitrogenous fertilizers
2874 Phosphatic fertilizers
2875 Fertilizers, mixing only
2879 Pesticides and agricultural chemicals, n.e.c.
2891 Adhesives and sealants
2892 Explosives
2893 Printing ink
2895 Carbon black
2899 Chemicals and chemical preparations, n.e.c.

29 Petroleum Refining and Related Industries

2911 Petroleum refining
2951 Asphalt paving mixtures and blocks
2952 Asphalt felts and coatings
2992 Lubricating oils and greases
2992 Products of petroleum and coal, n.e.c.

36 Electronic and Other Electrical Equipment and Components

3612 Power, distribution and specialty transformers
3613 Switchgear and switchboard appar.
3621 Motors and generators
3624 Carbon and graphite products
3625 Relays and industrial controls
3629 Electrical industrial appliances, n.e.c.
3631 Household cooking equipment
3632 Household refrigerators and home and farm freezers
3633 Household laundry equipment
3634 Electrical housewares and fans
3635 Household vacuum cleaners
3639 Household appliances, n.e.c.
3641 Electric lampbulbs and tubes
3643 Current carrying wiring devices
3644 Noncurrent carrying wiring devices
3645 Residential electric lighting devices
3646 Commercial, industrial, and institutional electric lighting fixtures
3647 Vehicular lighting equipment
3648 Lighting equipment, n.e.c.
3651 Household audio and video equip.
3652 Phonograph records and pre-recorded audio tapes and disks
3661 Telephone and telegraph apparatus
3663 Radio and television broadcasting and communications equipment
3669 Communications equipment, n.e.c.
3671 Electron tubes
3672 Printed circuit boards
3674 Semiconductors and related devices
3675 Electronic capacitors
3676 Electronic resistors
3677 Electronic coils, transformers, and other inductors
3678 Electronic connectors
3679 Electronic components, n.e.c.
3691 Storage batteries
3692 Primary batteries, dry and wet
3694 Electric equipment for internal combustion engines
3695 Magnetic and optical recording media
3699 Electrical machiner, equipment, and supplies, n.e.c.

Table VI. Examples of Pollution Prevention Methods

Recycling and reuse on-site
– Solvent Stills
– Vapor recovery
Recycling and reuse off-site
– Commercial recycling service
Equipment modifications
– Floating roof tanks
– Overflow alarms
Process modifications
– Change production schedule
– Segregate wastes
Substitution
Housekeeping
Other

III. Reliability of Data Entries

Substantial effort was devoted towards reviewing the literature regarding the SARA database, and interviewing individuals who were familiar with the database. The main focus of this literature search was to obtain a better idea of the accuracy and reliability of the numbers contained in the database.

A problem that became apparent from this literature review was that, due to the short deadline set for filing reports for 1987, many of the 20,000 facilities responding were only able to obtain extremely uncertain "engineering estimates" for their 1987 releases (Baram *et al.*, 1990; CMA, 1989). Many concerns were raised regarding the reliability of these first-year estimates (Fisher *et al.*, 1988; Thurm, 1990; Hanson, 1990), and preliminary examination of the reports submitted for 1988 supports the suspicion that some of the first year emissions estimates were overly high (Santa Clara County Manufacturing Group, 1990a, 1990b; Associated Press, 1990). Because the utility of this database to the planning and design of industrial facilities will be greatly affected by the reliability of the data, this finding was viewed as a potential problem.

EPA studied the accuracy of the TRI data collected for 1987 and 1988 by having investigators closely examine the reports filed by a representative sample of companies. For 1987, the filings of 156 companies were reviewed, including 44 in the chemical industry. For the 1988 reports, 89 facilities were reviewed, including 43 in the chemical industry and 15 in petroleum refining and electronics (EPA 1990b).

For each company, outside experts independently determined new estimates for comparison with those filed by the companies. Since the TRI legislation does not mandate any new measurements, the companies merely interpret data that they already have available to fill out the TRI reports. Likewise, the EPA investigators did not perform any new measurements; rather they independently reviewed the available data. The EPA's accuracy assessment, therefore, is a measure of the agreement of the TRI reports with expert engineering estimates based on the same measurement database; it does not measure the agreement of the TRI with actual emissions to the environment. Thus, the true accuracy of the TRI database remains uncertain.

The investigations found that some facilities omitted chemicals completely, and others made large errors in their reports; a report was considered in error if quantities were high or low by a factor of 10 or more. For the reports examined, 17% contained errors of underestimation, while 7% were erroneously overestimated, as shown in Figure 3. The aggregate overestimates tended to offset the underestimates, however, and the total air emissions were surprisingly accurate. The accuracy study estimated annual air releases totaling 2.7 billion pounds, while the original reports estimated 2.6 billion pounds (see Table VII).

Figure 3.
Accuracy of Individual Reports

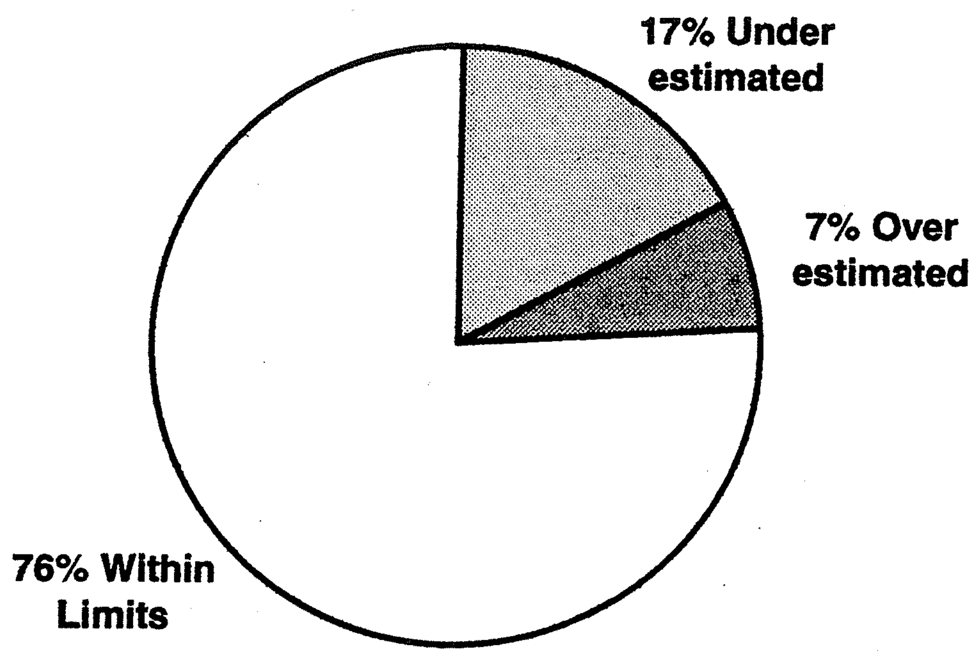


Table VII. Accuracy of Release Estimates

Medium	Fraction with large errors	Fraction of chemicals omitted	Total release in original reports [billion lbs]	Total corrected release [billion lbs]
Non-Point Emissions	13%	14%	0.8	0.8
Point Emissions	7%	16%	1.8	1.9

Total releases reported for all emissions were within 2% of the quality control estimates in 1987, and even better for selected industries in 1988. EPA believes that the quality of data has improved steadily through 1990, as facilities gain experience in making the estimates and filing the forms (Hanson, 1992).

In general, therefore, individual reports are likely to be erroneous. Many reports are in error by a factor of 10 or more and many chemicals are omitted. For large aggregates of data, however, the total emissions are quite accurate. Users of the data should rely on the largest number of reports possible to minimize the importance of errors. Furthermore, because the TRI does not require any new measurements, the accuracy of the reported quantities is limited by the data available to the companies.

In evaluating specifically the reliability of the air emissions estimates for this project, close examination was made of the category of estimation method listed. As shown in Table VIII, almost half of the estimates were based on "engineering judgement", which would not be expected to be very accurate. The chemical and petroleum industries also frequently utilized published emission factors, while the second most popular choice for the electrical/electronic industries was mass balance calculations.

Table VIII. Estimation Methods Used for Air Releases

	Chemical Industry		Electronic Industry		Petroleum Industry	
Method of Estimation	Non-Point	Point	Non-Point	Point	Non-Point	Point
Engineering Judgement	45%	51%	48%	42%	44%	45%
Published Emission Factors	24%	12%	3%	4%	35%	29%
Monitoring Data	7%	7%	4%	9%	3%	6%
Mass Balance Calculations	3%	8%	20%	21%	3%	5%
Not Specified	21%	21%	24%	24%	15%	16%

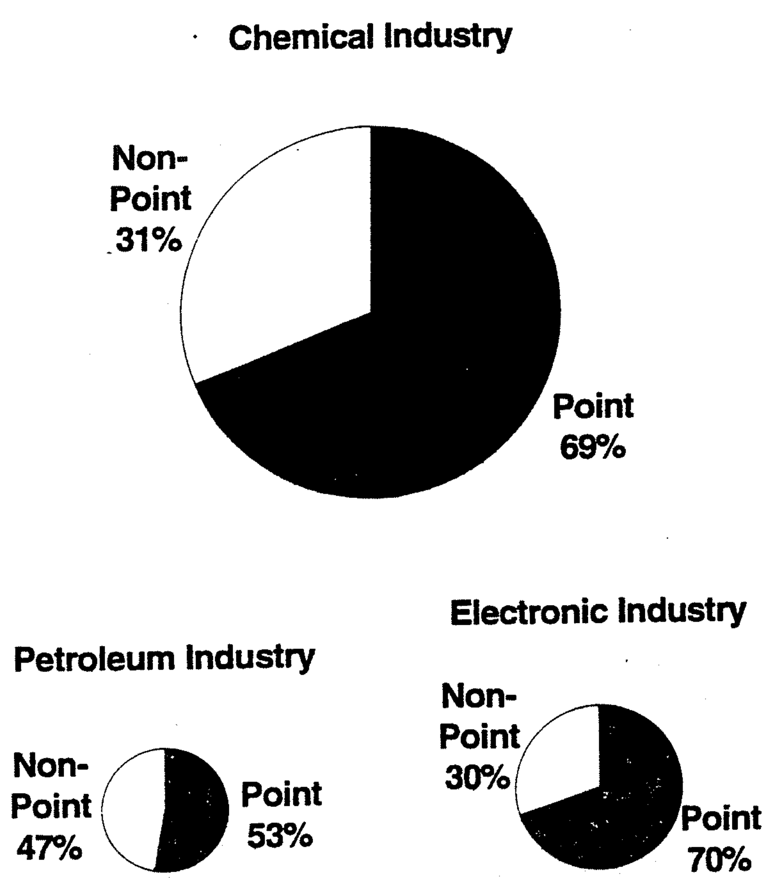
IV. Trends Observed for Industrial Groups

A summary of the 1987 TRI data for the chemical, electrical/electronic, and petroleum SIC classifications is shown in Table IX. Air emissions were listed for 70 to 75% of the filings, with one-third to one-half of the air emissions attributed to non-point sources (see Figure 4).

Table IX. Summary of Air Emissions Data in TRI

	Chemical Industry	Electronic Industry	Petroleum Industry	Total for 3 Industries
Total Air Emissions [ton/yr]	507,000	75,000	57,000	639,000
% Emissions from Point Sources	69%	70%	53%	
% Emissions from Non-Pt. Sources	31%	30%	47%	
Total Records Filed [no.]	23,000	6,300	3,900	33,000
Fraction of the Total TRI Reports	30%	8%	5%	43%
Fraction Showing Air Emissions	75%	72%	71%	
Fraction Showing Point Emissions	59%	55%	49%	
Fraction Showing Non-Pt. Emissions	65%	52%	61%	
Total Number of Toxic Chemicals Released to the Air	254	118	127	

Figure 4. Air Emissions for Each Industry



To examine whether a particular industrial group's air emissions were mainly due to a few large industries or to many comparably-sized facilities, the cumulative emissions vs. number of filings was examined. As shown in Figure 5, out of the 17,000 filings from the chemical industry that included air emissions, 100 (less than 1%) accounted for over half the air emissions. In the electrical/electronic industries, about 100 out of the 4500 filings with air emissions (2%) accounted for about half of the mass released to the air. Finally, for the petroleum industry, which had 2800 filings with air emissions, 100 (4%) accounted for more than half the total mass emissions.

A small number of toxic chemicals were primarily responsible for the air emissions reported by each of these industrial groups. Plotted on a cumulative basis (see Figure 6), four compounds accounted for greater than 50% of the emissions in each SIC group. For the chemical industry, ammonia (a fertilizer ingredient) accounted for over 20% of the air emissions, followed by carbon disulfide, acetone, and methanol (all solvents). For the electrical-electronic industry, all of the major toxic air emissions consist of degreasing solvents: trichloroethane and Freon 113 each constitute about 16% of the total air emissions, while dichloromethane and methyl ethyl ketone (MEK) also contribute substantially. For the petroleum industry, toluene (a fuel component), aluminum oxide, ammonia, and xylenes are the major toxic chemicals emitted.

The amount of chemical released to the atmosphere from point sources vs. non-point sources is an important factor in evaluating how easy or difficult it might be to reduce emissions. As shown in Figure 7, the type of emission source depends on the chemical considered. For the chemical industry, ammonia and carbon disulfide are emitted primarily from point sources. For the electrical/electronic industries, about half of the trichloroethane is emitted from fugitive (non-point) sources, as is about two-thirds of the Freon 113 emissions. For the petroleum industry, most of the toluene emissions are fugitive, while most of the aluminum oxide emissions originate from point sources.

V. Conclusions: Prospects for Future Use

An important goal of this seed project was to utilize the TRI database to assess priorities for environmental improvements and pollution prevention that would impact the design of industrial facilities. We expected that the TRI might enable identification of problem areas for a given size and type of industrial facility, and perhaps provide some guidance in implementing effective pollution prevention measures at the design stage.

Using TRI, it is possible to estimate the quantities of toxic chemicals that would be released by a proposed industrial facility based on TRI data for industries of the same SIC code(s). While selecting "comparably-sized" facilities is not straightforward, the reported maximum amount of chemical on-site can be used as a rough measure of the size of operations. Those chemicals posing a particular pollution problem can be identified for this subset of reports, providing a sounder basis for deciding where money might best be spent on control devices and/or treatments. However, the TRI database did not allow evaluation of toxic emissions on a process-by-process basis, so it does not allow identification of which processes pose particular environmental hazards.

Extreme caution must be utilized in drawing conclusions from the reports of one or a few companies, due to uncertainties regarding the accuracy of the data. EPA's finding that a substantial fraction of the reports contained order-of-magnitude errors or failed to report releases implies that a report from a single company can be quite misleading. However, the total or average releases for all companies in an industry is likely to be fairly accurate,

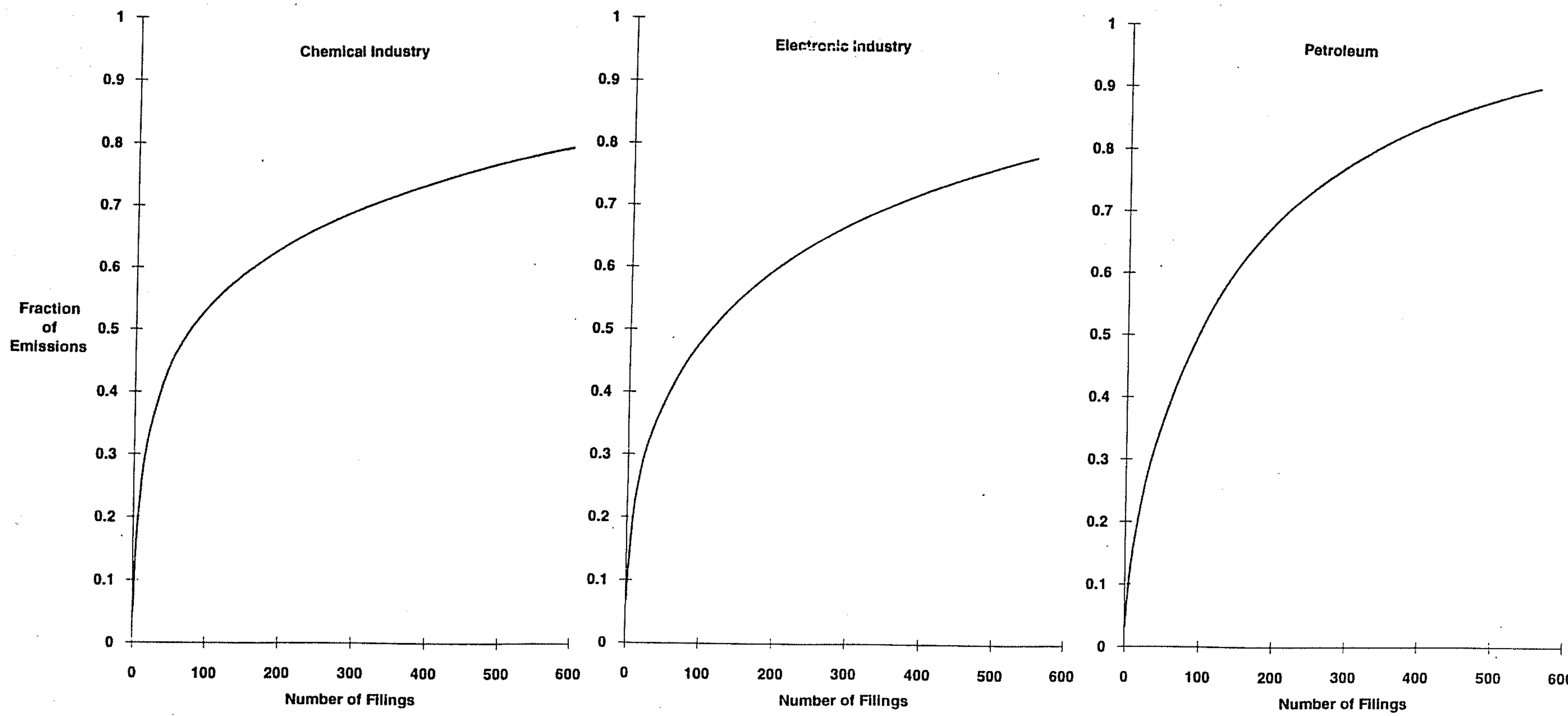


Figure 5. Cumulative Emissions for Each Industry vs. Number of Filings

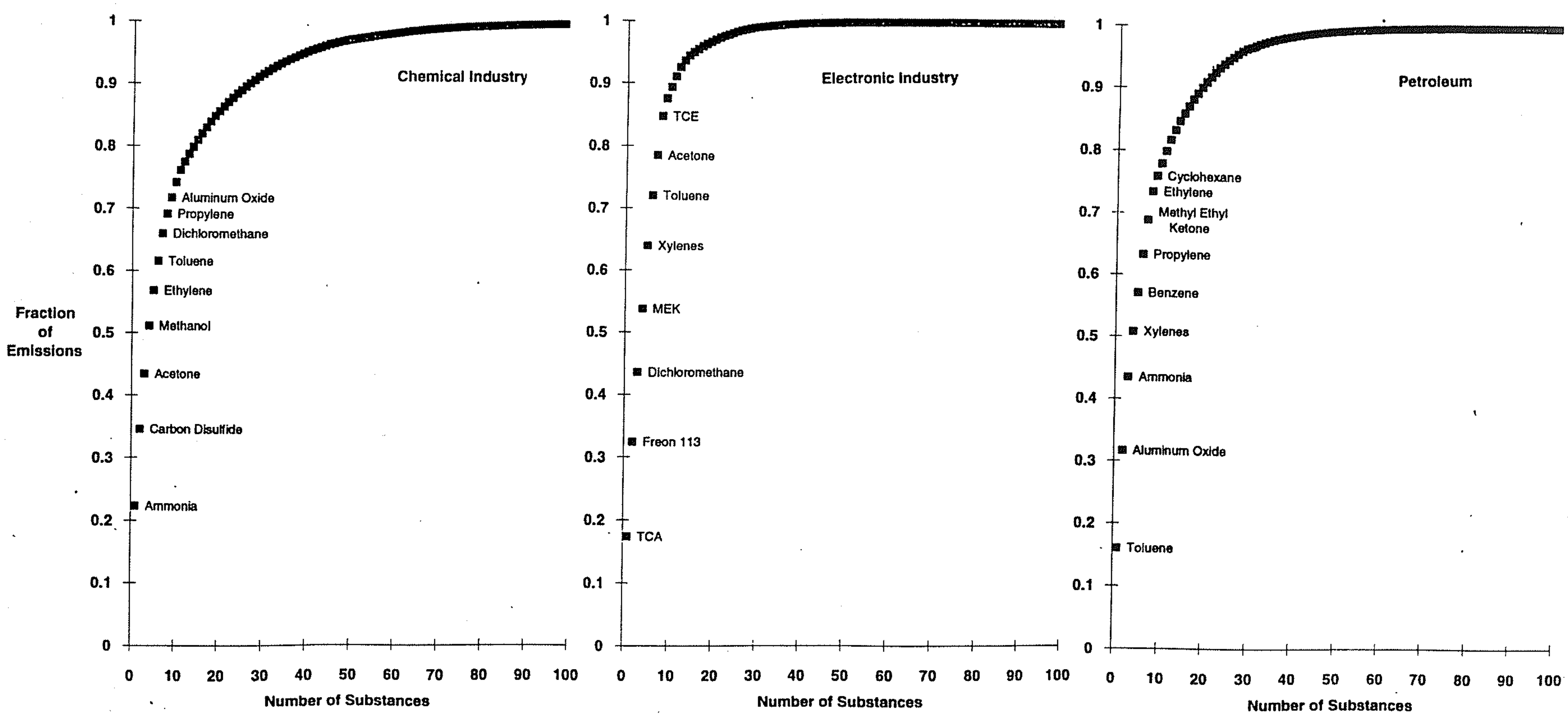


Figure 6. Cumulative Emissions for Each Industry vs. Chemical Compound

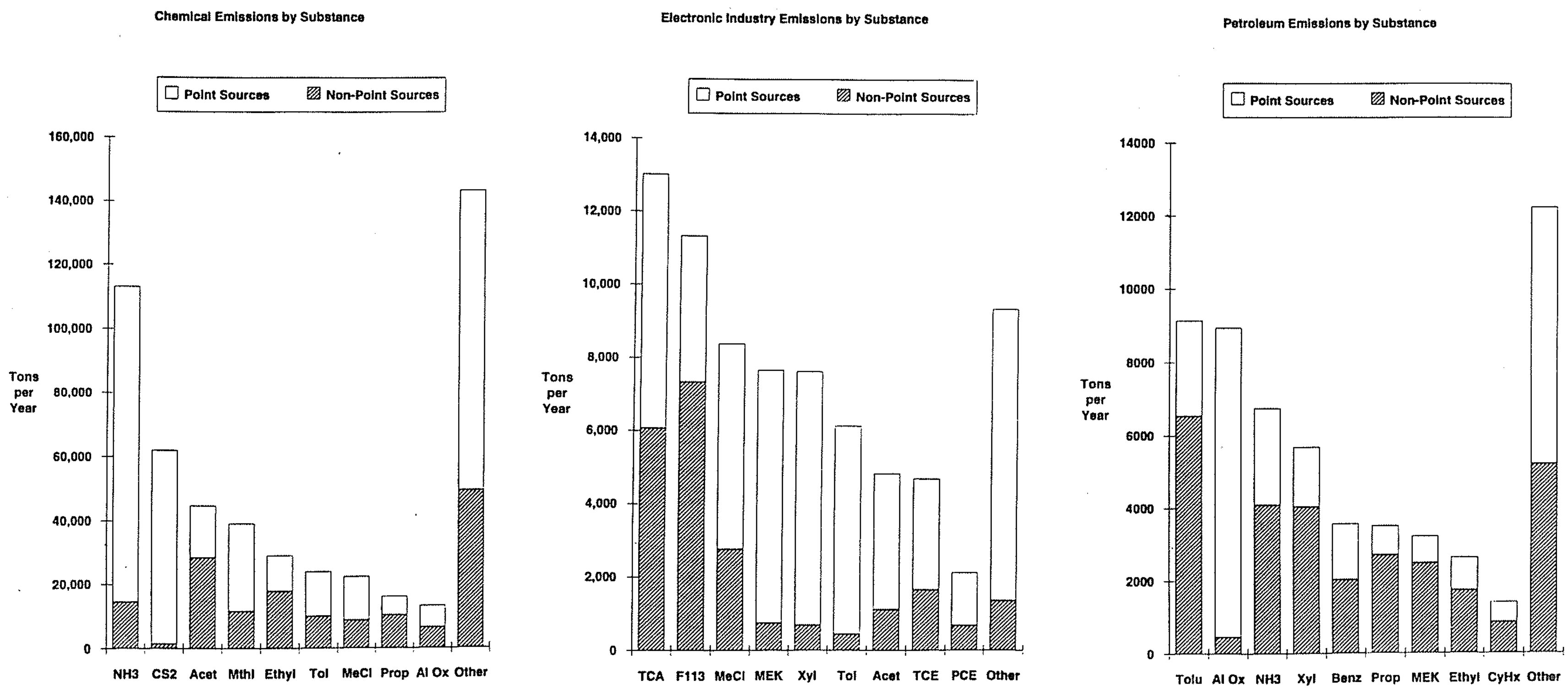


Figure 7. Type of Air Emissions for the Major Chemicals Released from Each Industry

due to the tendency of overestimates to be offset by other underestimates. Nonetheless, since data does not exist to compare the report estimates with new measurements, doubt remains about the true accuracy of the TRI data.

Few companies provided waste minimization data for the 1987 database. In later years, however, more companies have provided this information, perhaps because of publicity considerations (Lave and Omenn, 1989; Philbin, 1990). Even with complete information, unfortunately, the current TRI report format provides only a limited level of detail. This project was unable to examine the subsequent years of data, due to numerous delays in releasing the database in an affordable, accessable format (CD-ROM). Due to these delays, the TRI cannot currently be considered as an up-to-date source of information for construction planning.

References

Associated Press (Washington D.C.) (1990). "Reported Toxic Emissions Cuts Phony, Group Says", *Mercury News* (Aug. 17):8A.

Baram, M.S., P.S. Dillon and B. Ruffle (1990). "Managing Chemical Risks: Corporate Response to SARA Title III", published by The Center for Environmental Management, Tufts University, Medford, MA.

Chemical Manufacturer's Association (1989). "Title III: One Year Later, Plant Manager Interviews", June 1989.

Fisher, P.W., R.M. Currie and R.J. Churchill (1988). "Sara Title III, Section 313 — Looking Ahead", *JAPCA* 38(11):1376–1379.

Hanson, D. (1990). "Hazardous Chemicals Emission Data Drawing More Local, State Attention", *Chem. & Eng. News* (Sept. 3):23–24.

Hanson, D. (1992). "Toxic Release Inventory Data Show Steady Drop in Emissions", *Chem. & Eng. News* (June 15):13–14.

Krieger, J. (1989). "Hazardous Waste Management Database Starts to Take Shape", *Chem. & Eng. News* (Feb. 6):19–21.

Lave, L.B., and G.S. Omenn (1989). "Tons of Toxic Chemicals Above", *The Christian Science Monitor*, Tuesday April 11, 1989, p. 19.

Nowak, G. (1990). Personal communication, Office of Toxic Substances, U.S. Environmental Protection Agency.

Philbin, K. (1990). "SARA Title III Brings Chemical Risk Issues to the Forefront of Corporate Management", *CEM Report* 9:1.

Santa Clara County Manufacturing Group (1990a). "Environmental Releases for 1987, 1988 and 1989 Based on Data Submitted Under Superfund Amendments and Reauthorization Act (SARA), Title III — Community Right to Know", July 18, 1990, 18 pp.

Santa Clara County Manufacturing Group (1990b). "Background Information for Report. SARA Title III", July 18, 1990.

Sarokin, D. (1988). "The Toxics Release Inventory", *Environ. Sci. Technol.* 22(6):616–618.

Thurm, S. (1990). "Chemical Releases Cut, Study Says", *Mercury News* (July 18):2B.

U.S. Environmental Protection Agency (1990b). "Assessment of Data Quality in the 1987 TRI: Site Visit Program", prepared by Radian Corporation for the EPA Office of Toxic Substances, March 27, 1990.

U.S. Environmental Protection Agency (1990a). "Toxic Release Inventory CD-ROM User Guide", published by SearchExpress, Birmingham, AL.

U.S. Environmental Protection Agency (1989a). "The Toxics-Release Inventory: A National Perspective", report no. EPA 560/4-89-005.

U.S. Environmental Protection Agency (1989b). "The Toxics-Release Inventory. Executive Summary", report no. EPA 560/4-89-006.

Appendix A: Methodology for Recovery of 1987 Database

For the 1987 database, the software provided with the CD-ROM does not allow output of user-selected fields into user-specified reports or into a tabular format (USEPA, 1990a). This CD-ROM software also does not support arithmetic calculations. Thus, in order to easily analyze the data, it is necessary to process the data and export it to a spreadsheet, database, or other application.

Although the CD ROM was intended for use with IBM personal computers, the ASCII format is compatible with Macintosh computers. Because of the greater ease of storage device handling, a Macintosh was used for data processing.

Using Microsoft Word the format of the database was determined. Each type of data field is identified by a two-letter code. An application prepared in Think C (given in Appendix C) read the database and collected the data fields of interest one record at a time. This application wrote each record to a Macintosh text file if the company was in the correct industrial groups and if the company filed air emission data (see Figure A-1). Thus, the database was reduced to a reasonable size by filtering out companies in other industrial groups and removing zero entries. The resulting data file was stored as a standard tab-formatted Macintosh text file with return characters at the end of each record. Files of this type can be opened by almost all Macintosh applications for further analysis.

For this project, the output files were opened in 4th Dimension, a database application. 4th Dimension was used to sort the records by chemical type and to produce tables of emissions with totals for each chemical.

Summary tables from 4th Dimension were exported to Microsoft Excel to perform calculations and to prepare graphics for reports and presentations. A flow chart representing this process is shown in Figure A-2.

In summary, the CD-ROM is prepared for IBM PCs with supporting software. Because of the limited capabilities of the supplied software, however, it is advantageous to process the data directly from the CD. Almost any programming environment can be used to gather the ASCII data and translate it to a suitable file format for other applications. Database applications are ideally suited for handling the large files, but spreadsheets can be used to process smaller files.

Figure A-1.

Data Processing Flow Chart

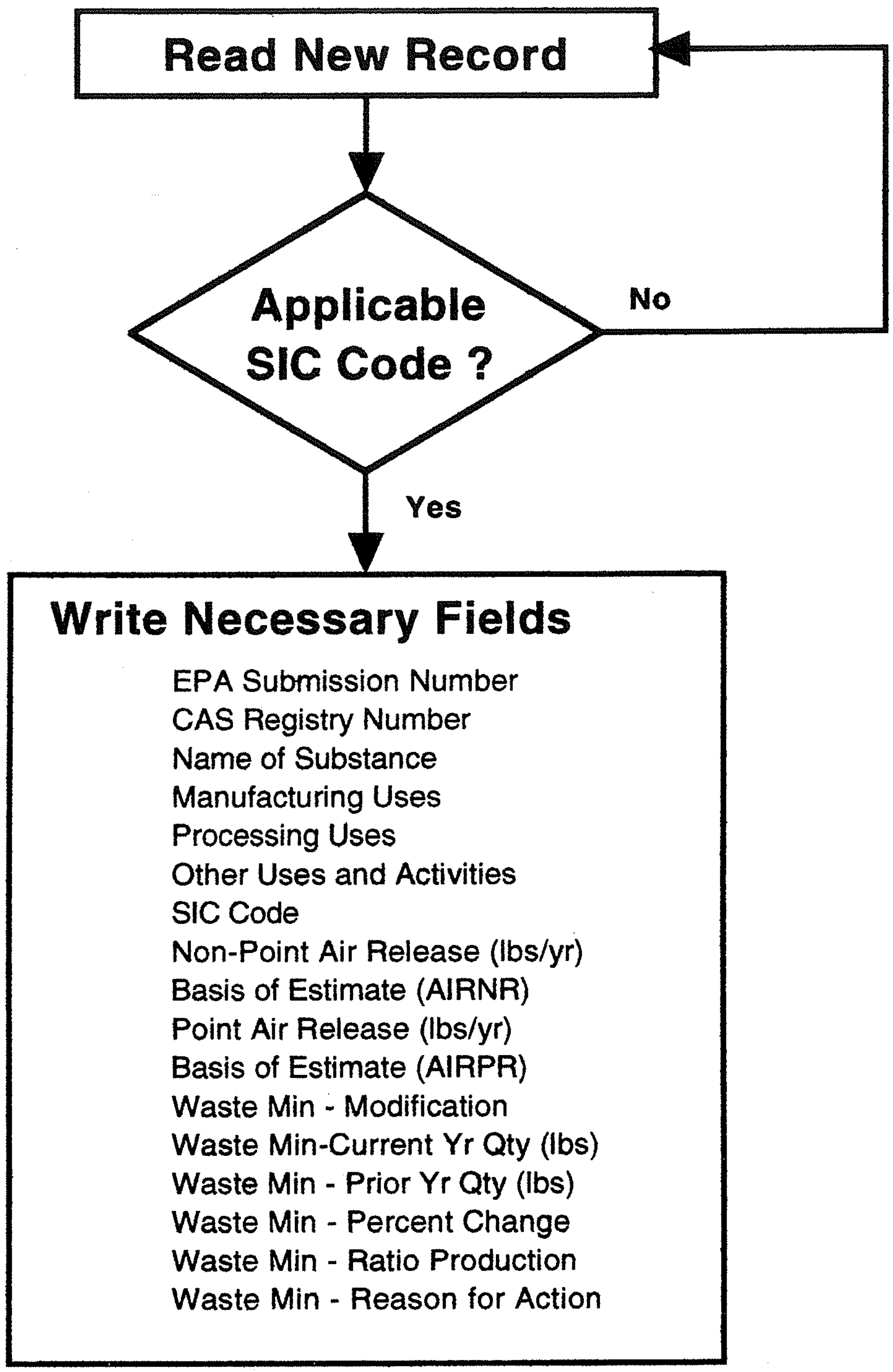
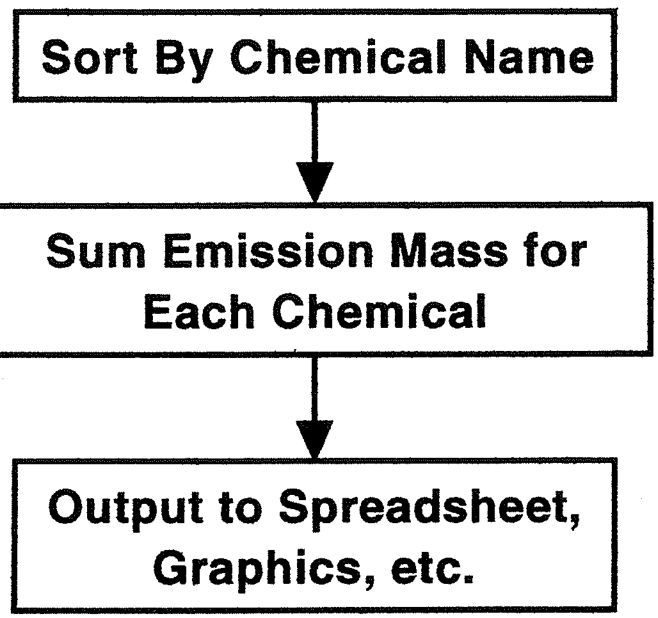


Figure A-2.

Data Processing Flow Chart



Appendix B: Sorting Strategies for Examining Data

For this report, data was sorted first by SIC Code and then by chemical name. A number of other strategies may be used for other purposes (See Table B-I). A simple strategy for use in facility planning is to examine the reports of similar facilities by company name. For example, the names of ethylene producers could be drawn from a chemical industry trade magazine or index. By examining the records for each facility directly in the IBM PC software package, a user can make a list of similar facilities releases. This may seem like a crude approach, but similar procedures are widely used for financial and stock market analysis. The advantage of such a selective approach is that the search can concentrate on companies for which financial data, production capacity, and other information is known from other sources.

A more detailed comparative analysis can be performed by linking data in the TRI with other data using various fields. Dun & Bradstreet numbers for each facility and each facility's parent company can be matched to financial databases. The Chemical Abstract (CAS) number for each chemical can be used to locate EPA Integrated Risk Information Service records. These records contain References Doses and Slope Factors for use in performing short-term and long-term exposure assessments. Facility EPA Identification Numbers can be used to match TRI records with those maintained by state agencies for hazardous waste disposal, permits, and violations. NPDES numbers can be used to match releases to receiving surface streams or watersheds. Geographical data such as latitude and longitude, state, county, zip code, and EPA region can be used to limit a search to a specific region.

Each facility may list up to 3 SIC codes for its operations. The codes designate broad industrial classifications, but there are distinctions within each class. A number of sub-classes divide up the petroleum industry, for example (see Table IV). Within an SIC code, each facility specifies a manufacturing use for the chemical. In addition to company data, the database contains fields for waste treatment and waste minimization methods for each stream. Thus, a user can specify a variety of values for a search to limit the scope of the result.

In analyzing any TRI data, users should remember that each report is an estimate provided by the facility based on existing data. A substantial fraction of estimates will be in error. Thus, aggregates of data should be used whenever possible. Analysis methods should consider these errors. For example, a user should generally calculate median values rather than means for emissions. Furthermore, analyses should be viewed in light of the fact that some companies fail to report releases and therefore skew the low end of the distribution.

Table B-I.

Example Search Types

Search Type	Key Field
Industry or Product Type	SIC Code
Specific Producer	Company Name
Producer Size	Cross Link with D & B Number Maximum Amount on Site Amount of Emissions
Classes of Compounds	Chemical Name CAS Number
Uses of Chemicals	Manufacturing Use SIC Code
Treatment Methods	Waste Stream Treatment Methods
Waste Minimization	Waste Minimization Methods
Geography	City, State, County, Zip Code EPA Region Longitude and Latitude
Other Cross Links	NPDES Number EPA ID Number Census Data

```
#include <stdio.h>                                                    /* include libraries */

#define INFILENAME "Tsunami 50-Q:Work:C Digestion:Examinations:420 sample records"
#define OUTFILENAME "HD20SC Main:testout"
#define MAXSTR 100                                                    /* define files, etc. */

typedef char TEXT[MAXSTR];


FILE *OpenAnInputFile()                                               /* open input files */
{
        FILE *infile;

        if ((infile = fopen(INFILENAME, "r")) > NULL )
                return(infile);
        else {
                printf("\nCouldn't open input file:  '%s'\n", INFILENAME);
                return(NULL);
        }
}

FILE *OpenAnOutputFile()                                              /* open output files */
{
        FILE *outfile;

        if ((outfile = fopen(OUTFILENAME, "w")) > NULL )
                return(outfile);
        else {
                printf("\nCouldn't open output file:  '%s'\n", OUTFILENAME);
                return(NULL);
        }
}



main()
{

        FILE *infile, *outfile;
        TEXT data, sn, ns, np, bn, pa, bp, wm, wy, wp, wc, wx, wa;
        int n, p, o, end, wmfilter;
        int filter = 0;

        sn[1] = '\0';                                                 /*clear data*/
        ns[1] = '\0';
        np[1] = '\0';
        bn[1] = '\0';
        pa[1] = '\0';
        bp[1] = '\0';
        wm[1] = '\0';
        wy[1] = '\0';
        wp[1] = '\0';
```

```
wc[1] = '\0';
wx[1] = '\0';
wa[1] = '\0';

wm[4] = '\0';
wy[4] = '\0';
wp[4] = '\0';
wc[4] = '\0';
wx[4] = '\0';
wa[4] = '\0';


if ((infile = OpenAnInputFile()) != NULL) {                                        /*open in*/
        if ((outfile = OpenAnOutputFile()) != NULL) {                              /*open out*/
                do {
                        for (n = 0; n < MAXSTR; n++) {
                                data[n] = '0';
/*clear data*/
                        }
                        end = 0;
                        for (n = 0; n < MAXSTR && end == 0; n++) {                 /*read to \n*/
                                data[n] = fgetc(infile);
                                if (data[n] == '\n') {
                                        data[n] = '\0';
                                        end = 1;
                                }
                        }

                        if (data[2] == 'S' && data[3] == 'N') {                    /*finds sn*/
                                if (filter == 1) {
                                        fputc('\n', outfile);
/*start new record*/

                                        end = 0;
                                        for (p = 5; p < MAXSTR && end == 0; p++) {   /*write sn*/
                                                fputc(sn[p], outfile);
                                                if (sn[p+1] == '\0') {
                                                        fputc('\t', outfile);
                                                        end = 1;
                                                }
                                        }

                                        end = 0;
                                        for (p = 4; p < MAXSTR && end == 0; p++) {   /*write ns*/
                                                if (ns[p] == '\0') {
                                                        fputc('\t', outfile);
                                                        end = 1;
                                                }
                                                else {
                                                        fputc(ns[p], outfile);
                                                }
```

```
}

end = 0;
for (p = 4; p < MAXSTR && end == 0; p++) {        /*write np*/
        fputc(np[p], outfile);
        if (np[p+1] == '\0') {
                fputc('\t', outfile);
                end = 1;
        }
}

end = 0;
for (p = 4; p < MAXSTR && end == 0; p++) {        /*write bn*/
        fputc(bn[p], outfile);
        if (bn[p+1] == '\0') {
                fputc('\t', outfile);
                end = 1;
        }
}

end = 0;
for (p = 4; p < MAXSTR && end == 0; p++) {        /*write pa*/
        fputc(pa[p], outfile);
        if (pa[p+1] == '\0') {
                fputc('\t', outfile);
                end = 1;
        }
}

end = 0;
for (p = 4; p < MAXSTR && end == 0; p++) {        /*write bp*/
        fputc(bp[p], outfile);
        if (bp[p+1] == '\0') {
                fputc('\t', outfile);
                end = 1;
        }
}

end = 0;
for (p = 4; p < MAXSTR && end == 0; p++) {        /*write wm*/
        if (wm[p] == '\0') {
                fputc('\t', outfile);
                end = 1;
        }
        else {
                fputc(wm[p], outfile);
        }
}

end = 0;
for (p = 4; p < MAXSTR && end == 0; p++) {        /*write wy*/
        if (wy[p] == '\0') {
```

```
                        fputc('\t', outfile);
                        end = 1;
                }
                else {
                        fputc(wy[p], outfile);
                }
        }

        end = 0;
        for (p = 4; p < MAXSTR && end == 0; p++) {      /*write wp*/
                if (wp[p] == '\0') {
                        fputc('\t', outfile);
                        end = 1;
                }
                else {
                        fputc(wp[p], outfile);
                }
        }

        end = 0;
        for (p = 4; p < MAXSTR && end == 0; p++) {      /*write wc*/
                if (wc[p] == '\0') {
                        fputc('\t', outfile);
                        end = 1;
                }
                else {
                        fputc(wc[p], outfile);
                }
        }

        end = 0;
        for (p = 4; p < MAXSTR && end == 0; p++) {      /*write wx*/
                if (wx[p] == '\0') {
                        fputc('\t', outfile);
                        end = 1;
                }
                else {
                        fputc(wx[p], outfile);
                }
        }

        end = 0;
        for (p = 4; p < MAXSTR && end == 0; p++) {      /*write wa*/
                if (wa[p] == '\0') {
                        fputc('\t', outfile);
                        end = 1;
                }
                else {
                        fputc(wa[p], outfile);
                }
        }
}
```

```
                        filter = 0;
/*reset filter*/
                        wmfilter = 0;
                        wm[4] = '\0';
/*flag w items*/
                        wy[4] = '\0';
                        wp[4] = '\0';
                        wc[4] = '\0';
                        wx[4] = '\0';
                        wa[4] = '\0';

                        for (p = 5; p < MAXSTR; p++)                    /*save sn*/
                                sn[p] = data[p];

                }

                else if (data[1] == 'N' && data[2] == 'S')              /*save ns*/
                        for (p = 4; p < MAXSTR; p++)
                                ns[p] = data[p];

                else if (data[1] == 'S' && data[2] == 'C') {            /*test SIC*/
                        if (data[4] == '2' && data[5] == '8')
                                filter = 1;
                }

                else if (data[1] == 'N' && data[2] == 'P')              /*save np*/
                        for (p = 4; p < MAXSTR; p++)
                                np[p] = data[p];

                else if (data[1] == 'B' && data[2] == 'N')              /*save bn*/
                        for (p = 4; p < MAXSTR; p++)
                                bn[p] = data[p];

                else if (data[1] == 'P' && data[2] == 'A')              /*save pa*/
                        for (p = 4; p < MAXSTR; p++)
                                pa[p] = data[p];

                else if (data[1] == 'B' && data[2] == 'P')              /*save bp*/
                        for (p = 4; p < MAXSTR; p++)
                                bp[p] = data[p];

                else if (data[1] == 'G' && data[2] == 'W') {            /*test gw for air*/
                        if (data[4] == 'A') {
                                wmfilter = 1;
                        }
                        else {
                                wmfilter = 0;
                        }
                }
```

```
                    else if (data[1] == 'W' && data[2] == 'M' && wmfilter == 1)
    /*save wm*/
                            for (p = 4; p < MAXSTR; p++)
                                    wm[p] = data[p];

                    else if (data[1] == 'W' && data[2] == 'Y' && wmfilter == 1)
    /*save wy*/
                            for (p = 4; p < MAXSTR; p++)
                                    wy[p] = data[p];

                    else if (data[1] == 'W' && data[2] == 'P' && wmfilter == 1)
    /*save wp*/
                            for (p = 4; p < MAXSTR; p++)
                                    wp[p] = data[p];

                    else if (data[1] == 'W' && data[2] == 'C' && wmfilter == 1)
    /*save wc*/
                            for (p = 4; p < MAXSTR; p++)
                                    wc[p] = data[p];

                    else if (data[1] == 'W' && data[2] == 'X' && wmfilter == 1)
    /*save wx*/
                            for (p = 4; p < MAXSTR; p++)
                                    wx[p] = data[p];

                    else if (data[1] == 'W' && data[2] == 'A' && wmfilter == 1)
    /*save wa*/
                            for (p = 4; p < MAXSTR; p++)
                                    wa[p] = data[p];


                }
                while (data[n-1] != EOF);
        }
    }

    fclose(outfile);
    fclose(infile);
}
```