EPA/600/A-96/116

Methods for Improving Emissions Estimates

H. L. Waters, V. E. McCormick, J. G. Cleland, and J. R. Youngberg
Research Triangle Institute
Research Triangle Park, NC 27709-2194

P. J. Chappell
U. S. Environmental Protection Agency
Air and Energy Engineering Research Laboratory
Research Triangle Park, NC 27711

ABSTRACT

The Environmental Protection Agency (EPA) is investigating ways to improve methods for estimating volatile organic compound (VOC) emissions from area sources. Using the automobile refinishing industry for a detailed area source case study, the Research Triangle Institute (RTI) and the EPA's Air and Energy Engineering Research Laboratory (AEERL) are developing an emission estimation method that utilizes both advanced computational techniques and updated, comprehensive, emissions-related information. This development includes a thorough characterization of the area source industry, an analysis of current emission estimation methods, the development and execution of a nationwide industry activity survey, and a compilation and analysis of the survey results and other explanatory variables. Results are to be captured in a personal computer-based VOC emissions estimation system called VOCEES. VOCEES is a dual-use tool, the users of which can both prepare VOC emissions inventories and analyze the impact of numerous factors on emissions. This methodology and VOCEES are readily extendible to other area sources.

INTRODUCTION

Stationary sources of pollutant emissions are designated as either point sources or area sources. Whereas point sources are inventoried on an individual basis, area source emissions emanate from processes, activities, or businesses that are too small or too numerous to be practically tracked as individual emission sources. The distinction between point and area sources is an annual emissions cutoff, such as 10 tons (9090 kg) of volatile organic compounds (VOCs) per year per source. The U.S. Environmental Protection Agency's (EPA's) Air and Energy Engineering Research Laboratory (AEERL) has initiated an effort to develop appropriate area source emission estimation methodologies for solvent area source categories using available technologies and to evaluate long-term informational needs. The Research Triangle Institute (RTI) and AEERL are working together to establish an emission estimation methodology development process that will result in accurate emission estimate methods for all area sources. Results of this work will either validate existing area source emission estimation methods or recommend alternative methods.

The required components of an emission estimation methodology are: 1) the calculation of the emission estimates (e.g., using emission and activity factors); 2) the temporal and spatial allocation of the emissions; 3) the validation of the emission estimates, and 4) the speciation of the emissions. The work presented in this paper concentrates on the first three components. Tools are available from EPA that can be used to develop a VOC inventory grouped into reactivity classes suitable for modeling [1]. The criteria used in developing an emission estimation methodology include: reasonable accuracy and cost; dynamic and robust behavior; use of readily available information; and ease of use.

Why New Methods?

This work has been initiated for two reasons: 1) the current emission estimation methods require further development to improve accuracy and ease of use; and 2) innovative tools have emerged which may contribute significantly to improving estimation methods. Among the issues and uncertainties are:

- Most solvents used by any one solvent area source category are also used by other industries, preventing use of solvent manufacturer production figures alone for estimating emissions from one solvent source category.
- There are a large number of very different solvent area sources categories. Therefore, reasonably accurate emission estimates require area-source-category-specific estimation methods.
- Current methods endorse the use of data sources whose published data are from 2 to 5 years old and may misrepresent a significant segment of the area source industry. For example, one-person establishments are a significant percentage of the estimated 65,000 automobile refinishing establishments in the U.S. Since most federal information sources are derived from business payroll records, there is often no record of these businesses in the endorsed information sources. Also, data at the county level for certain Standard Industrial Classification (SIC) codes are often not disclosed as policy.
- Current methods do not consider the dynamic factors that impact area sources such as local economics, changing technology, and regulatory influence. Nor do they consider human behavior and consumption patterns. For example, it is difficult to estimate automobile refinishing emissions from insurance payments because people are more likely to "pocket" insurance payments for vehicle repair if they have more pressing expenses.

With the emergence of new technologies, such as innovative computational techniques, often comes the opportunity to improve on existing methods. Part of the ongoing work being sponsored by AEERL is to investigate the appropriateness of applying advanced, inference-based computational intelligence techniques, such as neural networks, fuzzy logic, and genetic algorithms, to emissions estimation. These techniques provide the ability to determine and utilize relationships between two domains (e.g., industry variables and emission levels) without developing traditional mathematical models. They also provide the ability to incorporate expertise expressed in inexact, English-like terms, such as *big, small, fast,* or *slow*. Also, since all emissions have a geographic component in that each emission has a source, it is helpful to capture and represent emissions-related data using a geographic information system (GIS). A GIS supports the organization, analysis, and visualization of data by their geographic orientation. This is a powerful tool when determining the characteristics of a geographic region, such as personal income levels, that may impact emission levels.

Relevant Work By Others

Most EPA-endorsed methods for estimating solvent emissions from area sources have been derived from a methodology developed as part of the National Emissions Data System (NEDS) [2] on sources of airborne pollutants. This methodology determines national solvent emissions through material balance, allocating emissions to states and counties using *emission factors,* which are based on source-specific emission measurements as a function of source activity levels. The NEDS area source emission estimation method has been used in several emission inventory efforts. These efforts include the 1985 National Acid Precipitation Assessment Program (NAPAP) emissions inventory prepared by EPA [3], the Regional Ozone Modeling for Northeast Transport (ROMNET), and the Area and Mobile Source (AMS) Subsystem of the Aerometric Information Retrieval System (AIRS) [4], which will replace NEDS.

There have also been detailed studies performed on individual area sources in order to provide guidance on area source emission control technology. Methods for estimating emissions are often included in these studies. A 1988 study of VOC emissions from automobile refinishing area sources uses a rather complex calculation dependent upon the final thickness of applied coatings, in thousandths of an inch, while categorizing the entire industry into three types of establishments -small, medium, and volume shops [5]. This approach has been criticized as being inappropriate for *an industry* which has a *large* variance in operating characteristics [6].

METHODOLOGY DEVELOPMENT PROCESS

The goal of this work is to establish an emissions estimation methodology development process that can be applied to all solvent area source categories. The proposed process is shown in Figure 1 and includes the following functions:

1) Select the area source category for which the emission estimation method is to be developed.
2) Thoroughly characterize the chosen area source category so as to fully understand the issues affecting emission estimation for that category.
3) Study all methods currently used for estimating the category's emissions, so as to understand their strengths and weaknesses.
4) Develop an *explanatory variables* database -- data which directly or indirectly affect the area source category operations and assist in explaining its emission levels and fluctuations.
5) Assemble a tool set containing statistical tools, computational intelligence tools,

and geographic information system tools for analyzing and relating assembled data.

6) Conduct the nationwide area source survey using a questionnaire reviewed and endorsed by industry experts and by trade associations. Compile and analyze the survey results.
7) Select readily available explanatory variables that most significantly impact emission levels. Structure and complete the method which uses these variables to estimate emissions from the area source category, and which satisfies the specified criteria.

Area Source Selection

VOC-emitting area sources include gasoline distribution losses, stationary source solvent evaporation, bioprocess emission sources, catastrophic or accidental releases, solid waste incineration, and small stationary source fossil fuel use. Stationary solvent evaporation sources include dry cleaning, surface cleaning, surface coating, graphic arts, asphalt paving, pesticide application, commercial or consumer solvent use, and synthetic organic chemical storage tanks. Automobile refinishing, a surface coating category, was chosen for study because it is representative of VOC emission area source categories in terms of relative environmental significance, national prevalence, and an accessible information base representative of such sources.

Industry Characterization

A detailed characterization of the automobile refinishing industry was conducted in order to fully consider the issues and variables associated with the industry's VOC emissions. The number and employment of automobile refinishing establishments vary across information sources. According to industry representatives, the impact of technological and regulatory changes, along with economic factors, has provoked a steady decrease *in number of automobile refinishing establishments* nationwide over the past 20 years [5][7]. They attribute this reduction in establishments and paint use to: 1) improved vehicle safety and more stringent enforcement of traffic laws; 2) more corrosion-resistant finishes; 3) smaller cars with less surface area to paint per repair; and 4) more efficient spray guns which use less paint per job. Other information sources, however, have reported a steady, yet gradual increase in number of establishments. The difference may be explained in part by undocumented "backyard" establishments, which may represent an additional 25% to 40% of the total number of establishments. These "unseen shops" are potentially a large source of VOC emissions since they are less likely to use emissions control technology or comply with laws and regulations, and more likely to use high-VOC coatings.

With the passage of the Clean Air Act Amendments, and other EPA initiatives, new regulations for the automobile refinishing industry are emerging. Regulations requiring use of low-VOC coatings have forced the industry away from lacquer and enamel coatings and toward urethanes. Additional emission control methods include:

- Use of enclosed equipment cleaning devices that support solvent reuse.

- Increased paint (to surface) transfer efficiency through use of high-volume, low-pressure (HVLP) spray guns.
- Addition to paint spray booths of emission controls which use the following techniques: thermal incineration, catalytic incineration, and carbon adsorption. All three methods may be too expensive to be considered as universal, reasonable, control techniques.
- Better shop operations management, including: use of tight fitting containers; reduction of spills; rigid control of inventory; tracking of worker use rates; mixture of paint to need; provision of operator training; use of proper cleanup methods; and use of in-house or leased solvent recycling.

Current Methods Review

Methods for estimating solvent emissions from area sources are based on use of emission factors. Emission factors "are developed from only a limited sampling of the emissions source population for any given category, and the values reported are an average of those limited samples and may not be statistically representative of the population" [4]. The basic approach in estimating emissions is derived from a simple calculation that requires an estimate of an activity level, an emissions factor relating emissions to activities, and, if the source has a pollution control device, a control factor:

Emissions = activity level x emission factor x control factor

Emission factors can be found in a number of EPA references. Table 1 contains the emission and activity factors for automobile refinishing from *AP-42 Compilation of Air Pollutant Emission Factors* [8] and its Supplement D [91, the EPA State Implementation Plan (SIP) guidance document[10], and AIRS/AMS [1 1] [12]. In analyzing these data, note that:

- The AP-42 emission factors and national VOC emissions did not change for more than 10 years. They may not have changed since the AP-42 was first published in 1972.
- The SIP guidance document's *per capita* emission factor is 21% greater than that of the AP-42 document, while the *per employee* emission factor is 48% less than that of the AP-42 document. Some difference may be explained by the 1987-1988 SIC code changes (i.e., using SIC 7532, *Automotive Top and Body Repair and Paint Shops,* instead of SIC 7535, *Automotive Paint Shops).*

Examples of two problem areas are:

1) Difference in Emission Estimates. The two emission factors *(per capita vs. per employee)* seldom produce the same emission estimate, as shown in the example in Figure 2 for 26 California counties. These discrepancies arise from the static nature of emission factors and the different characteristics of individual geographic regions.

2) Missing Data. County Business Patterns [13] exercises disclosure protection for counties where revealing information may disclose details about the operations of individual businesses. In 1990, SIC 7532 employment for 47 of North Carolina's 100 counties was not disclosed. An emission estimation method should utilize activity factors that are available for all areas required to compile emission inventories.

Database Development

The success of any new method to estimate area source emissions is limited by the availability and accessibility of information. The objective is to identify data that are updated at least annually and represent county-level activity. Records should exist for at least the past 5 years and into the foreseeable future. The data must be statistically defensible, representative, "universal" (representative of national distributions and/or variations), and result in more accurate emissions estimates. The data are then related to the primary area source variables of both a) solvent use and b) emissions. These variables may be normalized -- for example, on the basis of per capita, per employee, or per operation.

The database assembled for this study has both geographic or spatial components (e.g., nation, state, county, or city) and temporal components (e.g., year or month). The current data set includes:

- 60 variables over 12 years for the United States
- 25 variables over 12 years for 51 states (including DC)
- 1 1 variables over 12 years for 3126 counties
- 10 variables for 1993 for 64,524 automobile refinishing establishments

Analytical Tool Set Selection

An analytical tool set has been assembled for the purpose of capturing the relationships between explanatory variables and emission levels. This set includes both Computational Intelligence tools and traditional statistical tools. *Computational Intelligence* is a term adopted by the Institute for Electrical and Electronics Engineers (IEEE) for innovative computational techniques that include artificial neural networks, fuzzy logic, and genetic algorithms.

Artificial Neural Networks. An artificial neural network (ANN) is an analysis tool that is modeled after the massively parallel structure of the brain. It simulates a highly interconnected, parallel computational structure with many relatively simple, individual processing elements or neurons. Feed-forward ANN paradigms are capable of learning or extracting a relationship between two domains. ANNs are best applied where there are no known rules or mathematical models that accurately relate the independent and dependent variables of interest, but there is an abundance of data representing the relationship.

Fuzzy Logic Expert System. An expert system is a knowledge-based system that captures human expertise in a specific knowledge domain. A fuzzy logic expert system is being developed to augment estimation of VOC emissions from the automobile refinishing area sources. Fuzzy logic is an approximate reasoning technique used in processing inexact information. While a typical expert system may be thought of as defining "true or false" conditions, fuzzy systems allow for varying degrees of truth, or "shades of gray," more like human reasoning.

For example, if climate is a factor in an area's emission levels, then it could be classified as *dry, moderate,* or *rainy.* The type of area might also be loosely classified into three fuzzy sets: *rural, suburban,* and *urban.* A typical fuzzy rule, based on expert opinion, may be expressed as, "IF the climate is *dry* AND the area is *rural* THEN emissions are low." Another rule may state "IF the climate is *wet* AND the area is *urban* then emissions are *high.*" These rules describe the increased likelihood of accidents and auto refinishing in a congested area with poor weather and vice versa. The fuzzy system uses the degree of membership of an input in a given set to determine to what degree the output belongs in any set (e.g., *low, medium, high).* This type of reasoning can augment the emissions prediction based on optimally correlated data.

Genetic Algorithms. Genetic algorithms (GAs) are a class of machine learning search algorithms based on the mechanics of natural selection and natural genetics. GAs combine Darwin's "survival of the fittest" with structured, yet randomized information exchange to form a search algorithm with some innovative flair of human search [141. Created by John Holland at the University of Michigan [15], GAs are robust, general-purpose problem solvers especially suited for optimization and classification. GAs can develop better solutions from thousands of choices more effectively than other techniques for a host of problems.

Use of GAs in the emissions problem will be confined to methodology development. Examples of using GAs would be in the design and training of neural networks or development of optimum fuzzy logic membership functions. To properly optimize a system, the GA must know what constitutes good performance. In estimating emissions, this would be a knowledge of the VOC emission level for a corresponding set of explanatory variables resulting from the nationwide industry survey.

Statistics. Simple statistical techniques have been employed as the preliminary step for analyzing variable relationships and selecting a priority explanatory variable set. Multivariate regression analyses (linear correlations) have been applied and correlation indicators derived. There is a high correlation between the automobile refinishing industry and several of the explanatory variables included in the database. The variables have been analyzed by examining their distribution across different regions of the country, their change over time, and their ratio one to another.

Geographic Information System. Emissions are characterized by their levels and by their distribution in space and time. A geographic information system (GIS) is being used to assign emissions-related explanatory variable values, such as number of employees and annual sales, to the actual location of an individual business. These values can then be aggregated to combine values within a ZIP Code, county, non-attainment area, metropolitan statistical area (MSA), state, EPA Region, or nation. For State Implementation Plan (SIP) emission inventory purposes, for example, area source emissions would be aggregated to county and to non-attainment area boundaries.

Nationwide Industry Survey

One of the most important steps in the emission estimation method development process is the nationwide survey of the chosen area sources. The Automobile Refinishing Solvent Use Survey (ARSUS) is the first survey of its kind. It is also essential for validation of the new emission estimation method. Important features of ARSUS are:

- The survey is designed based on a detailed knowledge of the industry built on numerous contacts with shop operators, paint manufacturers, and association representatives.
- Results will be statistically defensible, based on random probability sets, with the results represented by statistically correct accuracy estimations and confidence levels.

The map shown in Figure 3 represents the scope of coverage that the 5900 samples will provide. ARSUS includes local-area intensive surveys of six high-population areas. The survey data are divided into two independent sets with probability proportional to population, each containing 30 areas or Primary Sample Units (PSUs). One set is assigned for estimating model parameters (developing the method), while the second set is for developing a comparison variable (evaluating and validating the method). The estimated accuracies of solvent use data from the surveys are shown in Table 2.

Information services and computerized "Yellow Pages" are used to retrieve information abstracts on firms with a range of SIC codes for automobile refinishing in the counties selected. This file is stratified by SIC and number of employees, with a probability sample of organizations in each of the sample PSUS. A detailed file for each sample organization is then retrieved with the names and addresses for the mail survey and auxiliary data for use in the final statistical analyses.

The survey combines mail and telephone contacts to maximize response rates, minimize respondent burden, and complete the data collection efficiently. Survey results are entered using bar-coded identification labels and event codes which indicate the pending or final status. Results are to be tabulated in data files for analysis and inclusion in the developed method.

Method Implementation

The final step of the emission estimation method development consists of bringing together all explanatory variable and survey data, analyzing them, and processing them with the assembled tool set. More than one computational intelligence tool is needed to meet the specified method development criteria. The tool set's statistical components are being used to identify the best data for use in development. Also, fuzzy set clustering is being used during preprocessing to group the geographic regions (e.g., counties) with similar characteristics for additional analysis. The preprocessed survey results and explanatory variable data can be presented to a neural network as illustrated in Figure 4. Initially, half of the available data will be used to train the neural network while the other half will be used to evaluate that which the network has learned. Genetic algorithms will be used during this phase of development to identify the optimum neural network architecture, learning algorithm, and other critical parameters.

Finally, the entire database will be used to train the network, to continue to identify and minimize the input variable set which best estimates emission levels. Analysis of the trained network will identify the input variables which are most important in estimating emission levels, reducing the number of input variables required. A sensitivity analysis will also be performed to determine the error introduced by this minimized artificial neural network.

A personal-computer-based VOC emission estimation system (VOCEES) will automate the developed method. The components of the system are: 1) the essential explanatory variable database; 2) basic algorithms and possibly an artificial neural network-based computational component; 3) the supplemental fuzzy logic expert system; and 4) the GIS-based user interface.

PRELIMINARY RESULTS AND CONCLUSIONS

Practically all of the emission estimation method case study is complete. The hardware and software implementation of the method is called the Volatile Organic Compound Emission Estimation System or VOCEES. The system is a dual-use tool, enabling the user both to prepare standard VOC emission inventories for geographic regions of various sizes, and also to analyze how emission levels are influenced by various factors. VOCEES currently combines the emission-factor-based VOC estimation with the predicted impact of current and anticipated VOC-limiting regulations. The system supports tabular, bar graph, and map formatted analysis and reporting facilities. An example is presented in Figure 6.

Of the seven functions specified by the proposed development process shown in Figure 1, work performed to date has completed the first five functions and the preparation and review of the survey questionnaire as described in the sixth function. It is anticipated that all seven steps will be completed for the automobile refinishing case study and the process will be extended to other area source categories. No technical barriers to this process have been identified, and the early results demonstrate a more efficient and cost-effective emissions estimation methodology.

DISCLAIMER

This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication.

REFERENCES

1. *VOCIPM Speciation Data System Documentation and User's Guide, Version 1.32a;* EPA-450/291-002; U. S. Environmental Protection Agency: Research Triangle Park, November 1990.

2. Myers, J.P., Benesh, F.; *Methodologies for Countywide Estimation of Coal, Gas, and Organic Solvent Consumption;* EPA-450/3-75-086 (NTIS PB259909); U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, December 1975.

3. Demmy, J., Tax W., Wam, T.; *Area Source Documentation for the 1985 National Acid Precipitation Assessment Program Inventory;* EPA-600/8-88-106 (NTIS'PB89-151427); U.S. Environmental Protection Agency, Air and Energy Engineering Research Laboratory: Research Triangle Park, December 1988.

4. Seinfeld, J., Atkinson, R., Berglund, R., et al.; *Rethinking the Ozone Problem in Urban and Regional Air Pollution;* National Academy Press: Washington, 1991.

5. Athey, C., Hester, C., McLaughlin, M., et al.; *Reduction of Volatile Organic Compound Emissions from Automobile Refinishing;* EPA-450/3-88-009 (NTIS PB89-148282); U.S. Environmental Protection Agency, Control Technology Center: Research Triangle Park, October 1988.

6. *National Air Pollution Control Techniques Advisory Committee, Meeting Minutes, Volume 2;* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, November 19-21, 1991; p 1008.

7. M. Hunke, Carstar Automotive, Inc., Shawnee Mission, KS, personal communication, 1992.

8. *Compilation of Air Pollution Emission Factors, Volume I.- Stationary Point and Area Sources, AP-42, Fourth Edition* (GPO 055-000-00251-7); U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, September 1985; pp. 2, 4.2.1-1.

9. Joyner, W.; *Supplement D to Compilation of Air Pollution Emission Factors, Volume I.Stationary Point and Area Sources, AP-42, Fourth Edition* (GPO 055-00-00391-2); U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, September 199 1; pp 2, 4.2. 1 - 1.

10. *Procedures for the Preparation of Emission Inventories for Carbon Monoxide and Precursors Of Ozone, Volume I.- General Guidance For Stationary Sources;* EPA-450/4-91-016 (NTIS PB92112168); U.S. Environmental Protection Agency: Research Triangle Park, May 1991; pp. 4-24, 4-47.

11. Kimbrough, E.; *Documentation of AIRS AMS National Methodologies;* EPA-600-R-92-001 (NTIS PB92-132869); U.S. Environmental Protection Agency, Air and Energy

Engineering Research Laboratory: Research Triangle Park, January 1992.

12. *Aerometric Information Retrieval System (AIRS): "Short List" of AMS SCCs and Emission Factors;* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, July 1992.

13. *County Business Patterns;* U.S. Department of Commerce, Bureau of the Census: Washington; Annual Publication.

14. Goldberg, D.E.; *Genetic Algorithms in Search Optimization and Machine Leaming;* AddisonWesley Publishing Company: Reading, 1989.

15. Holland, J.H.; *Adaptation in Natural and Artificial Systems;* University of Michigan Press: Ann Arbor, 1973.

Table 1. Recommended emission factors for the automobile refinishing area source.

Reference	National VOC emissions (tons per year)	VOC emission factors (pounds/year)		U.S. population (millions)	SIC code	Number of U.S. automobile refinishing employees	Year derived
		Per capita	Per employee				
AP-42	199,000	1.9	5200	209.5 (est'd)		76,500 (est'd)	1972 (est'd)
AP-42 Supplement D	199,000	1.9	5200	209.5 (est'd)		76,500 (est'd)	1972 (est'd)
SIP Guidance Document	281,00 (est'd)	2.3	3519	245.7	7532	163,000 (est'd)	1989
AIRS / AMS			3519		7535, 7532		

Table 2. Estimated accuracies of solvent use survey data.

Survey	Sample size fielded	Precision estimate	Difference estimate
National	5,900	7%	10%
Six local intensives	3,400	10%	14%

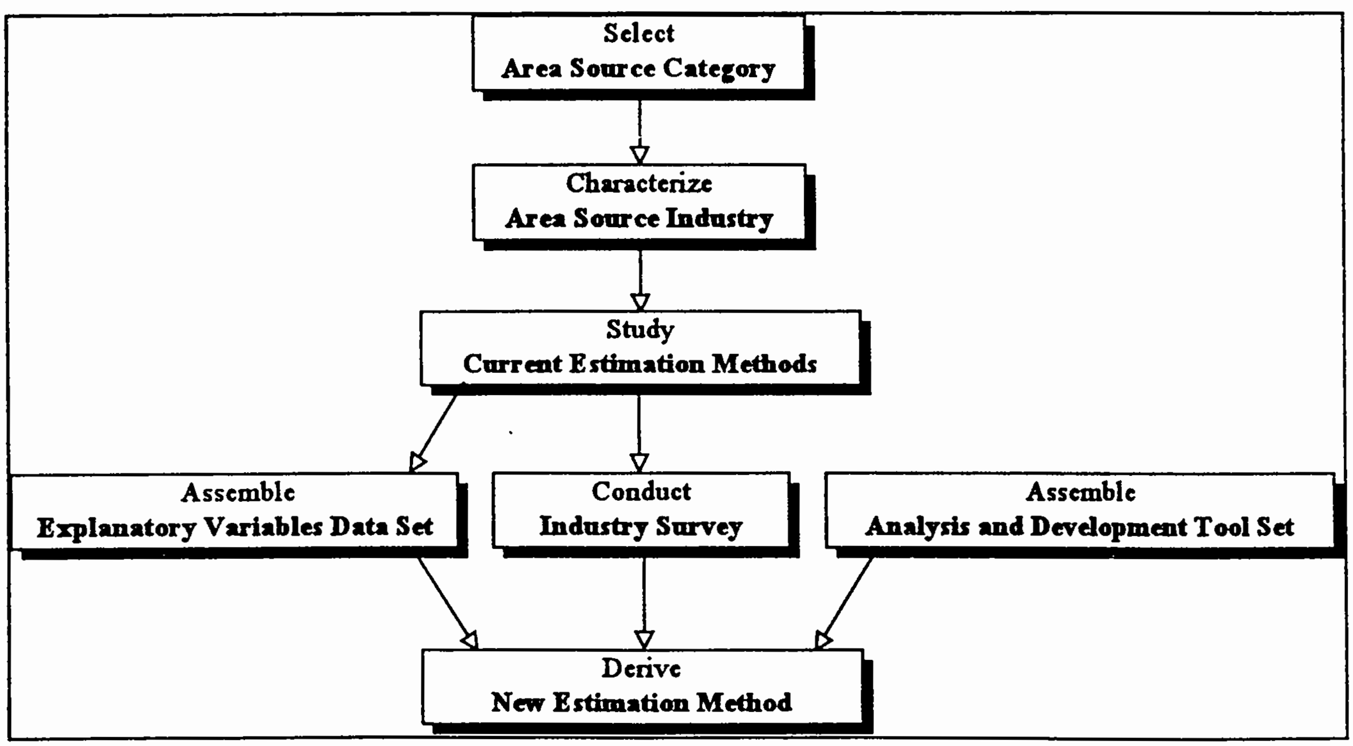

Figure 1. The proposed emission estimation methodology development process.

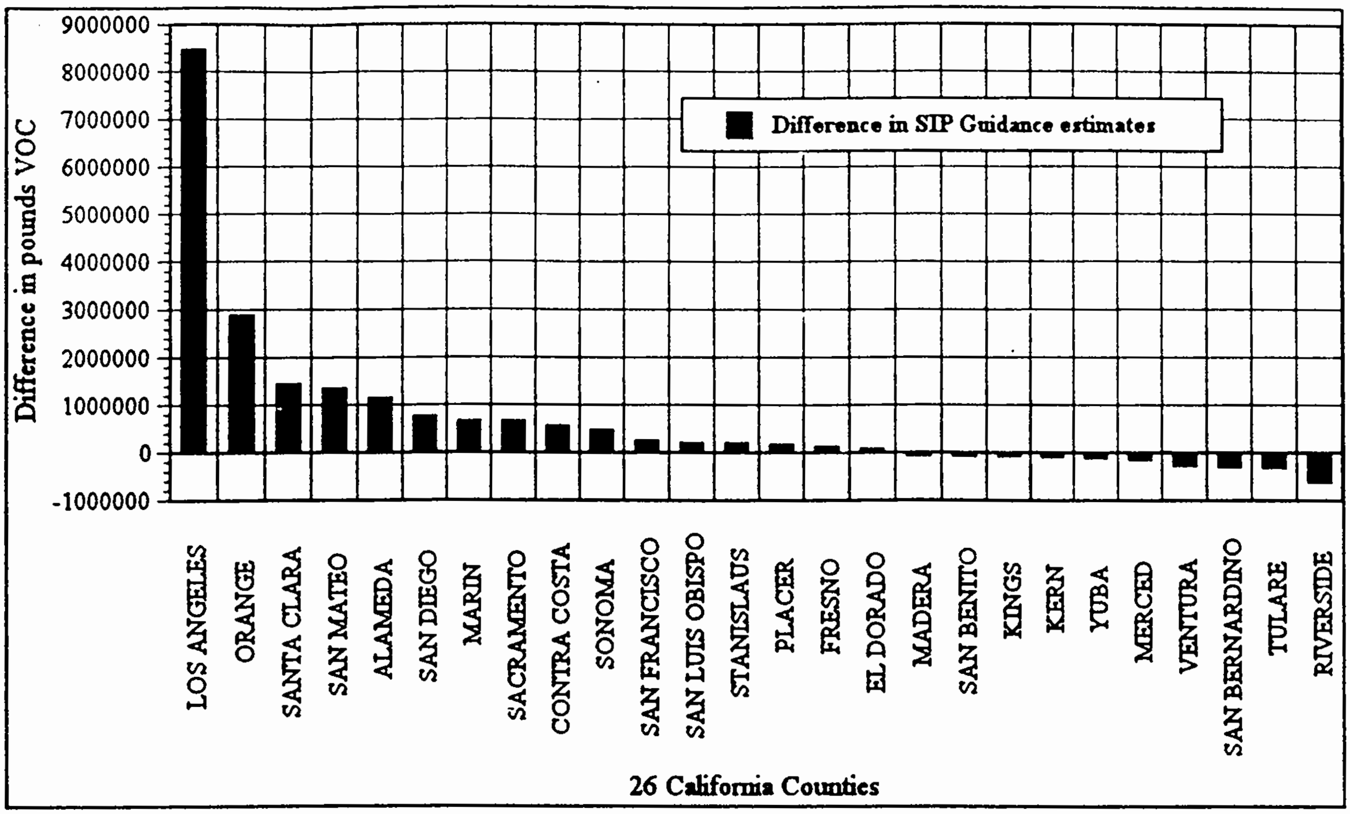

Figure 2. The difference in automobile refinishing VOC emission estimates, using the two SIP Guidance emission factors, *per employee* and *per capita.*

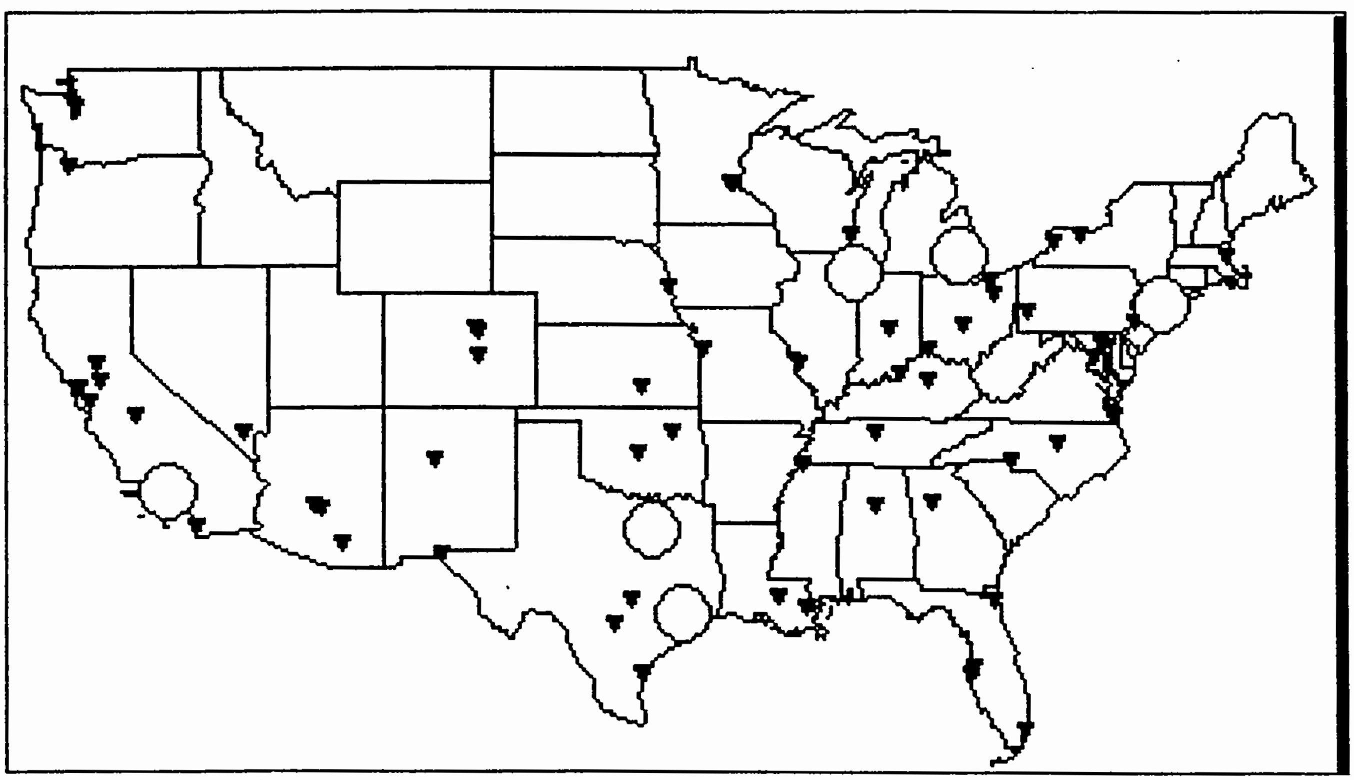

Figure 3. Preliminary targets of nationwide automobile refinishing survey: six highly populated areas and 60 primary sample units (PSUs).

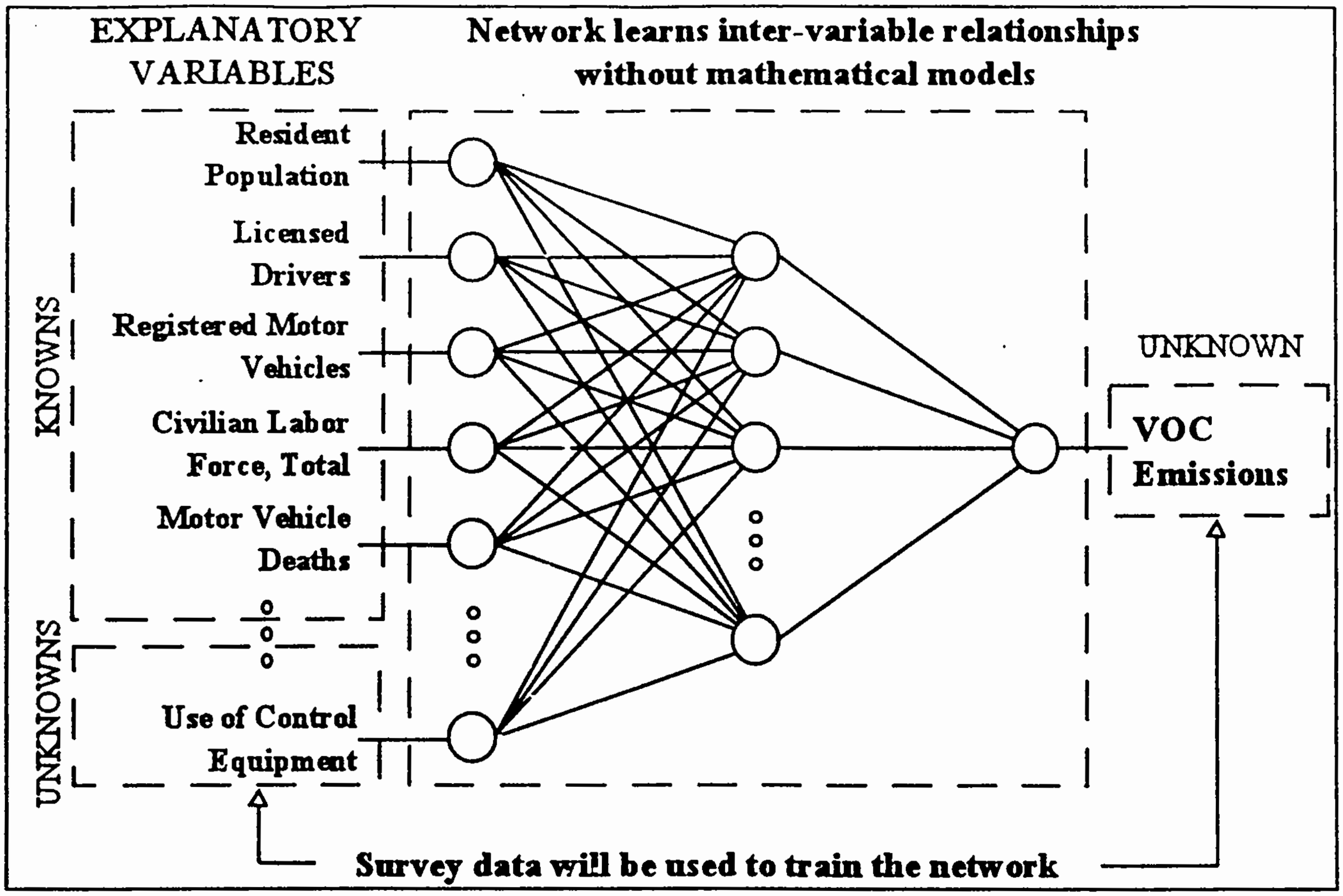

Figure 4. A use of neural networks in the estimation of VOC emissions.

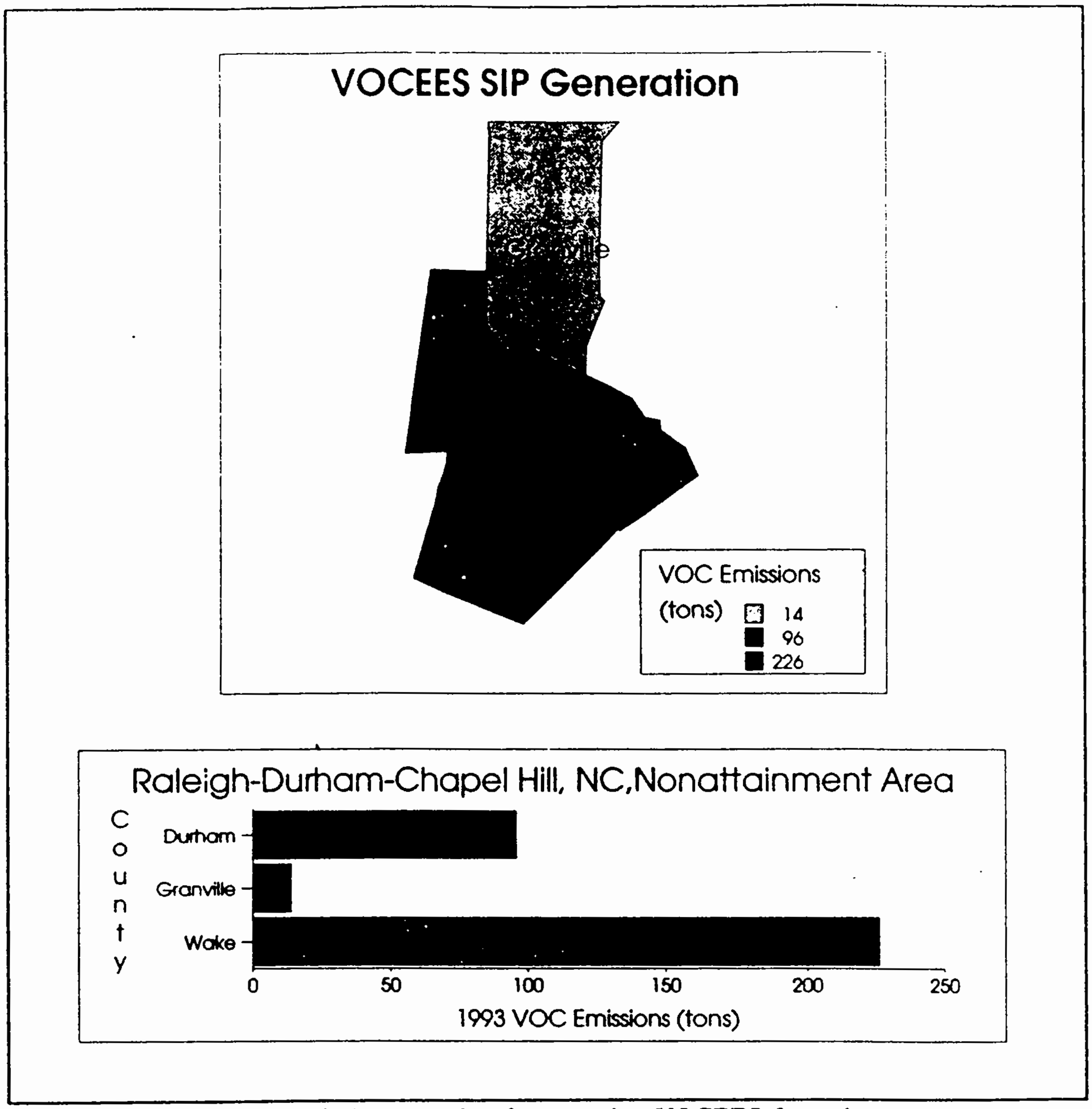

Figure 5. Example emission reporting format using VOCEES for a three-county nonattainment area in North Carolina.

EPA/600/A-96/116

Methods for Improving Emissions Estimates

H. L. Waters, V. E. McCormick, J. G. Cleland, and J. R. Youngberg
Research Triangle Institute
Research Triangle Park, NC 27709-2194

P. J. Chappell
U. S. Environmental Protection Agency
Air and Energy Engineering Research Laboratory
Research Triangle Park, NC 27711

ABSTRACT

The Environmental Protection Agency (EPA) is investigating ways to improve methods for estimating volatile organic compound (VOC) emissions from area sources. Using the automobile refinishing industry for a detailed area source case study, the Research Triangle Institute (RTI) and the EPA's Air and Energy Engineering Research Laboratory (AEERL) are developing an emission estimation method that utilizes both advanced computational techniques and updated, comprehensive, emissions-related information. This development includes a thorough characterization of the area source industry, an analysis of current emission estimation methods, the development and execution of a nationwide industry activity survey, and a compilation and analysis of the survey results and other explanatory variables. Results are to be captured in a personal computer-based VOC emissions estimation system called VOCEES. VOCEES is a dual-use tool, the users of which can both prepare VOC emissions inventories and analyze the impact of numerous factors on emissions. This methodology and VOCEES are readily extendible to other area sources.

INTRODUCTION

Stationary sources of pollutant emissions are designated as either point sources or area sources. Whereas point sources are inventoried on an individual basis, area source emissions emanate from processes, activities, or businesses that are too small or too numerous to be practically tracked as individual emission sources. The distinction between point and area sources is an annual emissions cutoff, such as 10 tons (9090 kg) of volatile organic compounds (VOCs) per year per source. The U.S. Environmental Protection Agency's (EPA's) Air and Energy Engineering Research Laboratory (AEERL) has initiated an effort to develop appropriate area source emission estimation methodologies for solvent area source categories using available technologies and to evaluate long-term informational needs. The Research Triangle Institute (RTI) and AEERL are working together to establish an emission estimation methodology development process that will result in accurate emission estimate methods for all area sources. Results of this work will either validate existing area source emission estimation methods or recommend alternative methods.

The required components of an emission estimation methodology are: 1) the calculation of the emission estimates (e.g., using emission and activity factors); 2) the temporal and spatial allocation of the emissions; 3) the validation of the emission estimates, and 4) the speciation of the emissions. The work presented in this paper concentrates on the first three components. Tools are available from EPA that can be used to develop a VOC inventory grouped into reactivity classes suitable for modeling [1]. The criteria used in developing an emission estimation methodology include: reasonable accuracy and cost; dynamic and robust behavior; use of readily available information; and ease of use.

Why New Methods?

This work has been initiated for two reasons: 1) the current emission estimation methods require further development to improve accuracy and ease of use; and 2) innovative tools have emerged which may contribute significantly to improving estimation methods. Among the issues and uncertainties are:

- Most solvents used by any one solvent area source category are also used by other industries, preventing use of solvent manufacturer production figures alone for estimating emissions from one solvent source category.
- There are a large number of very different solvent area sources categories. Therefore, reasonably accurate emission estimates require area-source-category-specific estimation methods.
- Current methods endorse the use of data sources whose published data are from 2 to 5 years old and may misrepresent a significant segment of the area source industry. For example, one-person establishments are a significant percentage of the estimated 65,000 automobile refinishing establishments in the U.S. Since most federal information sources are derived from business payroll records, there is often no record of these businesses in the endorsed information sources. Also, data at the county level for certain Standard Industrial Classification (SIC) codes are often not disclosed as policy.
- Current methods do not consider the dynamic factors that impact area sources such as local economics, changing technology, and regulatory influence. Nor do they consider human behavior and consumption patterns. For example, it is difficult to estimate automobile refinishing emissions from insurance payments because people are more likely to "pocket" insurance payments for vehicle repair if they have more pressing expenses.

With the emergence of new technologies, such as innovative computational techniques, often comes the opportunity to improve on existing methods. Part of the ongoing work being sponsored by AEERL is to investigate the appropriateness of applying advanced, inference-based computational intelligence techniques, such as neural networks, fuzzy logic, and genetic algorithms, to emissions estimation. These techniques provide the ability to determine and utilize relationships between two domains (e.g., industry variables and emission levels) without developing traditional mathematical models. They also provide the ability to incorporate expertise expressed in inexact, English-like terms, such as *big, small, fast,* or *slow.* Also, since all emissions have a geographic component in that each emission has a source, it is helpful to capture and represent emissions-related data using a geographic information system (GIS). A GIS supports the organization, analysis, and visualization of data by their geographic orientation. This is a powerful tool when determining the characteristics of a geographic region, such as personal income levels, that may impact emission levels.

Relevant Work By Others

Most EPA-endorsed methods for estimating solvent emissions from area sources have been derived from a methodology developed as part of the National Emissions Data System (NEDS) [2] on sources of airborne pollutants. This methodology determines national solvent emissions through material balance, allocating emissions to states and counties using *emission factors,* which are based on source-specific emission measurements as a function of source activity levels. The NEDS area source emission estimation method has been used in several emission inventory efforts. These efforts include the 1985 National Acid Precipitation Assessment Program (NAPAP) emissions inventory prepared by EPA [3], the Regional Ozone Modeling for Northeast Transport (ROMNET), and the Area and Mobile Source (AMS) Subsystem of the Aerometric Information Retrieval System (AIRS) [4], which will replace NEDS.

There have also been detailed studies performed on individual area sources in order to provide guidance on area source emission control technology. Methods for estimating emissions are often included in these studies. A 1988 study of VOC emissions from automobile refinishing area sources uses a rather complex calculation dependent upon the final thickness of applied coatings, in thousandths of an inch, while categorizing the entire industry into three types of establishments -small, medium, and volume shops [5]. This approach has been criticized as being inappropriate for *an industry* which has a *large* variance in operating characteristics [6].

METHODOLOGY DEVELOPMENT PROCESS

The goal of this work is to establish an emissions estimation methodology development process that can be applied to all solvent area source categories. The proposed process is shown in Figure I and includes the following functions:

1) Select the area source category for which the emission estimation method is to be developed.
2) Thoroughly characterize the chosen area source category so as to fully understand the issues affecting emission estimation for that category.
3) Study all methods currently used for estimating the category's emissions, so as to understand their strengths and weaknesses.
4) Develop an *explanatory variables* database -- data which directly or indirectly affect the area source category operations and assist in explaining its emission levels and fluctuations.
5) Assemble a tool set containing statistical tools, computational intelligence tools,

and geographic information system tools for analyzing and relating assembled data.

6) Conduct the nationwide area source survey using a questionnaire reviewed and endorsed by industry experts and by trade associations. Compile and analyze the survey results.
7) Select readily available explanatory variables that most significantly impact emission levels. Structure and complete the method which uses these variables to estimate emissions from the area source category, and which satisfies the specified criteria.

Area Source Selection

VOC-emitting area sources include gasoline distribution losses, stationary source solvent evaporation, bioprocess emission sources, catastrophic or accidental releases, solid waste incineration, and small stationary source fossil fuel use. Stationary solvent evaporation sources include dry cleaning, surface cleaning, surface coating, graphic arts, asphalt paving, pesticide application, commercial or consumer solvent use, and synthetic organic chemical storage tanks. Automobile refinishing, a surface coating category, was chosen for study because it is representative of VOC emission area source categories in terms of relative environmental significance, national prevalence, and an accessible information base representative of such sources.

Industry Characterization

A detailed characterization of the automobile refinishing industry was conducted in order to fully consider the issues and variables associated with the industry's VOC emissions. The number and employment of automobile refinishing establishments vary across information sources. According to industry representatives, the impact of technological and regulatory changes, along with economic factors, has provoked a steady decrease *in number of automobile refinishing establishments* nationwide over the past 20 years [5][7]. They attribute this reduction in establishments and paint use to: 1) improved vehicle safety and more stringent enforcement of traffic laws; 2) more corrosion-resistant finishes; 3) smaller cars with less surface area to paint per repair; and 4) more efficient spray guns which use less paint per job. Other information sources, however, have reported a steady, yet gradual increase in number of establishments. The difference may be explained in part by undocumented "backyard" establishments, which may represent an additional 25% to 40% of the total number of establishments. These "unseen shops" are potentially a large source of VOC emissions since they are less likely to use emissions control technology or comply with laws and regulations, and more likely to use high-VOC coatings.

With the passage of the Clean Air Act Amendments, and other EPA initiatives, new regulations for the automobile refinishing industry are emerging. Regulations requiring use of low-VOC coatings have forced the industry away from lacquer and enamel coatings and toward urethanes. Additional emission control methods include:

- Use of enclosed equipment cleaning devices that support solvent reuse.

- Increased paint (to surface) transfer efficiency through use of high-volume, low-pressure (HVLP) spray guns.
- Addition to paint spray booths of emission controls which use the following techniques: thermal incineration, catalytic incineration, and carbon adsorption. All three methods may be too expensive to be considered as universal, reasonable, control techniques.
- Better shop operations management, including: use of tight fitting containers; reduction of spills; rigid control of inventory; tracking of worker use rates; mixture of paint to need; provision of operator training; use of proper cleanup methods; and use of in-house or leased solvent recycling.

Current Methods Review

Methods for estimating solvent emissions from area sources are based on use of emission factors. Emission factors "are developed from only a limited sampling of the emissions source population for any given category, and the values reported are an average of those limited samples and may not be statistically representative of the population" [4]. The basic approach in estimating emissions is derived from a simple calculation that requires an estimate of an activity level, an emissions factor relating emissions to activities, and, if the source has a pollution control device, a control factor:

Emissions = activity level x emission factor x control factor

Emission factors can be found in a number of EPA references. Table 1 contains the emission and activity factors for automobile refinishing from *AP-42 Compilation of Air Pollutant Emission Factors* [8] and its Supplement D [91, the EPA State Implementation Plan (SIP) guidance document[10], and AIRS/AMS [1 1] [12]. In analyzing these data, note that:

- The AP-42 emission factors and national VOC emissions did not change for more than 10 years. They may not have changed since the AP-42 was first published in 1972.
- The SIP guidance document's *per capita* emission factor is 21% greater than that of the AP-42 document, while the *per employee* emission factor is 48% less than that of the AP-42 document. Some difference may be explained by the 1987-1988 SIC code changes (i.e., using SIC 7532, *Automotive Top and Body Repair and Paint Shops,* instead of SIC 7535, *Automotive Paint Shops).*

Examples of two problem areas are:

1) Difference in Emission Estimates. The two emission factors *(per capita vs. per employee)* seldom produce the same emission estimate, as shown in the example in Figure 2 for 26 California counties. These discrepancies arise from the static nature of emission factors and the different characteristics of individual geographic regions.

2) Missing Data. County Business Patterns [13] exercises disclosure protection for counties where revealing information may disclose details about the operations of individual businesses. In 1990, SIC 7532 employment for 47 of North Carolina's 100 counties was not disclosed. An emission estimation method should utilize activity factors that are available for all areas required to compile emission inventories.

Database Development

The success of any new method to estimate area source emissions is limited by the availability and accessibility of information. The objective is to identify data that are updated at least annually and represent county-level activity. Records should exist for at least the past 5 years and into the foreseeable future. The data must be statistically defensible, representative, "universal" (representative of national distributions and/or variations), and result in more accurate emissions estimates. The data are then related to the primary area source variables of both a) solvent use and b) emissions. These variables may be normalized -- for example, on the basis of per capita, per employee, or per operation.

The database assembled for this study has both geographic or spatial components (e.g., nation, state, county, or city) and temporal components (e.g., year or month). The current data set includes:

- 60 variables over 12 years for the United States
- 25 variables over 12 years for 51 states (including DC)
- 1 1 variables over 12 years for 3126 counties
- 10 variables for 1993 for 64,524 automobile refinishing establishments

Analytical Tool Set Selection

An analytical tool set has been assembled for the purpose of capturing the relationships between explanatory variables and emission levels. This set includes both Computational Intelligence tools and traditional statistical tools. *Computational Intelligence* is a term adopted by the Institute for Electrical and Electronics Engineers (IEEE) for innovative computational techniques that include artificial neural networks, fuzzy logic, and genetic algorithms.

Artificial Neural Networks. An artificial neural network (ANN) is an analysis tool that is modeled after the massively parallel structure of the brain. It simulates a highly interconnected, parallel computational structure with many relatively simple, individual processing elements or neurons. Feed-forward ANN paradigms are capable of learning or extracting a relationship between two domains. ANNs are best applied where there are no known rules or mathematical models that accurately relate the independent and dependent variables of interest, but there is an abundance of data representing the relationship.

Fuzzy Logic Expert System. An expert system is a knowledge-based system that captures human expertise in a specific knowledge domain. A fuzzy logic expert system is being developed to augment estimation of VOC emissions from the automobile refinishing area sources. Fuzzy logic is an approximate reasoning technique used in processing inexact information. While a typical expert system may be thought of as defining "true or false" conditions, fuzzy systems allow for varying degrees of truth, or "shades of gray," more like human reasoning.

For example, if climate is a factor in an area's emission levels, then it could be classified as *dry, moderate,* or *rainy.* The type of area might also be loosely classified into three fuzzy sets: *rural, suburban,* and *urban.* A typical fuzzy rule, based on expert opinion, may be expressed as, "IF the climate is *dry* AND the area is *rural* THEN emissions are low." Another rule may state "IF the climate is *wet* AND the area is *urban* then emissions are *high.*" These rules describe the increased likelihood of accidents and auto refinishing in a congested area with poor weather and vice versa. The fuzzy system uses the degree of membership of an input in a given set to determine to what degree the output belongs in any set (e.g., *low, medium, high).* This type of reasoning can augment the emissions prediction based on optimally correlated data.

Genetic Algorithms. Genetic algorithms (GAs) are a class of machine learning search algorithms based on the mechanics of natural selection and natural genetics. GAs combine Darwin's "survival of the fittest" with structured, yet randomized information exchange to form a search algorithm with some innovative flair of human search [141. Created by John Holland at the University of Michigan [15], GAs are robust, general-purpose problem solvers especially suited for optimization and classification. GAs can develop better solutions from thousands of choices more effectively than other techniques for a host of problems.

Use of GAs in the emissions problem will be confined to methodology development. Examples of using GAs would be in the design and training of neural networks or development of optimum fuzzy logic membership functions. To properly optimize a system, the GA must know what constitutes good performance. In estimating emissions, this would be a knowledge of the VOC emission level for a corresponding set of explanatory variables resulting from the nationwide industry survey.

Statistics. Simple statistical techniques have been employed as the preliminary step for analyzing variable relationships and selecting a priority explanatory variable set. Multivariate regression analyses (linear correlations) have been applied and correlation indicators derived. There is a high correlation between the automobile refinishing industry and several of the explanatory variables included in the database. The variables have been analyzed by examining their distribution across different regions of the country, their change over time, and their ratio one to another.

Geographic Information System. Emissions are characterized by their levels and by their distribution in space and time. A geographic information system (GIS) is being used to assign emissions-related explanatory variable values, such as number of employees and annual sales, to the actual location of an individual business. These values can then be aggregated to combine values within a ZIP Code, county, non-attainment area, metropolitan statistical area (MSA), state, EPA Region, or nation. For State Implementation Plan (SIP) emission inventory purposes, for example, area source emissions would be aggregated to county and to non-attainment area boundaries.

Nationwide Industry Survey

One of the most important steps in the emission estimation method development process is the nationwide survey of the chosen area sources. The Automobile Refinishing Solvent Use Survey (ARSUS) is the first survey of its kind. It is also essential for validation of the new emission estimation method. Important features of ARSUS are:

- The survey is designed based on a detailed knowledge of the industry built on numerous contacts with shop operators, paint manufacturers, and association representatives.
- Results will be statistically defensible, based on random probability sets, with the results represented by statistically correct accuracy estimations and confidence levels.

The map shown in Figure 3 represents the scope of coverage that the 5900 samples will provide. ARSUS includes local-area intensive surveys of six high-population areas. The survey data are divided into two independent sets with probability proportional to population, each containing 30 areas or Primary Sample Units (PSUs). One set is assigned for estimating model parameters (developing the method), while the second set is for developing a comparison variable (evaluating and validating the method). The estimated accuracies of solvent use data from the surveys are shown in Table 2.

Information services and computerized "Yellow Pages" are used to retrieve information abstracts on firms with a range of SIC codes for automobile refinishing in the counties selected. This file is stratified by SIC and number of employees, with a probability sample of organizations in each of the sample PSUS. A detailed file for each sample organization is then retrieved with the names and addresses for the mail survey and auxiliary data for use in the final statistical analyses.

The survey combines mail and telephone contacts to maximize response rates, minimize respondent burden, and complete the data collection efficiently. Survey results are entered using bar-coded identification labels and event codes which indicate the pending or final status. Results are to be tabulated in data files for analysis and inclusion in the developed method.

Method Implementation

The final step of the emission estimation method development consists of bringing together all explanatory variable and survey data, analyzing them, and processing them with the assembled tool set. More than one computational intelligence tool is needed to meet the specified method development criteria. The tool set's statistical components are being used to identify the best data for use in development. Also, fuzzy set clustering is being used during preprocessing to group the geographic regions (e.g., counties) with similar characteristics for additional analysis. The preprocessed survey results and explanatory variable data can be presented to a neural network as illustrated in Figure 4. Initially, half of the available data will be used to train the neural network while the other half will be used to evaluate that which the network has learned. Genetic algorithms will be used during this phase of development to identify the optimum neural network architecture, learning algorithm, and other critical parameters.

Finally, the entire database will be used to train the network, to continue to identify and minimize the input variable set which best estimates emission levels. Analysis of the trained network will identify the input variables which are most important in estimating emission levels, reducing the number of input variables required. A sensitivity analysis will also be performed to determine the error introduced by this minimized artificial neural network.

A personal-computer-based VOC emission estimation system (VOCEES) will automate the developed method. The components of the system are: 1) the essential explanatory variable database; 2) basic algorithms and possibly an artificial neural network-based computational component; 3) the supplemental fuzzy logic expert system; and 4) the GIS-based user interface.

PRELIMINARY RESULTS AND CONCLUSIONS

Practically all of the emission estimation method case study is complete. The hardware and software implementation of the method is called the Volatile Organic Compound Emission Estimation System or VOCEES. The system is a dual-use tool, enabling the user both to prepare standard VOC emission inventories for geographic regions of various sizes, and also to analyze how emission levels are influenced by various factors. VOCEES currently combines the emission-factor-based VOC estimation with the predicted impact of current and anticipated VOC-limiting regulations. The system supports tabular, bar graph, and map formatted analysis and reporting facilities. An example is presented in Figure 6.

Of the seven functions specified by the proposed development process shown in Figure 1, work performed to date has completed the first five functions and the preparation and review of the survey questionnaire as described in the sixth function. It is anticipated that all seven steps will be completed for the automobile refinishing case study and the process will be extended to other area source categories. No technical barriers to this process have been identified, and the early results demonstrate a more efficient and cost-effective emissions estimation methodology.

DISCLAIMER

This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication.

REFERENCES

1. *VOCIPM Speciation Data System Documentation and User's Guide, Version 1.32a;* EPA-450/291-002; U. S. Environmental Protection Agency: Research Triangle Park, November 1990.

2. Myers, J.P., Benesh, F.; *Methodologies for Countywide Estimation of Coal, Gas, and Organic Solvent Consumption;* EPA-450/3-75-086 (NTIS PB259909); U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, December 1975.

3. Demmy, J., Tax W., Wam, T.; *Area Source Documentation for the 1985 National Acid Precipitation Assessment Program Inventory;* EPA-600/8-88-106 (NTIS'PB89-151427); U.S. Environmental Protection Agency, Air and Energy Engineering Research Laboratory: Research Triangle Park, December 1988.

4. Seinfeld, J., Atkinson, R., Berglund, R., et al.; *Rethinking the Ozone Problem in Urban and Regional Air Pollution;* National Academy Press: Washington, 1991.

5. Athey, C., Hester, C., McLaughlin, M., et al.; *Reduction of Volatile Organic Compound Emissions from Automobile Refinishing;* EPA-450/3-88-009 (NTIS PB89-148282); U.S. Environmental Protection Agency, Control Technology Center: Research Triangle Park, October 1988.

6. *National Air Pollution Control Techniques Advisory Committee, Meeting Minutes, Volume 2;* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, November 19-21, 1991; p 1008.

7. M. Hunke, Carstar Automotive, Inc., Shawnee Mission, KS, personal communication, 1992.

8. *Compilation of Air Pollution Emission Factors, Volume I.- Stationary Point and Area Sources, AP-42, Fourth Edition* (GPO 055-000-00251-7); U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, September 1985; pp. 2, 4.2.1-1.

9. Joyner, W.; *Supplement D to Compilation of Air Pollution Emission Factors, Volume I.Stationary Point and Area Sources, AP-42, Fourth Edition* (GPO 055-00-00391-2); U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, September 199 1; pp 2, 4.2. 1 - 1.

10. *Procedures for the Preparation of Emission Inventories for Carbon Monoxide and Precursors Of Ozone, Volume I.- General Guidance For Stationary Sources;* EPA-450/4-91-016 (NTIS PB92112168); U.S. Environmental Protection Agency: Research Triangle Park, May 1991; pp. 4-24, 4-47.

11. Kimbrough, E.; *Documentation of AIRS AMS National Methodologies;* EPA-600-R-92-001 (NTIS PB92-132869); U.S. Environmental Protection Agency, Air and Energy

Engineering Research Laboratory: Research Triangle Park, January 1992.

12. *Aerometric Information Retrieval System (AIRS): "Short List" of AMS SCCs and Emission Factors;* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, July 1992.

13. *County Business Patterns;* U.S. Department of Commerce, Bureau of the Census: Washington; Annual Publication.

14. Goldberg, D.E.; *Genetic Algorithms in Search Optimization and Machine Leaming;* AddisonWesley Publishing Company: Reading, 1989.

15. Holland, J.H.; *Adaptation in Natural and Artificial Systems;* University of Michigan Press: Ann Arbor, 1973.

Table 1. Recommended emission factors for the automobile refinishing area source.

Reference	National VOC emissions (tons per year)	VOC emission factors (pounds/year)		U.S. population (millions)	SIC code	Number of U.S. automobile refinishing employees	Year derived
		Per capita	Per employee				
AP-42	199,000	1.9	5200	209.5 (est'd)		76,500 (est'd)	1972 (est'd)
AP-42 Supplement D	199,000	1.9	5200	209.5 (est'd)		76,500 (est'd)	1972 (est'd)
SIP Guidance Document	281,00 (est'd)	2.3	3519	245.7	7532	163,000 (est'd)	1989
AIRS / AMS			3519		7535, 7532		

Table 2. Estimated accuracies of solvent use survey data.

Survey	Sample size fielded	Precision estimate	Difference estimate
National	5,900	7%	10%
Six local intensives	3,400	10%	14%

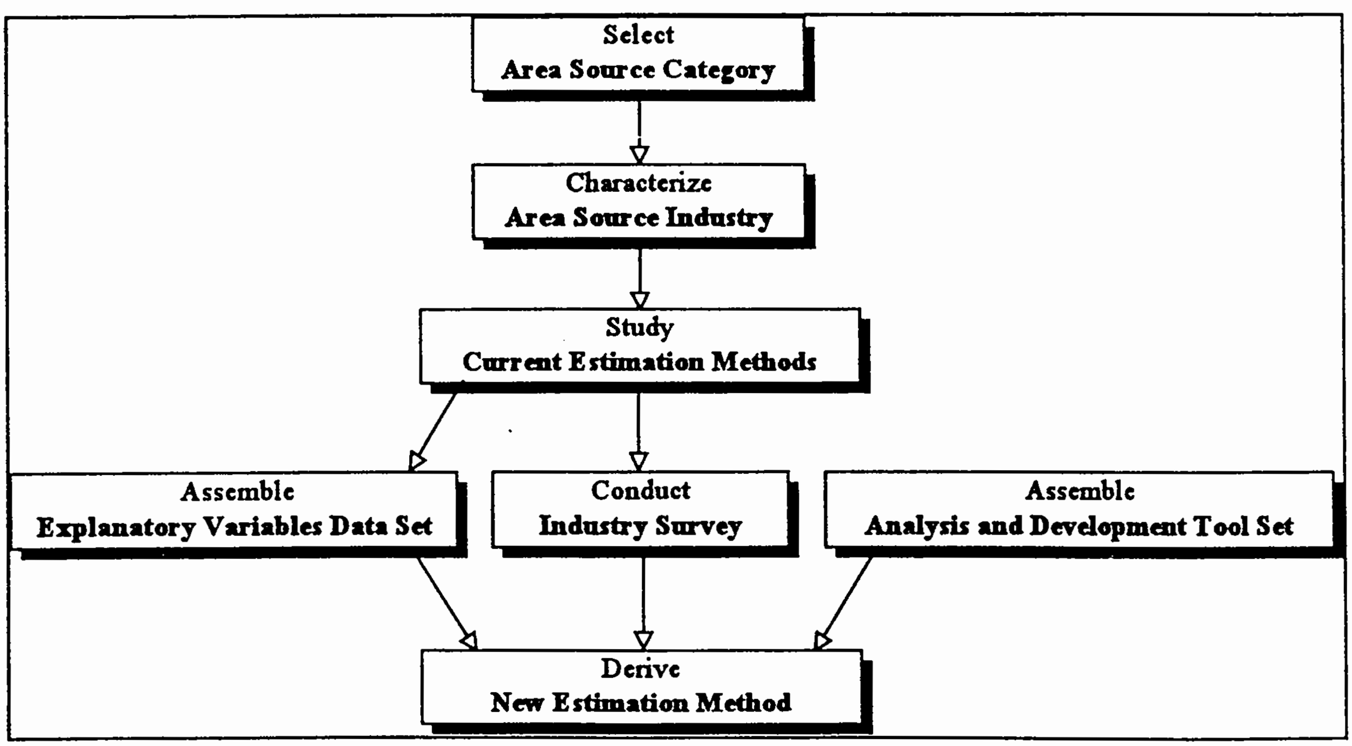

Figure 1. The proposed emission estimation methodology development process.

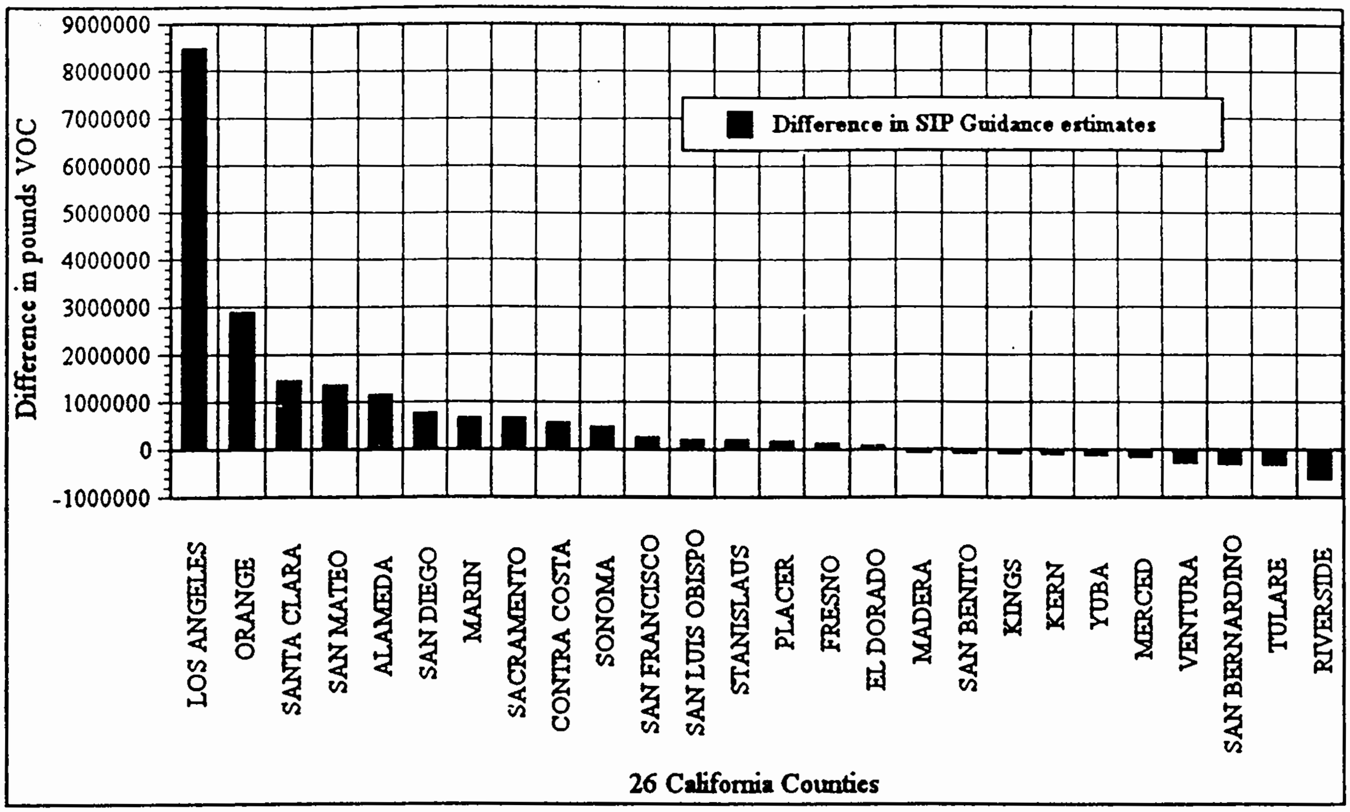

Figure 2. The difference in automobile refinishing VOC emission estimates, using the two SIP Guidance emission factors, *per employee* and *per capita*.

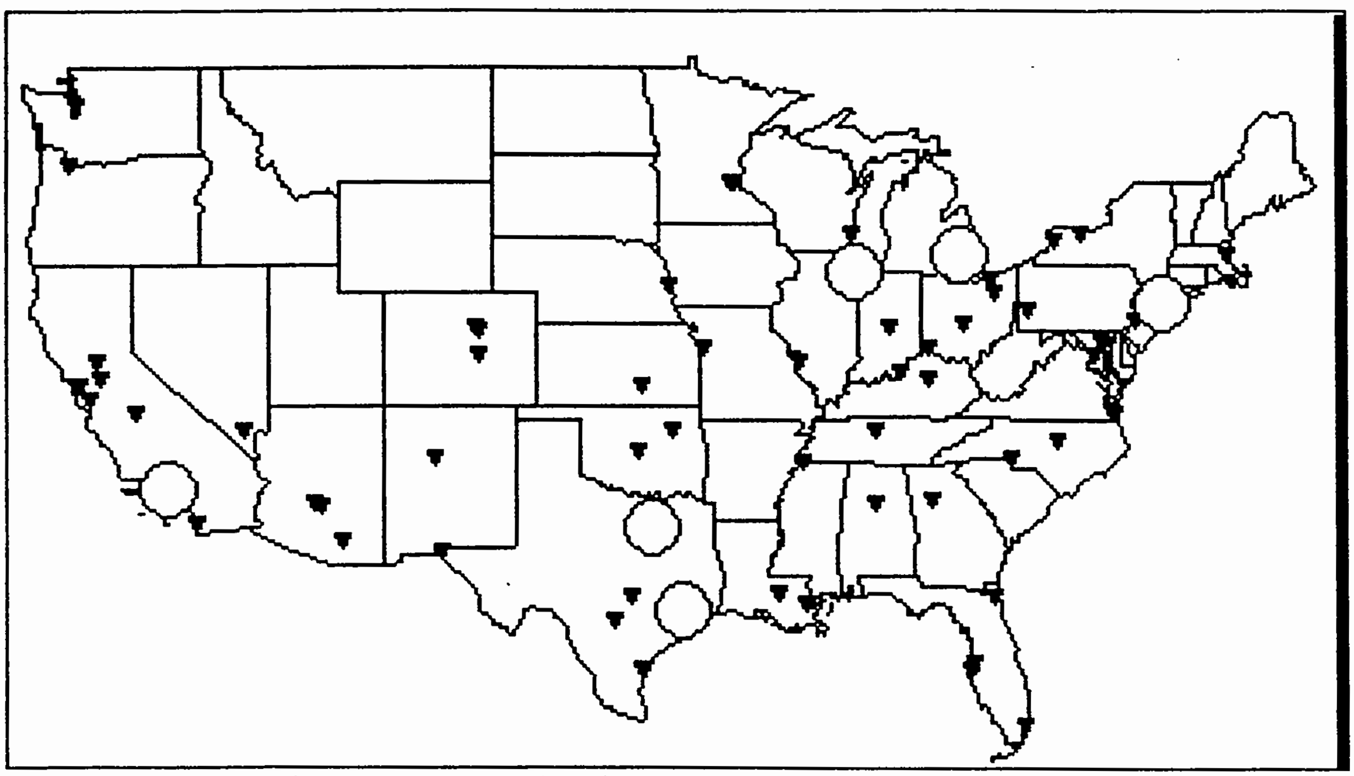

Figure 3. Preliminary targets of nationwide automobile refinishing survey: six highly populated areas and 60 primary sample units (PSUs).

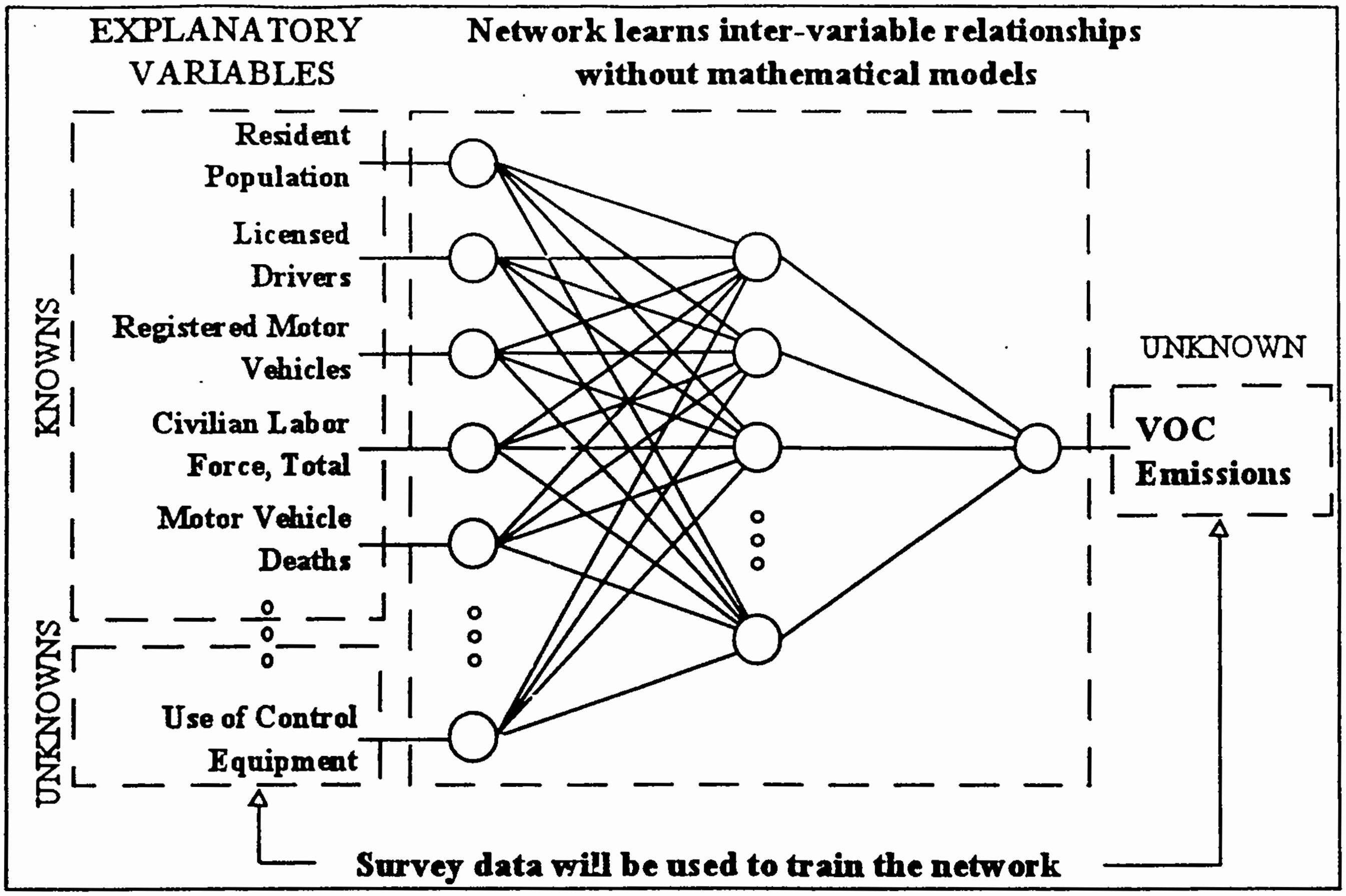

Figure 4. A use of neural networks in the estimation of VOC emissions.

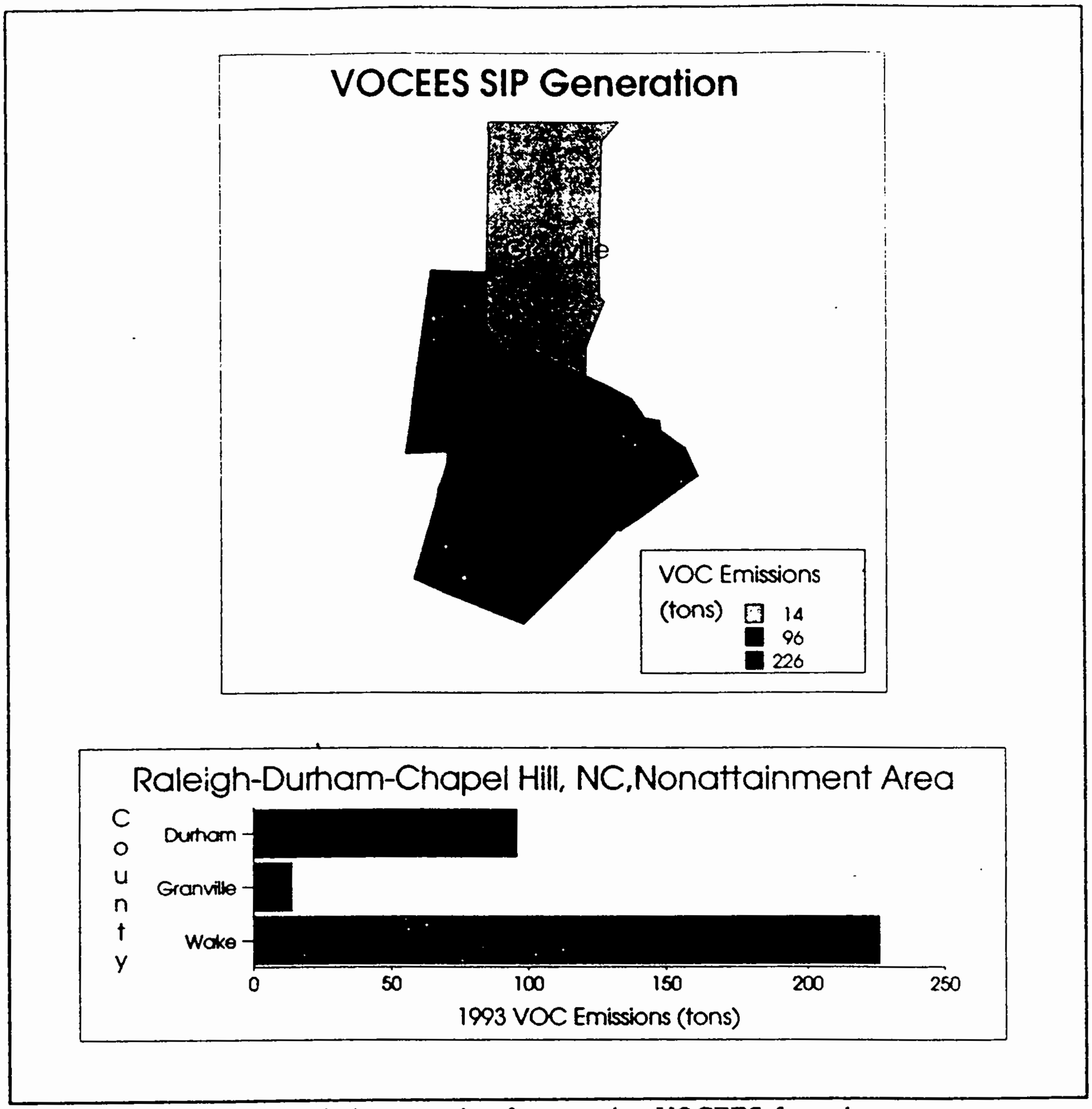

Figure 5. Example emission reporting format using VOCEES for a three-county nonattainment area in North Carolina.