United States Environmental Protection Agency
Research and Development

National Risk Management Research Laboratory
Research Triangle Park, NC 27711

EPA/600/SR-97/005 July 1997

EPA

Project Summary

Improving Emissions Estimates with Computational Intelligence, Database Expansion, and Comprehensive Validation

J.G. Cleland, V.E. McCormick, H.L. Waters, J.R. Youngberg, and J.A. Zak

The EPA is investigating techniques 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, an emission estimation method is being developed that uses advanced computational techniques and updated, comprehensive, emissions-related information. New computational techniques contributing to the estimation method are fuzzy logic, neural networks, and genetic algorithms. This method development requires a thorough characterization of the area sources, 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 will be captured in the personal-computer-based emissions estimation system, VOCEES (VOC Emissions Estimation System). VOCEES has been developed as a dual-use tool that prepares VOC emissions inventories and analyzes the impact of numerous factors on emissions. This methodology can easily be extended to other area sources.

This Project Summary was developed by EPA's National Risk Management Research Laboratory's Air Pollution Prevention and Control Division, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

EPA's emission inventory identifies the types of emission sources in a geographic area, the amount of each pollutant emitted by each type of source, and any emission control devices being employed. In developing these inventories, state agencies may use either emission estimation methods endorsed and provided by EPA or their own methods. New methods of estimation are sought which will be more accurate, efficient, cost effective, dynamic, and robust.

Stationary sources of pollutant emissions are designated as either point sources or area sources. While point sources are inventoried on an individual basis, area sources are processes, activities, or businesses that are too small or too numerous to be practically tracked as individual emission sources.

The required components of an emission estimation methodology are (1) calculation of the emission estimates, (2) temporal and spatial allocation of the emissions, (3) validation of the emission estimates, and (4) speciation of the emissions. This project concentrates on the first three components. Improving existing emission estimation methods addresses such issues as

- Accurate emission estimates require area source-specific estimation methods.

- Current methods use data from 2 to 5 years old which may also miss significant segments of the area source industry or have disclosure restrictions.
- Current methods do not consider such dynamic factors as local economics and consumption patterns, changing technology, and regulatory influence.

This project investigates advanced, inference-based computational intelligence techniques for emissions estimation, augmenting or supplanting traditional mathematical models. Also, since all emissions have a spatial distribution, emissions information is captured and represented using a geographic information system (GIS).

Procedure

The research concentrates on a case study of a single area source, *automobile refinishing*, in order to demonstrate the feasibility of developing an improved method for deriving VOC emission estimates from area sources. The project plan incorporates several new or expanded features for emissions estimation, listed in Table 1.

The new area source emissions estimation method development process involves

1) area source characterization, examining materials balances, process operations and controls, and economic influences,
2) review of current emissions estimation methods,
3) database development, including retrieval and screening of the best accessible data correlating with area source characteristics and emissions,
4) selection of analytical methods (the "tool set") for calculating or inferring final emissions levels and their distributions,
5) system configuration for data input, computation, and reporting, and
6) validation of results (for this case study, emphasizing a national survey of automobile refinishing shops).

For the case study, steps 1 and 2 are complete, steps 3 to 5 are near completion, and a plan for step 6 is complete.

Results and Discussion

Area Source Characterization

A detailed characterization of the automobile refinishing industry was conducted. The number of U.S. cars increases each year, but the volume of auto refinishing solvent use has stagnated because of reduced numbers of accidents, more corrosion resistant paints, and more efficient paint spray guns. According to the National Paint and Coatings Association, over 36 million gallons of coatings were sold in the U.S. in 1989. That number has re-

Table 1. Project Initiatives for New Method Development

Initiative	Description
Annual updating	The method will use information and logic that is continuously reviewed and revised.
Improved validation	New validation activities include: (a) preparation of the most thorough nationwide *survey of the automobile refinishing industry (i.e., auto refinishing shops*) to date. The survey will obtain industry answers about activity levels(e.g., number and types of employees or repair jobs), solvent usage, and emission control. In addition, extensive product distribution data have been requested from all major automotive *paint manufacturers;* (b) selective *sampling and chemical analysis* of local sources has been designed to establish actual mass rates of shop emissions; and (c) *literature review and consulting with experts* to provide new insight into major influences on emissions.
Intensive area source characterization	Characterization includes main industry variables influencing emission levels. In addition to the national survey, the research team has met with shop managers and industry consultants, attended national conferences, contacted associations, and analyzed the literature.
More extensive data	Data analysis will extend down to the county level. Data from federal and state agencies, trade associations, and industry sources have significantly increased the EPA information base related to this area source.
Improved data recovery and structuring	One result is a systematic and efficient approach for using information, emphasizing data sources which are regularly updated (e.g., annually) and which are readily accessible at minimal cost to EPA and state agencies.
Improved data correlations	This involves applying statistics, for prescreening of the best data relationships.
Application of imprecise information and expert opinion	Such "artificial intelligence" techniques as fuzzy logic, genetic algorithms, and rule-based expert systems can augment the use of pertinent information which had been neglected to this point.
Regulatory policy consideration	The method incorporates regulatory influence on emissions at local, state, and national levels.
Alternative methods review	Methods (used by states), other than the standard EPA guidelines and emission factors, have been examined.

mained basically constant over the past ten years.

In addition to more than 60,000 licensed auto refinishing shops, unlicensed "backyard" shops may make up to 25%-40% of the industry. These are potentially a large source of VOC emissions, since their control technology is more likely to be primitive and they are more likely to disregard laws and regulations.

Solvent-based coatings (lacquers, enamels, and urethanes) are by far the most common in automobile refinishing. Waterborne coatings are only now becoming more popular in the refinishing industry. Typical VOC contents range from 4.8 lb/gal (575 kg/m^3) for acrylic urethanes to 6.8 lb/gal (815 kg/m^3) for acrylic lacquers.

Auto refinishing accounts for 3% of VOC emissions in the U.S. The three main approaches to control are 1) use of lower-VOC coatings, 2) use of enclosed cleaning devices, and 3) increase of paint-to-surface transfer efficiency. EPA is considering implementation of a national rule that would limit VOC contents in the various automotive refinishing coatings, just as some states have done.

Review of Current Estimation Methods

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). 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 estimates of effectiveness of environmental controls and regulation:

Emissions = Activity level x Emission factor x [1-(CE x RE x RP)]

where: CE = Control efficiency percent/100
RE = Rule effectiveness percent/100
RP = Rule penetration percent/100

CE is related to control technology (e.g., carbon filtration of VOCs from auto refinishing paint booths). RE is an adjustment to reflect that air pollution rules are not able to ensure full compliance with regulatory requirements at all times. RP is a measure of the extent to which a rule applies to a source category.

Per capita and *per employee* activity levels (e.g., gallons of solvent per capita) are provided to the state pollution control agencies by the EPA in State Implementation Plan (SIP) guidance documents.

Figure 1 shows typical discrepancies between different methods and factors. A national value for VOC emissions from auto refinishing has been calculated from information provided by the seven different sources. The activity factor used in each case is based on population. The two state studies are most closely based on actual contacts with body shops within specific regions of the country. Without further validation, it is not possible to proselytize for one method or the other.

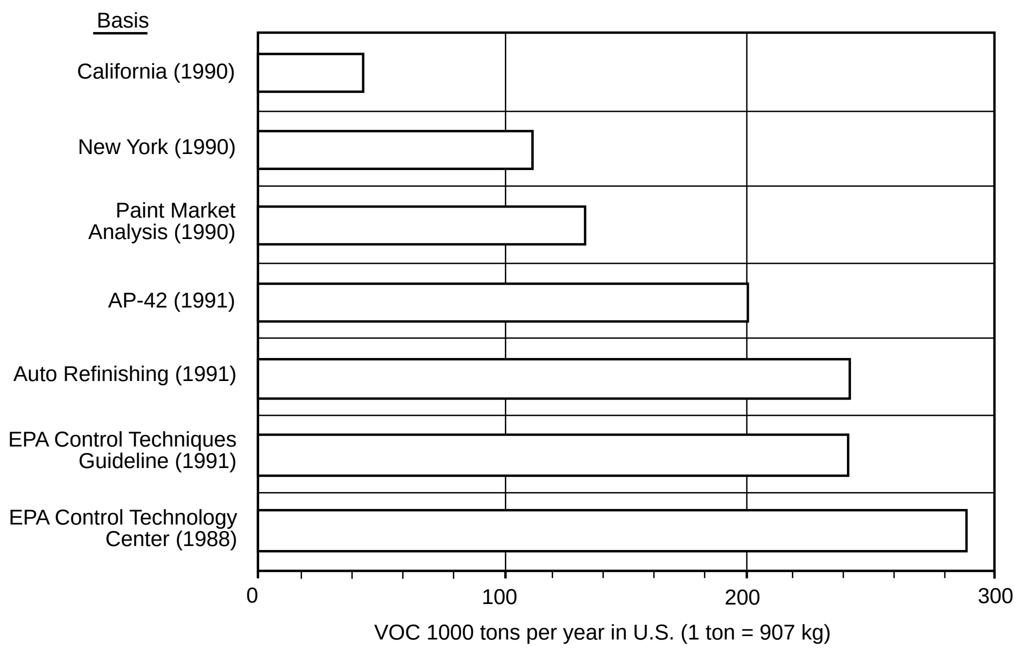

Figure 1. National VOC estimations by different methods.

Database Development

The database that has been assembled has both geographic "spatial" components (e.g., nation, state, county, or city) and temporal components (e.g., year or month).

The current data set assembled for this study includes

- 60 variables over 12 years for the U.S.,
- 25 variables over 12 years for 51 states (including DC),
- 11 variables over 12 years for 3,126 counties, and
- 10 variables for 1993 for 64,524 automobile refinishing establishments.

Variables selection and final arrangement of the desired databases are necessary to tailor a true analysis and estimation system for optimal emission predictions. The current information remains *surrogate data* until validation is provided by comparing predicted emissions to actual solvent use information. Nevertheless, some interesting preliminary results and selections have come from the evaluations to date.

The surrogate data sources include 1) census population data, 2) automobile sales and accident data, 3) general economics data, 4) statistical data from the *Statistical Abstract of the United States,* Bureau of Economic Analysis, the Bureau of the Census, and the Bureau of Labor Statistics, 5) annual income and employment information covering 1969 to 1990 for states, counties, and metropolitan areas, 6) Regional Economic Information System (REIS), 7) direct marketing information, 8) state regulations for VOC emissions from automobile refinishing, and 9) Occupational Safety and Health Administration regulations. Data subsets were first selected based on availability, frequency of revision, and expert option on relation to auto refinishing emissions.

To provide working surrogate variables and data for the emission estimation method, several data subsets were further screened using *statistical regression analysis.* Annual data collected at the state level have proven to be the most useful for analyzing trends and regional variations.

Most variables were considered as independent, or explanatory. A few variables (i.e., number of auto refinishing employees, receipts for the auto refinishing Standard Industrial Classification (SIC), and sales for the paints and allied products SIC) were selected as most closely representing area source activity. Using these as surrogates for solvent use, the independent variables were correlated.

The best explanatory variables for use in estimating derived state-level emission estimates are total civilian labor force, licensed drivers, employed civilian labor force, and resident population. The best correlations with paint sales were for urban vehicle miles traveled, licensed drivers, and civilian labor force.

Analytical Tool Set Selection

The current tool set includes computational intelligence (CI) tools, statistical tools, and a geographic information system. CI is a term adopted by the Institute for Electrical and Electronics Engineers (IEEE) for innovative computational techniques like expert systems, artificial neural networks, fuzzy logic, and genetic algorithms. CI is being used to

- Select the best initial data matrix to become the VOCEES resident datafile,
- Continuously update the best data by a training function,
- Verify and/or improve statistical correlations,
- Quantitatively interpret qualitative responses from a nationwide survey,
- Interpolate to fill in data gaps,
- Develop rules and produce if/then scenarios using expert opinion,
- Assess geographic and time dependent influences to establish trends,
- Provide a structured technique for adding and evaluating new factors, and
- Support optimal information displays and information transmittal.

An expert system is a computer-based system that contains human expertise or reasoning capabilities. A new expert system for area source emissions must evolve if one is to be applied.

The rules derived from industry and emissions experts thus far have been, at best, very general. Since "crisp-logic," rule-based expert systems do not handle "approximate reasoning" well, a fuzzy logic expert system is being developed. 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. Fuzzy logic will supply a set of secondary emission factors (V1, V2, V3,...Vn) based on qualitative or uncertain influences to augment the best quantitative data correlations between emissions and independent variables. These would contribute by such a relationship as

Total Emission Factor = $EF_{quantitation}$ x V1 x V2 x V3....

where a specific example might be

if{county=suburban} and {winter=average} then {VOC emissions factor V3=moderate}

The linguistic result is then "defuzzified" to provide a numerical value for V3.

An artificial neural network is an analysis tool that is modeled after the massively parallel structure of the brain. The neural network does not have to be programmed, but learns from example. A neural network's ability to generalize will prove beneficial in data interpolation. Once a network has been trained, it provides an instantaneous output, without iterating, for each set of inputs.

Genetic algorithms select optimum rules and data by an evolutionary process analogous to a "survival of the fittest" approach.

A personal-computer (PC)-based VOC Emission Estimation System (VOCEES) automates the data management, computation, and displays/reports under the new method. The main components of VOCEES are 1) a variable database screened by neural nets and/or genetic algorithms, 2) basic computational algorithms, 3) the supplemental fuzzy logic expert system, and 4) the GIS-based user interface, that presents data in both map and graph displays. Users have a choice of examining counties, nonattainment areas, states, or EPA regions. Also, since VOCEES can be used to examine emissions for different years, temporal changes can be observed.

Extended Method Development

Five areas of new method development have begun under this case study, but not completed. These five areas should be extended to completion in the future, in suggested order of importance:

1) Nationwide Automobile Refinishing Solvent Use Survey (ARSUS) for Validation of Emissions and Verified Correlations with Explanatory Variables — this activity is essen-

tial to the project's success, confirming solvent use and variable correlations. ARSUS is almost certainly the most significant area source emissions estimation validation undertaken to date. Important features of ARSUS are a) results will be statistically defensible, based on random probability sets and represented by accuracy estimations and confidence levels, b) proper survey techniques will be applied to assure a high percentage response, c) results are expected to increase available validation data by three orders of magnitude and improve accuracy by 20% to 200%, and d) the survey is designed based on a detailed knowledge of the industry.

2) Application of Computational Intelligence (Fuzzy Logic Expert System, Neural Nets, and Genetic Algorithms) to Emissions Estimation Using Validated Data — these tools allow new kinds of information to be used, accelerate the data selection process, and provide more accurate estimates.

3) Potential Negotiation of the Use of VOC-Containing Product Manufacturers Data for Validation and Database Updates — important to provide an accurate estimate of geographic distribution of product categories, to compare with survey data, and to extend survey results to an annual validation update.

4) Sampling and Chemical Analysis of Selected VOC Species at a Limited Number of Area Source Sites — important to improve the credibility of emissions predictions and the VOCEES extrapolation of product use data to actual annual volume of emissions of VOC.

5) Graphical Interpretation of Past Emission Estimation Data — important for comparison with VOCEES estimates, to illustrate needs for estimating improvement, and to provide a baseline perspective of techniques to date.

A method for applying the techniques is shown in Figure 2.

Results Summary

The results of the case study cannot be complete until validation data become available and emissions estimates can be better substantiated. However, results of the study to date include

- A comprehensive information search and development of a database;
- Assessment of current emission estimation methods related to the area source and their limitations;
- A PC Windows- and GIS-graphics-based system with computational techniques that provide reasonable examples of system function and output;
- Demonstration of the VOCEES system;
- An examination of computational intelligence and recommendations for incorporation into the overall estimation method;

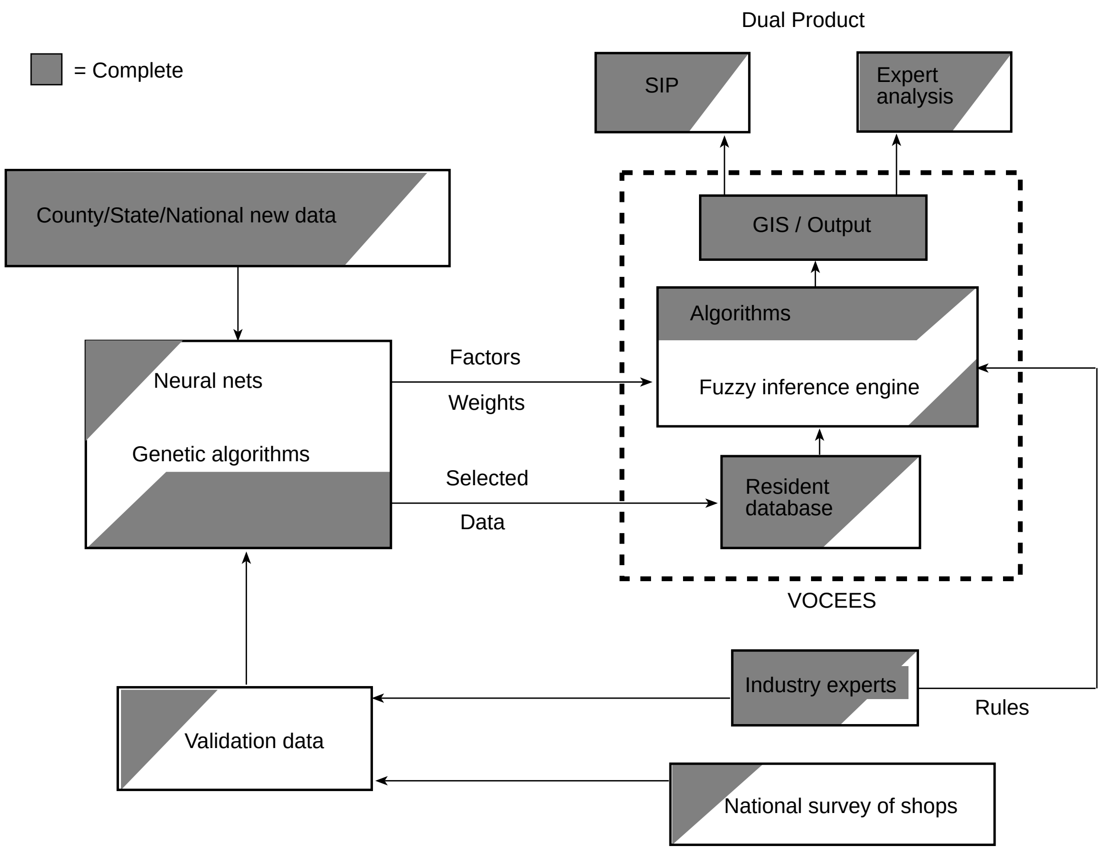

Figure 2. New method for estimating emissions.

- Preliminary screening of explanatory variables using statistical regressions;
- Preparation of a survey (questionnaire, sampling population, and data storage and analysis system), and completion of a presurvey; and
- Development of preliminary source sampling recommendations.

Conclusions and Recommendations

- Automobile refinishing is similar to other major area sources associated with VOC use in terms of establishment size, customer interaction, environmental compliance attitudes, materials suppliers, and demographic influences.
- The auto refinishing industry expects to be using lower-VOC materials and to expand the use of high transfer efficiency equipment like high-volume, low-pressure (HVLP) paint guns.
- Difficulties exist for obtaining product distribution data from automotive paint manufacturers because of confidentiality issues. A system for tracking solvent distribution is needed. Manufacturers and original equipment manufacturers (OEMs) control the product mix. For auto refinishing, sales records of about seven large companies could characterize more than 90% of the VOC distribution.
- Discrepancies in current emission estimation methods exist for auto refinishing, because of unsubstantiated activity factors and of fundamental inaccuracies in activity factor data (e.g., numbers of individual shop employees).
- PC-based analysis and GIS computer graphics displays are a good combination for an accurate, easy-to-use, low-cost emissions estimation system.
- No expert approach exists upon which to base an expert system (i.e., no agency is consistently providing reliable, accessible, and continuously available emission estimates). Therefore, an expert system is needed, one that is built through new expertise and then captured in software.
- Once validation data are obtained for VOC use by an area source, genetic algorithms and neural networks should be efficient for completing selection and weighting of the best explanatory variables, and for training the system to optimally integrate new information.
- Fuzzy logic is appropriate for manipulating rules to apply inferential estimates in augmenting the correlation of VOC usage variables.
- State and county databases should be used as activity factor data wherever possible. The best ones are readily accessible and typically updated annually. EPA would continuously refine the techniques and tools for applying these databases and could be responsible for centralized validation of the method.
- Preliminary indications show that licensed drivers and registered vehicles are better explanatory variables for auto refinishing emissions than population or number of shop employees.
- Data related to the area source appear to be best for emissions estimation purposes, including data related to materials volumes and product users' levels of activity (e.g., registered vehicles).
- An estimate of the impact of regulations and standards, and of their level of enforcement, requires more accurate emissions estimation and prediction.
- The current VOCEES design could evolve into an expert system. Validation will confirm an ever-improving technique that will result in a highly accurate system of rules.
- Validation of data and variable relationships using industry responses is essential to completion of a new estimation method. The best validation is through a national survey of the users of the pertinent VOC-containing materials and/or a distribution of the materials obtained from solvent product manufacturers.

J. G. Cleland, V.E. McCormick, H.L. Waters, J.R. Youngberg, and J.A. Zak are with Research Triangle Institute, Research Triangle Park, NC 27709.
***P. Jeff Chappell** is the EPA Project Officer (see below).*
The complete report, entitled "Improving Emissions Estimates with Computational Intelligence, Database Expansion, and Comprehensive Validation," (Order No. PB97-152565; Cost: $31.00, subject to change) will be available only from:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officer can be contacted at:
Air Pollution Prevention and Control Division
National Risk Management Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States
Environmental Protection Agency
Center for Environmental Research Information
Cincinnati, OH 45268

Official Business
Penalty for Private Use $300

EPA/600/SR-97/005

BULK RATE
POSTAGE & FEES PAID
EPA
PERMIT NO. G-35