Volume 3 Number 2
September 2000
ISSN 1094-8848

Special Issue on the Statistical Analysis and Modeling of Automotive Emissions

BUREAU OF TRANSPORTATION STATISTICS UNITED STATES DEPARTMENT OF TRANSPORTATION

JOURNAL OF TRANSPORTATION AND STATISTICS

JOURNAL OF TRANSPORTATION AND STATISTICS

DAVID GREENE Editor-in-Chief
DAVID BANKS Deputy Editor-in-Chief
SUSAN LAPHAM Associate Editor
DEBORAH MOORE Managing Editor
DEBBIE BAIN Assistant to the Editor-in-Chief
MARTHA COURTNEY Editor
MARSHA FENN Editor

EDITORIAL BOARD

WILLIAM ANDERSON Boston University
MOSHE BEN-AKIVA Massachusetts Institute of Technology
KENNETH BUTTON George Mason University
MARK DELUCCHI University of California at Davis
STEPHEN FIENBERG Carnegie Mellon University
DAVID GILLEN Wilfred Laurier University
GENEVIEVE GIULIANO University of Southern California
JOSE GOMEZ-IBANEZ Harvard University
KINGSLEY HAYNES George Mason University
DAVID HENSHER University of Sydney
PATRICIA HU Oak Ridge National Laboratory
RICHARD JOHN Volpe National Transportation Systems Center, USDOT
GARY MARING Federal Highway Administration, USDOT
MICHAEL MEYER Georgia Institute of Technology
PETER NIJKAMP Free University
ALAN PISARSKI Consultant
EMILE QUINET Ecolé Nationale des Ponts et Chaussées
JEROME SACKS National Institute of Statistical Sciences
ROBERT SKINNER Transportation Research Board
MARTIN WACHS University of California at Berkeley
C. MICHAEL WALTON University of Texas at Austin
TOMMY WRIGHT U.S. Census Bureau

The views presented in the articles in this journal are those of the authors and not necessarily the views of the Bureau of Transportation Statistics. All material contained in this journal is in the public domain and may be used and reprinted without special permission; citation as to source is required.

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 3 Number 2
September 2000
ISSN 1094-8848

BUREAU OF TRANSPORTATION STATISTICS UNITED STATES DEPARTMENT OF TRANSPORTATION

U.S. DEPARTMENT OF TRANSPORTATION

RODNEY E. SLATER
Secretary

MORTIMER L. DOWNEY
Deputy Secretary

BUREAU OF TRANSPORTATION STATISTICS

ASHISH K. SEN
Director

RICHARD R. KOWALEWSKI
Deputy Director

SUSAN J. LAPHAM
Associate Director for Statistical Programs and Services

Bureau of Transportation Statistics

Our mission is to lead in developing transportation data and information of high quality and to advance their effective use in both public and private transportation decisionmaking.

Our vision for the future: Data and information of high quality will support every significant transportation policy decision, thus advancing the quality of life and the economic well-being of all Americans.

The *Journal of Transportation and Statistics* is published by the

Bureau of Transportation Statistics
U.S. Department of Transportation
Room 3430
400 7th Street SW
Washington, DC 20590
USA
journal@bts.gov

Subscription information

To receive a complimentary subscription or a BTS Product Catalog:

mail	Product Orders Bureau of Transportation Statistics U.S. Department of Transportation Room 3430 400 7th Street SW Washington, DC 20590 USA
phone	202.366.DATA
fax	202.366.3640
internet	www.bts.gov

Cover and text design Susan JZ Hoffmeyer
Layout Gardner Smith of OmniDigital
Cover photo Image® copyright 2000 PhotoDisc, Inc.

The Secretary of Transportation has determined that the publication of this periodical is necessary in the transaction of the public business required by law of this Department.

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 3 Number 2
September 2000

Contents

Introduction to the Special Issue on Statistical Analysis and Modeling of Automotive Emissions

Air pollutants generated through the combustion of fossil fuels present a difficult environmental challenge to society. Transportation, which depends heavily on fossil fuels as an energy source, is a prominent contributor to the problem. Emissions of carbon monoxide, volatile organic compounds, and nitrogen oxides, all of which affect local air quality and may cause public health problems, are at least partly attributable to transportation; and transportation is thought to be responsible for a large portion of the greenhouse gas emissions (e.g., carbon dioxide) that have recently been linked to global climate change.

Automobile usage represents a substantial portion of transportation in all industrialized countries, and the demand for automotive transportation absorbs much of the world's energy resources. Consequently, concerns about emissions on the part of the public, the media, and various governing bodies and health-related organizations have recently been more directly focused on the automotive sector. Over the past two decades, major public and private efforts have been undertaken in an attempt to more fully understand the complexity of automotive emissions; yet, despite these initiatives, the extent to which they harm the atmosphere and degrade public health, and the mechanisms by which they do so, are still not completely known. Nonetheless, policymakers in many industrialized countries continue to tighten the restrictions on automotive emissions in an ongoing effort to dampen their environmental impact. In some quarters, particularly within the business and economic communities, such initiatives have been met with determined resistance which, in turn, has contributed to the international debate concerning the ultimate social impact of the pollutants and policies designed to counteract them.

Unresolved air quality issues, the potential for climate change, and political opposition to increased regulation are prompting additional scientific and engineering investigation of automotive emissions and the methods used to control them. At the most basic level, the ability of policymakers to arrive at sound and reasoned decisions about an issue as complex as this depends on the availability of good data in sufficient quantities to support valid interpretations and conclusions. Unfortunately, for various reasons (not the least of which is cost), good emissions data can be difficult to come by, both in terms of quality and amount. Admittedly, an extensive body of information about automotive emissions exists. All kinds of emissions data are generated for various purposes through a wide range of experimental and operational programs, and the collection of such information is ongoing, with funding from various sources. Yet it frequently is the case that the right data are seemingly unavailable when it comes time to make a difficult public policy decision. There are always interacting and confounding factors to consider, which policymakers who are not experts in all the physical mechanisms pertaining to automotive emissions can find difficult to disentangle. Further, emissions measurements exhibit some particular characteristics that often make their analysis less than straightforward.

In response to the obstacles presented by emissions data, statistical science is playing an expanding role in the establishment of emissions standards. A general recognition has emerged that statistical methods and models can provide the basis for making sound inferences about emissions effects and processes and that statistical thinking can provide the framework through which appropriate data collection is planned and executed. Nevertheless, statistical methods themselves can be misused or misapplied, thereby diminishing the benefits derived from their use and possibly even exacerbating the problems they are employed to resolve.

In recognition of the increasing role of statistical methods and modeling in the emissions arena, the Committee on Statistics and Statistical Software of the Transportation Research Board (TRB) of the National Research Council organized a mini-symposium of two technical sessions for presentation at the 1999 Joint Statistical Meetings (JSM) in Baltimore, Maryland. The American Statistical Association (ASA) sections on Statistics in the Physical and Engineering Sciences and Statistics in the Environment co-sponsored the two sessions with the TRB committee, chaired by Dr. Timothy Coburn of Abilene Christian University and Dr. Robert Mason of the Southwest Research Institute.

The principal goal of this mini-symposium was to foster interaction among individuals working on various statistical aspects of the emissions puzzle. The sense of the committee has been that, while much is being accomplished, it is being done by statisticians and engineers working in isolation who would greatly benefit from communication and exchange of ideas. A related objective was to promote greater overall consistency in the statistical treatment of emissions data and the use of appropriate methods. Finally, the committee sought to increase the dialog about automotive emissions within the professional statistical community for the purposes of stimulating new analytical approaches and enlightening policymakers about the need for statistical rigor.

This special issue of The *Journal of Transportation and Statistics* is devoted to the statistical analysis and modeling of automotive emissions. It contains many of the papers presented in the mini-symposium last August and also includes one additional manuscript submitted after the conference. The articles here represent the efforts of approximately 20 authors and co-authors from across industry, government, and academia and cover a diverse array of topics regarding fundamental methodological issues, advanced statistical techniques, and specific case studies. Two papers included in the mini-symposium but published elsewhere involved the assessment of sulfur in diesel fuel on the performance of emissions control devices and the forecasting of ozone standard exceedances that occur partly in response to vehicular traffic volume and dispersion.

The statistical analysis of automotive emissions is clearly a topic of current interest: the 1999 JSM is not the only recent venue to focus on it. For example, statistical applications were a major theme of the 2000 Spring International Fuels and Lubricants Meeting in Paris, co-sponsored by the Society of Automotive Engineers (SAE) and the Coordinating European Council (CEC), and several papers on the statistical analysis of emissions were presented there. Conferences sponsored by SAE and other professional organizations regularly include talks of this nature although they are not frequently organized into a single session or topical series. Likewise, since the early 1990s a substantial number of statistically oriented reports about automotive emissions have been published, and numerous related papers have appeared in a wide variety of technical journals. The growth in publications of this nature reflects both an increase in emissions research and the heightened emphasis on integrating the methods of statistical science. While by no means exhaustive, the bibliography provided on pages viii and ix is intended to be representative of the kinds and amount of statistical work that have been accomplished in this arena over the last ten years.

The use of statistical methods has contributed a great deal to an understanding of the origin and impact of automotive emissions, yet some important data-oriented problems remain largely unresolved. Many of these have to do with the sampling of vehicle populations. For example, determining the number and characteristics of all the vehicles in a large city so that a truly representative subset of those vehicles can be emissions tested is major undertaking. Other equally challenging problems include determining the correct sample size (the right number of vehicles) in the face of uncertainty and the spiraling cost of emissions testing, resolving the statistical requirements and operational difficulties of vehicle recruitment necessary to obtain an adequate sample, and knowing how to appropriately weight vehicle characteristics on the basis of actual usage or vehicle-miles traveled. Unresolved analytical questions exist as well, such as knowing how and when to mathematically transform emissions data to preserve distributional assumptions (e.g., use of the lognormal transformation), how to appropriately treat duplicate (or replicate) emissions measurements, and how to dispense with fixed and random factors in statistical emissions models.

It is the hope of the TRB committee that publication of this special issue of the *Journal* will serve to raise the level of awareness of these and similar issues. The goals are to foster development of the most realistic solutions possible to emissions-related problems and to provide the kinds of information necessary to produce improved decisionmaking and policy setting in the cross-disciplinary complex of energy, transportation, business, public health, and the environment.

TIMOTHY C. COBURN

Guest Editor

Abilene Christian University

Selected Bibliography

Alberini, A., W. Harrington, and V. McConnell. 1996. Estimating an Emissions Supply Function from Accelerated Vehicle Retirement Programs. *The Review of Economics and Statistics* 78:251–65.

Barth, M., T. Younglove, T. Wenzel, G. Scora, A. Feng, M. Ross, and J. Norbeck. 1997. Analysis of Modal Emissions from Diverse In-Use Vehicle Fleet. *Transportation Research Record* 1587:73–84.

Bazzani, R., M. Brown, K. Kuck, Y. Kwon, and M. Schmidt. 2000. The Effects of Driveability on Exhaust Emissions in European Gasoline Vehicles. SAE Technical Paper Series, 2000–01–1884. Warrendale, PA: Society of Automotive Engineers.

Bishop, G.A., D.H. Stedman, and L. Ashbaugh. 1996. Motor Vehicle Emissions Variability. *Journal of the Air & Waste Management Association* 46:667–75.

Burns, V.R., L.A. Rapp, W.J. Koehl, J.D. Benson, A.M. Hochhauser, J.C. Knepper, W.R. Leppard, L.J. Painter, B.H. Rippon, R.M. Reuter, and J.A. Rutherford. 1995. Gasoline Reformulation and Vehicle Technology Effects on Emissions—The Auto/Oil Air Quality Improvement Research Program. SAE Technical Paper Series, 952509. Warrendale, PA: Society of Automotive Engineers.

Cheng, W.K., S. Hockgreb, M.G. Norris, and K.-C. Wu. 1994. Auto-Oil Program Phase II Heavy Hydrocarbon Study: Analysis of Engine-Out Hydrocarbon Emissions Data. SAE Technical Papers Series, 941966. Warrendale, PA: Society of Automotive Engineers.

CleanFleet. 1995. Vehicle Emissions: Statistical Analysis Report no. 6. Columbus, OH: Battelle Memorial Institute.

Coburn, T.C. 1998. *Statistical Analysis of Particulate Matter Emissions from Light-Duty and Heavy-Duty Diesel Vehicles*. NREL/TP-540–24942. Golden, CO: National Renewable Energy Laboratory.

Coburn, T.C. 2000. Statistical Analysis of On-Road Particulate Matter Emissions from Diesel Vehicles. *Inhalation Toxicology* 12, suppl. 2:23–34.

Coburn, T.C., K.J. Kelly, and B.K. Bailey. 2000. Reduction in Vehicle Emissions Attributable to Alternative Transportation Fuels and its Prospective Impact on Air Quality and Public Health. *Journal of Occupational and Environmental Hygiene* 13:395–405.

Cohen, J. 1991. A Statistical Analysis of the Effect of Time-Shifting Automobile Emissions on Ambient Ozone Concentrations. *Transactions, 5th International Specialty Conference on Ozone* 19:850–861. Edited by R.L. Berglund, D.R. Lawson, and D.J. McKee. Pittsburgh, PA: Air & Waste Management Association.

Deaton, M.L. and J.J. Winebrake. 2000. Comparing Lifetime Emissions of Natural Gas and Conventional Fuel Vehicles: An Application of the Generalized ANCOVA Model. *Journal of the Air & Waste Management Association* 50:162–8.

Eide, I. and E. Johansson. 1994. Statistical Experimental Design and Projections to Latent Structures Analysis in the Evaluation of Fuel Blends with Respect to Particulate Emissions. *Chemometrics and Intelligent Laboratory Systems* 22:77–86.

Fomunung, I.S., S. Washington, and R. Guensler. 1999. A Statistical Model for Estimating Oxides of Nitrogen Emissions from Light-Duty Motor Vehicles. *Transportation Research D: Transport and Environment* 4/5:333–52.

Kelly, K., L. Eudy, and T. Coburn. 1999. *Light-Duty Alternative Fuel Vehicles: Federal Test Procedure Emissions Results*. NREL/TP-540–25818. Golden, CO: National Renewable Energy Laboratory.

Mason, R.L., A.C. Matheaus, T.W. Ryan, R.A. Sobotowski, J.C. Wall, C. Hobbs, G.W. Passavant, and T.J. Bond. 2000. EPA HDEWG Program—Statistical Analysis. SAE Technical Paper Series, 2000–01–1859. Warrendale, PA: Society of Automotive Engineers.

McAdams, H.T. 1995. A Random Balance Procedure for Simplifying a Complex Model, in *Proceedings of the American Statistical Association Section on Statistics and the Environment,* Orlando, FL, 13–17 August. Alexandria, VA: American Statistical Association.

McAdams, H.T., R.W. Crawford, and G.R. Hadder. 2000. A Vector Approach to Regression Analysis and its Application to Heavy-Duty Diesel Emissions. SAE Technical Paper Series, 2000–01–1961. Warrendale, PA: Society of Automotive Engineers.

Neill, W.S., W.L. Chippior, Ö.L. Gülder, J. Cooley, E.K. Richardson, K. Mitchell, and C. Fairbridge. 2000. Influence of Fuel Aromatics Type on the Particulate Matter and NOx Emissions of a Heavy-Duty Diesel Engine. SAE Technical Paper Series, 2000–01–1856. Warrendale, PA: Society of Automotive Engineers.

Orban, J.E. and H.-C. Tsai. 2000. Statistical Issues in the Evaluation of the Impact of Sulfur in Diesel Fuel on the Performance of Diesel Particulate Filter Emissions Control Devices. SAE Technical Paper Series, 2000–01–1958. Warrendale, PA: Society of Automotive Engineers.

Ouenou-Gamo, S., M. Ouladsine, and A. Rachid. 1998. Measurement and Prediction of Diesel Engine Exhaust Emissions. *ISA Transactions* 37:135–40.

Painter, L.J. and J.A. Rutherford. 1992. Statistical Design and Analysis Methods for the Auto/Oil Air Quality Research Program. SAE Technical Paper Series, 920319. Warrendale, PA: Society of Automotive Engineers.

Polanco, I. 1995. Using Design of Experiments to Control Auto Emissions. *IIE Solutions* 27:22–5.

Rozum, M.A. 1995. Appropriate Variable Selection in the EPZ Complex Model and CARB Predictive Model, in *Proceedings of the Section on Statistics and the Environment,* Orlando, FL, 13–17 August. Alexandria, VA: American Statistical Association.

Rutherford, J.A. and T. Crosby. 1993. The Statistical Value of Duplicates in Vehicle Emissions Testing. Paper presented at the Annual Meetings of the American Statistical Association, San Francisco, CA.

Schleyer, C.H., K.D. Eng, R.A. Gorse, and R.F. Gunst. 1999. Reversibility of Sulfur Effects on Emissions of California Low Emission Vehicles. SAE Technical Series Papers, 941544. Warrendale, PA: Society of Automotive Engineers.

Schuetzle, D., W.O. Siegl, T.E. Jensen, M.A. Dearth, E.W. Kaiser, R. Gorse, W. Kreucher, and E. Kulik. 1994. The Relationship Between Gasoline Composition and Vehicle Hydrocarbon Emissions: A Review of Current Studies and Future Research Needs. *Environmental Health Perspectives* 102, suppl. 4:1–12.

Sjögren, M., H. Li, U. Rannug, and R. Westerholm. 1996. Multivariate Analysis of Exhaust Emissions from Heavy-Duty Diesel Fuels. *Environmental Science and Technology* 30: 38–49.

Spreen, KB., T.L. Ullman, and R.L. Mason. 1995. Effects of Cetane Number, Aromatics, and Oxygenates on Emissions from a 1994 Heavy-Duty Diesel Engine with Exhaust Catalyst. SAE Technical Paper Series, 950250. Warrendale, PA: Society of Automotive Engineers.

Ullman, T.L, K.B. Spreen, and R.L. Mason. 1994. Effects of Cetane Number, Cetane Improver, Aromatics, and Oxygenates on 1994 Heavy-Duty Diesel Engine Emissions. SAE Technical Paper Series, 941020. Warrendale, PA: Society of Automotive Engineers.

_____. 1995. Effects of Cetane Number on Emissions from a Prototype 1998 Heavy-Duty Diesel Engine. SAE Technical Paper Series, 950251.Warrendale, PA: Society of Automotive Engineers.

Utts, J.M. and D.A. Niemeier. 1999. Multinomial Prediction Intervals for Micro-Scale Highway Emissions. *IMA Volumes in Mathematics and its Applications* 114:65–74.

Wang, W.G., D.W. Lyons, N.N. Clark, M. Gautam, and P.M. Norton. 2000. Emissions from Nine Heavy Trucks Fueled by Diesel and Biodiesel Blend Without Engine Modification. *Environmental Science and Technology* 34:933–99.

Washington, S. and R. Guensler. 1995. Statistical Assessment of Vehicular Carbon Monoxide Emission Prediction Algorithms. *Transportation Research Record,* no. 1472.

Webster, W.J. and C. Shih. 1996. A Statistically-Derived Metric to Monitor Time-Speed Variability in Practical Emissions Testing, in Proceedings of the Sixth CRC On-Road Vehicle Emissions Workshop, San Diego, CA. Atlanta, GA: Coordinating Research Council.

Westerholm, R. and K.-E. Egebäck. 1994. Exhaust Emissions from Light-and Heavy-Duty Vehicles: Chemical Composition, Impact of Exhaust and Treatment, and Fuel Parameters. *Environmental Health Perspectives* 102, suppl. 4:13–23.

Westerholm, R. and H. Li. 1994. A Multivariate Statistical Analysis of Fuel-Related Polycyclic Aromatic Hydrocarbon Emissions from Heavy-Duty Diesel Vehicles. *Environmental Science and Technology* 28:965–72.

Whalen, P., K. Kelly, and M.John. 1999. *Barwood CNG Cab Fleet Study: Final Results*, NREL/TP-540–26035. Golden, CO: National Renewable Energy Laboratory.

Winebrake, J.J. and M.L. Deaton. 1997. A Comparative Analysis of Emissions Deterioration for In-Use Alternative Fuel Vehicles. *Journal of the Air & Waste Management Association* 47:1291–96.

Winebrake, J.J., M.L. Deaton, T.C. Coburn, and K.J. Kelly. 2000. Statistical Analysis of Emissions and Deterioration Rates from In-Use, High Mileage CNG and Gasoline Ford Crown Victoria Taxi Cabs. SAE Technical Paper Series, 2000–01–1959. Warrendale, PA: Society of Automotive Engineers.

Wolf, J., R. Guensler, S.Washington, and W.Bachman. 1998. High-Emitting Vehicle Characterization Using Regression Tree Analysis. *Transportation Research Record* no. 1641:58–65.

Wright, C.M., D.E. Booth, S. Prasad, T.L. Isenhour, and K.N. Berk. 1996. A Joint Estimation Approach to Monitoring the Quality of Automobile Emissions, in *Proceedings of the American Statistical Association Section on Statistics and the Environment,* Chicago, IL, 4–8 August. Alexandria, VA: American Statistical Association.

Wu, D.-Y. and R.D. Matthews. 1998. The Texas Project: Conversions of Light-Duty Vehicles to CNG and LPG. Final Report. Golden, CO: National Renewable Energy Laboratory.

Yamaguchi, M. 2000. Japan Clean Air Program (JCAP)—Step I, Study of Gasoline Vehicle and Fuel Influence on Emissions. SAE Technical Paper Series, 2000–01–1972. Warrendale, PA: Society of Automotive Engineers.

Yanowitz, J., M.S. Graboski, L.B.A. Ryan, T.L. Alleman, and R.L. McCormick. 1999. Chassis Dynamometer Study of Emissions from Twenty-One In-Use Heavy-Duty Diesel Vehicles. *Environmental Science and Technology* 33:209–16.

Zemroch, P.J., P. Schimmerling, G. Sado, and C.T. Gray. 1996. European Programme on Emissions, Fuels and Engine Technologies (EPEFE)—Statistical Design and Analysis Techniques. SAE Technical Paper Series, 961069. Warrendale, PA: Society of Automotive Engineers.

Some Issues in the Statistical Analysis of Vehicle Emissions

TOM WENZEL

BRETT C. SINGER

Environmental Energy Technologies Division

Lawrence Berkeley National Laboratory

ROBERT SLOTT

Private Consultant

ABSTRACT

This paper presents some of the issues that complicate the statistical analysis of real-world vehicle emissions and the effectiveness of emissions control programs. The following issues are discussed: 1) inter- and intra-vehicle emissions variability, 2) skewness of the distribution of emissions from in-use vehicles, 3) the difficulty of obtaining statistically representative vehicle samples, 4) the influence of repeat testing on only a subset of the vehicle fleet, and 5) differences among common test methods and pollutant measurement devices. The relevance of these issues is discussed in light of three regulatory purposes: testing the compliance of in-use vehicles with certification standards, evaluating the effectiveness of vehicle inspection and maintenance programs, and estimating emissions inventories for air quality modeling and compliance planning. A brief history and description of common vehicle emissions tests is also provided.

INTRODUCTION

In recent years, emphasis on the measurement of vehicle emissions has shifted from laboratory testing towards the analysis of "real-world" emissions. The term "real-world" is used to differentiate

Tom Wenzel, Mail Stop 90–4000, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720. Email: TPWenzel@lbl.gov.

between the carefully controlled and limited environmental conditions and driving patterns associated with laboratory testing and those encountered on road. The "real-world" vehicle fleet is composed of new and aging vehicles with widely varying maintenance and operational histories and includes unregistered and out-of-state vehicles. By contrast, laboratory testing is often performed solely on new or well maintained vehicles that represent only a portion of the on-road fleet. Since the ultimate goal of vehicle emissions control devices and programs is to improve ambient air quality, analyses of program and technology effectiveness should focus as much as possible on real-world emissions reductions. Likewise, motor vehicle emissions inventories developed for air quality modeling and planning should accurately represent real-world fleets and conditions. This paper describes five major statistical issues that complicate the development of real-world vehicle emissions inventories, the evaluation of emissions control program effectiveness, and the process by which manufacturers certify that their vehicles are in compliance with emissions standards. Examples are given of how each of these statistical issues can complicate the analysis of emissions data. The presentation begins with a summary of the primary method used to measure vehicle emissions, the Federal Test Procedure, and alternative measurement techniques that have been developed in the last two decades.

MEASUREMENT TECHNIQUES

Several different techniques have been developed to measure vehicle emissions. Each of these techniques has strengths and weaknesses which should be considered when analyzing emissions measurements.

Federal Test Procedure

The first large-scale sampling of vehicle emissions was for the purpose of certifying manufacturer compliance with new-car emissions standards prescribed in the Clean Air Act Amendments (CAAA) of 1970. The U.S. Environmental Protection Agency (EPA) established an elaborate testing protocol, called the Federal Test Procedure (FTP), so that all vehicles could be tested under identical preparation and driving conditions. The FTP begins with overnight storage of the vehicle at a prescribed temperature in order to ensure that the engine and catalytic converter begin the test at this temperature. The vehicle is then rolled onto a treadmill-like device called a "dynamometer," where the vehicle is driven through a standard 30-minute speed/time trace, or "driving cycle." The FTP was designed in the early 1970s to simulate combined highway and city driving in urban Los Angeles. A top speed of only 57 mph and a top acceleration of only 3.3 mph per second were set to accommodate limitations of the dynamometers available when the test was developed. Tailpipe exhaust is mixed with a specified amount of dilution air and collected in large bags over three distinct portions of the driving cycle. The first bag captures the initial "cold start." "Hot stabilized" operation is captured in the second bag, and emissions following a "warm start" are measured in the third bag. Gas analyzers measure the concentrations of hydrocarbons (HC), carbon monoxide (CO), oxides of nitrogen (NO_x), and carbon dioxide (CO_2) in each bag. Concentration units relate the amount of each pollutant to the amount of total air collected (e.g., in percent or part per million (ppm) units). Mass emissions during each portion of the driving schedule are calculated as the product of the molecular mass and measured concentration of each pollutant and the total volume of air collected. Mass emissions are then related to the simulated distance traveled to yield gram per mile (gpm) emissions factors for each bag. The bag gpm emissions are then averaged together, weighted by the relative amount of driving under each section of the cycle, to achieve a composite gpm exhaust emissions rate. The FTP includes measurement of fuel evaporation during the driving cycle (running losses), for a short period after driving ceases (hot soak), and as the vehicle sits in an enclosed chamber during a multi-hour temperature cycle (diurnal).

Idle Testing

An idle emissions test measures pollutant concentrations in the tailpipe exhaust of a stationary vehicle. The test was proposed in the 1970 CAAA as a quick and inexpensive means to identify in-use vehicles with irregularly high emissions. Unlike the

FTP, idle testing includes no transient vehicle operation and no engine load. Idle testing is not used for NO_x emissions testing since NO_x emissions are always low during idle. HC and CO emissions during idle also may not be representative of emissions when a vehicle is driven under load. The 1977 CAAA required that all urban areas with poor air quality use idle testing in vehicle inspection and maintenance (I/M) programs. The first I/M programs used tailpipe probes to measure the concentrations of HC, CO, and CO_2 in the exhaust of idling vehicles. An enhancement of the basic idle test involves putting the car in neutral and revving the engine to 2500 rpm in an attempt to simulate the vehicle's emissions under loaded conditions.

IM240

The IM240 test uses 240 seconds of the FTP driving schedule to measure hot stabilized emissions during transient and loaded mode vehicle operation. It is the centerpiece of guidelines developed by the EPA to meet the Enhanced I/M program mandate of the 1990 CAAA. Enhanced I/M was designed to address several shortcomings of original I/M programs by 1) measuring emissions, including NOx, during loaded mode vehicle operation and 2) separating vehicle testing from vehicle repair by requiring a centralized network of contractor-run test-only facilities. Although desired for Enhanced I/M, no practical tests are available to measure evaporative HC emissions in an I/M setting. In the IM240 test, exhaust emissions are run directly through gas analyzers and can be quantified on a test-composite or a second-by-second basis. It was envisioned that the capability of analyzing second-by-second emissions would assist mechanics in properly diagnosing and repairing malfunctions of the emissions control system. Of the alternative emissions measurement techniques, the IM240 most closely resembles FTP testing. However, it is also the most time-consuming and costly test.

Acceleration Simulation Mode (ASM)

Many states resisted the use of centralized IM240 testing, citing the length of the test and the inconvenience to motorists of driving further to a small number of centralized test stations.[1] The California Bureau of Automotive Repair (BAR) developed an alternative test method to the IM240 called the Acceleration Simulation Mode (ASM) test. During an ASM test the vehicle is placed on a dynamometer and run at one or more distinct operating modes. These modes are defined as a certain vehicle load at a given speed; for instance, the California program gives each vehicle a 2525, 25% of the maximum vehicle load encountered on the FTP at 25 miles per hour, and a 5015, 50% of the maximum vehicle load encountered on the FTP at 15 miles per hour, ASM test. Emissions are measured in exhaust concentration using a tailpipe probe, just as in the idle test. The ASM test can be considered an improvement over the idle test in that emissions are measured when a vehicle is under load. However, the ASM does not measure emissions under varying loads and speeds, as does the IM240. In addition, NO_x emissions, which are not measured during idle testing, are measured under the ASM test. Eventually, EPA relaxed its requirement of centralized IM240 testing and allowed states to use alternative test methods such as the ASM if they could demonstrate that their alternative method would achieve the same reduction in emissions as the IM240.

Remote Sensing

In the late 1980s, researchers at the University of Denver developed a device to remotely measure the emissions of a vehicle as it is driven on the road (Bishop et al. 1989; Zhang et al. 1993). Remote sensors measure the changing intensity of a light beam directed across a roadway as the beam interacts with a passing vehicle's exhaust plume. The first generation sensors used an infrared source and a series of filters to isolate specific wavelengths that are absorbed by the CO, HC, and CO_2 in vehicle exhaust. A video camera placed alongside the remote sensor records each vehicle's license plate information, which is stored together with the emissions measurement. The license number can be

[1] Another factor was the political power wielded by the test and repair industry, which foresaw a centralized system displacing independent service stations that relied on I/M testing for a large portion of their business.

used to retrieve information about each vehicle (age, type, and perhaps mileage) from registration records. Remote sensors measure pollutant ratios, such as CO/CO_2 and HC/CO_2, but cannot measure absolute concentrations because the amount of exhaust dilution is not known. However, since more than 99% of fuel carbon atoms are emitted as CO, HC, or CO_2, the emissions ratios can be combined with known fuel properties (e.g., fuel carbon content) to calculate the mass of each pollutant emitted per gallon of fuel burned (Bishop et al. 1989; Zhang et al. 1993). Fuel-normalized emissions factors can be calculated for any emissions test, including the FTP, IM240, and ASM, as long as measurements of both CO and CO_2 are available. In recent years, remote sensors have been developed for the measurement of on-road emissions of NO and individual hydrocarbons or other emissions gases such as ammonia (Zhang et al. 1996; Jimenez et al. 1999b; Popp et al. 1997).

A single remote sensing instrument can measure emissions of thousands of vehicles per day for a fraction of the cost of conducting a similar number of idle, ASM, IM240, or FTP tests. In addition, the testing is unscheduled, so with an appropriately designed monitoring program actual on-road emissions can be measured from a large fraction of vehicles regularly in use without drivers taking steps prior to testing that would lower their vehicle's emissions. Remote sensors thus provide valuable data for estimating actual on-road emissions. Fuel-normalized emissions factors have been measured for tens of thousands of vehicles throughout the Los Angeles area. These factors have been combined with fuel sales data to estimate total exhaust emissions of the on-road vehicle fleet (Singer and Harley 1996; Singer and Harley 2000). However, there are limitations to remote sensing. The instrument accurately measures the emissions of a given vehicle as it is being driven for a fraction of a second only, and, therefore, overall emissions for the measured vehicle may differ considerably from those measured by one remote measurement. As a result, a single remote sensing measurement should not be regarded as indicative of typical emissions for any individual vehicle. In general, remote sensing is most valuable at providing data on fleet-average emissions or typical emissions from a certain vehicle model or type. Repeat measurements of individual vehicles can be used to identify high- or low-emitting vehicles.

One concern about the use of remote emissions data is that the vehicle driving condition (or load) at the time of measurement is unknown. To address this issue, remote sensors have been sited to measure emissions from vehicles under a known driving condition, often while driving uphill under moderate load. It is also becoming commonplace to measure roadway grade, along with vehicle speed and acceleration, at the time of each remote emissions measurement. The driving mode can be estimated by a calculation of the physical load encountered by the vehicle as a result of aerodynamic drag, tire rolling resistance, inertial and gravitational acceleration forces, and engine friction (Ross 1994; Jimenez et al. 1999a; Singer 1998). To address concerns about measuring emissions during cold start driving, remote sensors are sometimes located on highway off-ramps or on surface thoroughfares that cannot be accessed directly from residential streets.

On-Board Diagnostics

A new technology that has the potential to contribute important information about vehicle emissions is the on-board diagnostic (OBD) computer system required on all new cars sold after 1995. These systems were designed by manufacturers in response to regulations by the California Air Resources Board (CARB) and EPA. The OBD system is designed to monitor over 50 parameters of vehicle and engine operation. If the on-board computer detects malfunctions or operations that would lead to tailpipe emissions greater than 1.5 times the certification standard, the system stores a "fault" code in the computer and turns on a "malfunction indicator light" (MIL) on the dashboard to alert the driver. The intent of the OBD regulations is twofold: to encourage drivers to bring their vehicles in for inspection and repair as soon as problems are detected and to record engine parameters to assist mechanics in diagnosing and repairing malfunctions. The regulations have had additional beneficial results. The OBD systems have encouraged manufacturers to design better and more durable engine and emissions controls, including more extensive monitoring and backup systems.[2] In addition, OBD systems are identifying manufacturing flaws on individual vehicles before they leave the plant. In the next few years, EPA is expected to require that all states operating Enhanced I/M programs fail vehicles with illuminated MILs. CARB anticipates that OBD will eventually replace the periodic emissions testing in conventional I/M programs. A drawback to OBD systems is that they do not measure tailpipe emissions directly; rather, they predict when emissions are likely to exceed standards, based on extensive monitoring of engine and emissions control parameters. Therefore, the usefulness of OBD data is currently limited to determining failure rates of the vehicle fleet. However, there is some discussion about eventually requiring later generations of OBD systems to directly measure tailpipe emissions.

STATISTICAL ISSUES

This section describes and discusses five major statistical issues that complicate the analysis of in-use vehicle emissions.

Inherent Variability in Vehicle Emissions

Real-world vehicle emissions are highly variable. Emission variability from vehicle to vehicle spans several orders of magnitude, while the emissions of most vehicles will vary substantially with environmental and driving conditions. Emissions of some vehicles are unrepeatable: different emissions occur from one test to another, even when test conditions are carefully controlled.

Vehicle-emission variability is a consequence of the way emissions are generated and how they are controlled. Exhaust emissions are formed in the engine as a result of unburned fuel, HC, and partially burned fuel, HC and CO, and from undesirable side reactions, NO_x. Emissions control systems are designed to reduce pollutant formation in the engine and to chemically convert engine-out pollutants to less harmful products in the catalytic converter. When functioning properly, modern vehicle-emissions controls reduce tailpipe emissions levels to five percent or less of those observed from pre-control vehicles produced in the late 1960s. However, if the engine or the emissions control system fails to operate as designed, exhaust emissions may rise by orders of magnitude.

■ ■ ■

There are numerous factors affecting the variability in emissions across different vehicles. Several of these are discussed below.

Vehicle Technology. The increasingly stringent new-car emissions standards specified in the CAA Amendments of 1970, 1977, and 1990 have been met primarily through technological improvements. Emission control technologies incorporated into vehicles over the past 30 years included the use of exhaust gas recirculation to reduce NO_x formation in the engine, the addition of catalytic converters for exhaust gas treatment, the replacement of carburetors with throttle-body and port fuel injection, and computer control of air-fuel mixing and spark timing. In most cases, these and other vehicle-emission control improvements have been introduced to the entire new car fleet over just a few model years. Real-world emissions are sensitive to vehicle technology independent of vehicle age.

Vehicle Age and Mileage Accumulation. As vehicles age and accumulate mileage, their emissions tend to increase. This is both a function of the normal degradation of emissions controls of properly functioning vehicles, resulting in moderate emissions increases, and malfunction or outright failure of emissions controls on some vehicles, possibly resulting in very large increases in emissions, particularly CO and HC.

Vehicle Model. Some vehicle models are simply designed and manufactured better than others. Some vehicle models and engine families are observed to have very low average emissions while others exhibit very high rates of emissions control failure (Wenzel 1997). The design of a particular emissions control system affects both the initial

[2] Manufacturers wish to prevent the MILs from turning on since they want to maintain customer satisfaction and are responsible for the cost of repairing emissions malfunctions in new cars under warranty.

effectiveness and the lifetime durability of the system, which in turn contributes to a model-specific emissions rate.

Maintenance and Tampering. The degree to which owners maintain their vehicles by providing tune-ups and servicing according to manufacturer schedules can affect the likelihood of engine or emissions control system failure and therefore tailpipe emissions. Outright tampering with vehicles, such as removing fuel tank inlet restrictors to permit fueling with leaded fuel that will degrade the catalytic converter or tuning engines to improve performance, can have a large impact on emissions. Early I/M programs relied on visual inspection to discourage tampering. The advent of sophisticated on-board computers and sensors has greatly reduced the incentive to improve vehicle performance through tampering. In fact, tampering with the sophisticated electronics installed on today's vehicles will likely reduce performance as well as increase emissions. Requirements for extended manufacturer warranties have led to vehicle designs that are less sensitive to maintenance, at least within the warranty period. Nonetheless, there is evidence that maintenance can still affect real-world emissions from new vehicles, at least on some models (Wenzel 1997). Improper maintenance or repair can also lead to higher emissions.

Misuse. The cumulative effect of hard driving, or "misuse," of a vehicle can also increase emissions. For example, prolonged high power driving, such as repeated towing of a trailer up mountain grades, leading to high engine temperatures can cause premature damage to the catalytic converter, resulting in dramatic increases in emissions.

Type of Malfunction. There are many emissions control components that can malfunction or fail. Some of these malfunctions are interrelated; for instance, the onboard computer of a vehicle with a failed oxygen sensor may command a constant fuel enrichment, which can eventually lead to catalyst failure. Different component malfunctions result in very different emissions consequences. In general, malfunctioning vehicles with high CO emissions tend also to have high HC emissions, while vehicles with high NO_x emissions tend to have relatively low CO and HC emissions (Wenzel and Ross 1998).

Socioeconomics. Correlations have been observed between average vehicle emissions and socioeconomic indicators, such as the median household income in the zip code where vehicles are registered (see Singer and Harley forthcoming). This relationship results in part because the vehicle fleet is older in lower income areas. However, even after accounting for vehicle age, average emissions are higher in lower income areas than higher income areas. Even vehicles of the same age and engine family exhibit different failure rates and average emissions when tested at I/M stations located in lower vs. higher income areas (Wenzel 1997). There are three possible explanations for this phenomenon: 1) individual vehicles that have been poorly manufactured (i.e. perform poorly or frequently require repairs) are selectively sold by higher income individuals and eventually wind up in lower income areas, 2) less money is spent on maintenance and repairs in lower income areas, and 3) vehicles with higher mileage are more likely to "migrate" to lower income owners.

■ ■ ■

Different factors account for the variability in an individual vehicle's emissions.

Intermittent Emissions Control Failure. While some emissions control failures, such as a completely degraded catalyst, can lead to high emissions during all vehicle operation, other failures can be intermittent. For example, a vehicle with a partially degraded catalyst may have lower emissions under higher loads because the catalyst may be effective only at very high temperatures. Oxygen sensor, fuel delivery system, and computer malfunctions can also be intermittent. Intermittent control system malfunction can cause large changes in emissions from test to test, even when all of the factors listed below are held constant. This results in uncertainty in the average emissions from such a vehicle.

Driving Mode/Engine Load. Vehicle emissions can vary greatly with changing engine load. The relationship between emissions and load depends on the fuel-delivery and emissions-control technology, but as a general rule NO_x emissions almost always increase with increasing load. Under high speed and acceleration requirements, today's vehicles are designed to have excess fuel injected into the engine cylinder. This "enrichment" of the air/fuel mixture leads to elevated CO and HC formation during combustion, with no oxygen available for pollutant conversion to CO_2 and water in the catalyst. The result is a temporary "puff" of high tailpipe CO and HC emissions (Goodwin and Ross 1996). In some vehicles, fuel injection is cut off during rapid decelerations. This can lead to cylinder misfire and a temporary "puff" of high HC emissions (An et al. 1997). Roadway grade and accessory use, such as air conditioning and heaters, put additional loads on the engine and can affect emissions. Small changes in how a vehicle is driven can also affect emissions. For instance, how a driver shifts gears on a vehicle with a manual transmission or how smoothly a driver depresses and releases the accelerator, may affect emissions rates (Shih et al. 1997).

Engine and Catalyst Temperature. When a vehicle is initially started after more than a few minutes of nonoperation, emissions are temporarily high because both the catalytic converter and oxygen sensor are ineffective at low temperatures. Heated by vehicle exhaust, the devices reach the high temperatures required for their operation after one to four minutes of driving. The temporary control system ineffectiveness at start-up is exacerbated by higher pollutant formation in "cold" engines and commanded fuel enrichment designed to facilitate ignition. The magnitude of cold start emissions depends on the time since the vehicle was last operated, ambient temperature, and the operation of the vehicle after starting.

Ambient Temperature and Humidity. Ambient temperature has a large direct effect on evaporative HC emissions. Very low ambient temperatures (e.g., below 20 degrees Fahrenheit) can influence emissions at ignition and cause the catalysts of some vehicles to cool during short stops. Very high ambient temperatures can have a secondary influence on exhaust emissions because engine load is increased by air conditioner use. Effects can include higher NO_x and an increase in the frequency of commanded enrichment. The amount of water vapor in air can affect NO_x emissions in older and malfunctioning vehicles, but it appears to have less effect on new vehicles with computer engine control.

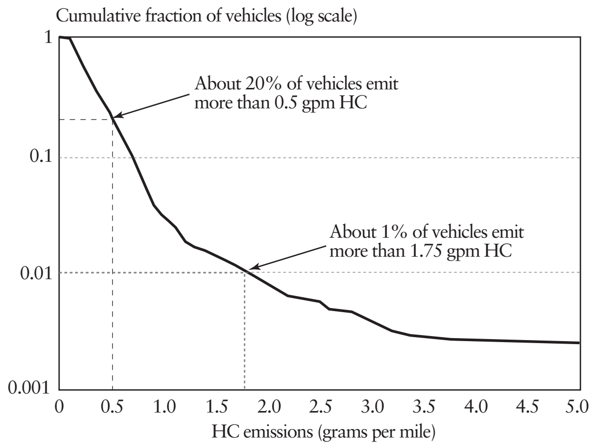
Fuel Quality. Fuel composition can have a substantial impact on vehicle tailpipe and evaporative emissions. Regulations may require changes in fuel composition by season within a region as a strategy to reduce emissions. For instance, some urban areas introduce oxygenates in fuel to reduce CO emissions in the winter and decrease the volatility to reduce evaporative HC emissions in the summer. Fuel composition can vary spatially since some regions in the country have been required or have chosen to adopt year-round reformulated gasoline standards as an emissions-control strategy.

■ ■ ■

The FTP calls for careful control of fuel and the conditions under which vehicles are tested to control for each of these factors (some factors are more tightly controlled than others). Even under these carefully controlled conditions, vehicle emissions can be quite variable (Bishop et al. 1996). A study of repeat FTP tests on the same vehicles found that CO and HC emissions from malfunctioning vehicles can change by over a factor of seven on independent FTP tests although the uncertainty is much less for properly functioning vehicles (Knepper et al. 1993). The emissions of vehicles exhibiting high uncertainty are difficult to characterize for regulatory and modeling purposes.

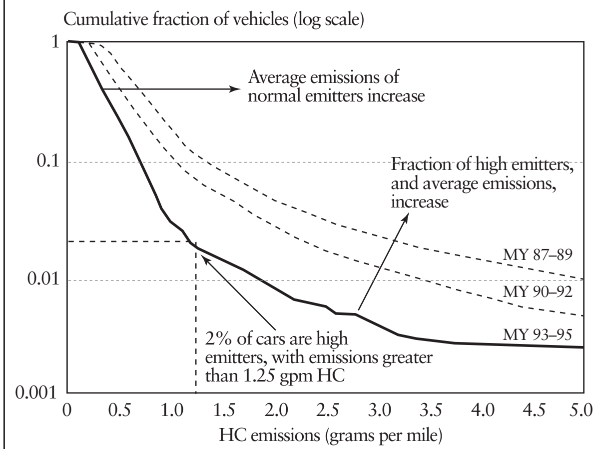
Most of the factors affecting variability and uncertainty in vehicle emissions are widely recognized. However, the degree to which some of these factors affect emissions has not been adequately quantified.

FIGURE 1 An Example of the Distribution of Vehicle HC Emissions



HC emissions from MY 1993–95 fuel-injected cars with 100,000 to 150,000 miles, from 1996–98 Wisconsin IM240 tests (fast pass emissions adjusted to full IM240 emissions).

FIGURE 2 Trends in HC Emissions Distributions as Vehicles Age



HC emissions from MY 1993–95 fuel-injected cars with 100,000 to 150,000 miles, from 1996–98 Wisconsin IM240 tests (fast pass emissions adjusted to full IM240 emissions).

Distributional Assumptions about Emissions Data

The distributions of emissions from large numbers of vehicles are highly skewed. The majority of vehicles have relatively low emissions, while a relatively small number of malfunctioning vehicles have extremely high emissions (Lawson et al. 1990; Stephens 1994; Bishop et al. 1996; Barth et al. 1999; Schwartz 2000). To overcome this difficulty, analysts have typically used the forms of the log-normal (Stephens 1994) and gamma (Zhang et al. 1994) distributions to model vehicle emissions data.

One graphical tool for analyzing this kind of data is to plot emissions as a function of the cumulative fraction of vehicles, as shown in figure 1. The figure shows the fraction of vehicles, on the y-axis, with emissions above a given level on the x-axis. For example, in figure 1, about 20% of the vehicles have HC emissions greater than about 0.5 grams per mile (gpm), while 1% of the vehicles have HC emissions greater than 1.75 gpm. With degradation of emissions controls, the average emissions of normal emitters increase, as shown in figure 2; that is, the upper/left part of the distribution shifts to the right. An increase in the fraction of high emitters, as well as an increase in the average emissions of high emitters, causes the lower/right segment of the distribution to shift upward and become flatter. The change in the shape of the distribution, shown in figure 2 approximately at 1.25 gpm for MY 93–95, can be taken as a cut point for dividing vehicles of the same age and model year into "normal emitters" and "high emitters." The shape of the emissions distribution may vary by pollutant, vehicle type, vehicle age, and so forth.

Since, in many cases, vehicle emissions approximately follow a log-normal or gamma distribution, confidence intervals on the mean emissions level are not symmetric. Also, statistical tests, such as t-tests, which depend on normality cannot be used to determine whether the difference in mean emissions from two groups of vehicles is statistically significant unless sample sizes are large. Further, the emissions of different pollutants, or different samples of vehicles, may not necessarily follow the same type of distribution.

As previously suggested, the logarithmic transformation is frequently used to account for the non-normality of the data; yet this may not be the appropriate approach to take. For example, emissions inventory models developed by EPA and CARB multiply estimates of the mean emissions of a group of vehicles by estimates of activity, such as miles driven and number of starts, of that group of vehicles. However, if the mean emissions are calculated based on the logarithmic transformation, then the emissions of any high emitting vehicles in the sample are given much less weight in the estimated mean emissions level, and the models tend to underestimate fleet emissions (Pollack et al. 1999b).

Other approaches to the problem of non-normality have been taken, with varying degrees of success. One way to construct an approximately normal distribution is to consider a collection of average values representing fairly large, unbiased subsets of emissions measurements. Stedman et al. (1997) demonstrated the usefulness of this method in the context of remote sensing measurements taken over a five-day period. First, the average emissions measured by remote sensing for each day were calculated. On the basis of the well-known Central Limit Theorem, the five averages should be approximately normally distributed if the samples measured over each of the five days were unbiased and sufficiently large. The five averages were then averaged to obtain an estimate of fleet-average emissions about which a symmetric confidence interval could be constructed. Normal statistical tests, such as the t-test, were then applied (Stedman et al. 1997). Some researchers are beginning to use nonparametric techniques, such as bootstrap sampling (Pollack et al. 1999a; Frey et al. 1999), since such techniques do not require an assumption regarding the distribution of the underlying population.

Although the skewed nature of vehicle emissions distributions is generally acknowledged, proper statistical tools are not always used to characterize the uncertainty associated with mean emissions levels.

Representativeness of Test Vehicles

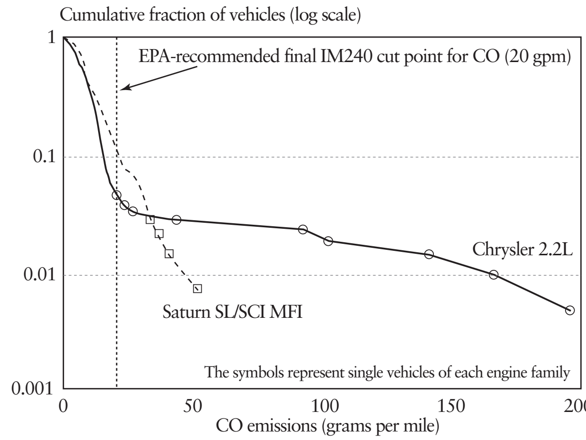
The skewed nature of vehicle emissions also has important implications for drawing a representative sample of vehicles from a population for testing. Because vehicle emissions vary by the factors discussed above (vehicle age, technology, make and model, owner socioeconomic characteristics, etc.), a representative sample of vehicles would account for all of these factors. There are two issues to consider when seeking a vehicle sample that is representative of the in-use fleet: the number of vehicles and selection/response bias.

Number of Vehicles

Because there are relatively few high emitters in the population, the sample needs to be large enough so that a number of high emitters is included. As noted above, the number of vehicles required depends on the constituent of interest and the shape of its distribution, as well as the statistical hypothesis to be tested.

The issue of inadequate sample size is demonstrated in EPA's in-use compliance program, which

FIGURE 3 CO Emissions Distributions of Two Car Models



142 Saturn CL/SCI cars with multiport fuel injection and 217 Chrysler cars with 2.2-liter engine. From 1996–97 Arizona IM240 tests (fast pass emissions adjusted to full IM240 emissions).

attempts to identify vehicle engine families that have high average in-use emissions for recall and repair by the manufacturer. Under this program, a very small sample (not more than a dozen) of three-to five-year-old in-use vehicles is recruited and its emissions tested under FTP conditions.[3] An engine family with average emissions in excess of the new-car certification standards may be subject to an emissions-related recall.

One limitation of the program is that not enough vehicles of a particular engine family are tested to identify models with a small number of extremely high emitters. Figure 3 demonstrates this situation, using cumulative vehicle distributions of CO emissions for two model-year 1991 engine families, the Saturn SL/SCI with multi-port fuel injection and the Chrysler 2.2 liter engine. The figure shows the emissions distribution of at least 100 individual vehicles from each of these engine families, tested on the IM240 in the Arizona I/M program. Both the Saturn and the Chrysler engine families have the same average CO emissions, 12 grams per mile. However, the figure indicates that the Saturns have relatively high emissions across all

[3] To satisfy legal requirements, vehicles must be "well-maintained and used." This requirement introduces a sampling bias into the test program because manufacturers consult the service history of individual vehicles, using data supplied by their dealers, before agreeing to include vehicles in the test sample.

vehicles tested but no extremely high emitters. The Chryslers, in contrast, have relatively low emissions on the low end of the distribution but several vehicles with extremely high emissions. The Saturn engine family shown in figure 3 failed EPA in-use compliance testing for CO and was recalled, whereas the Chrysler engine family did not. However, the potential reduction in emissions from repairing the high-emitting Chryslers is greater than that of repairing the Saturns. The design of the in-use compliance program identifies for recall engine families with marginally high emissions across all vehicles, rather than engine families with a small number of vehicles with extremely high emissions.

Selection/Response Bias

For detailed FTP testing, agencies usually recruit vehicles by mailing solicitations to potential participants. Participation in such testing programs typically is voluntary although incentives are frequently provided to encourage participation. Both CARB and EPA primarily use mail solicitation to obtain vehicles for data to feed their emissions inventory models. The output from these models is input into regional air quality models to forecast the effect of emissions control programs on future air quality. One large source of uncertainty in the vehicle emissions inventory models is the potential for selection bias in voluntary vehicle recruitment.[4]

The makeup of the vehicle sample is likely affected by the perceived rewards and penalties for participating in the study. Rewards typically include cash, use of a rental vehicle, and, sometimes, repairs to the vehicle. Penalties may include inconvenience, risk of damage to the vehicle, and possible future requirements to repair the vehicle at the owner's expense. It can be argued that voluntary recruitment programs where high-emitting vehicles are repaired free of charge may attract a disproportionate number of dirty vehicles. Voluntary recruitment programs using mail solicitation typically achieve a response rate of only 10 to 15% (CARB 1997). In fact, one study where the registration of vehicles not responding to mail solicitations was to be suspended still achieved a response rate of only 60% (CARB 1996).

The recruitment of vehicles with high emissions is particularly difficult. Many recruited vehicles likely to have high emissions cannot be tested because the condition of the vehicle (e.g., bald tires or fuel, oil, coolant, or exhaust leaks) would threaten the safety of the technicians performing the test. The degree to which testing programs change the condition of the vehicle prior to testing may also affect the emissions test results. For example, a long list of restorative maintenance procedures (such as replacing spark plugs, air filters, mufflers, distributor caps and rotors, and adjusting ignition timing) is performed on cars to be tested for compliance with in-use emissions standards (CARB 1994). In contrast, very little of this restorative maintenance is performed on vehicles recruited for "as received" emissions testing.

EPA, CARB, and others acknowledge the possibility of selection bias in voluntary vehicle procurement programs, but few studies have been conducted to estimate its effect. An analysis of remote sensing readings of emissions from vehicles whose drivers were asked to participate in a roadside pullover program conducted in California in 1994 found that vehicles whose owners refused to participate had, on average, 2.5 times higher on-road emissions than those of owners who did agree to participate (Stedman et al. 1994). However, a similar experiment conducted as part of a more recent roadside pullover testing program found that vehicles whose owners declined to participate had the same average remote sensing emissions by model year as those whose owners agreed to participate (Wenzel et al. 2000).

One method of estimating the selection bias in vehicle recruitment via mail solicitation is to conduct a formal experiment where the emissions of vehicles whose owners volunteer to participate are compared with emissions of vehicles that are unavailable or whose owners decline to participate. The emissions for this comparison would come from each vehicle's next (or last) regularly scheduled I/M test. The experiment would need to be conducted in a state that has a loaded mode,

[4] The National Academy of Sciences recently published a review of EPA's emissions inventory model summarizing this source of bias and other limitations of the model (NAS 2000).

centralized I/M program, such as Arizona, Colorado, Illinois, or Wisconsin. For example, if the expected acceptance rate of a voluntary mail solicitation test program is 10% and 100 vehicles are desired, 1,000 invitations would normally be mailed. To determine the effect of selection bias, one could instead mail 10,000 invitations. Owners of about 9,000 vehicles would not be available or would decline to participate; 1,000 owners would agree to participate, 100 of whose vehicles would actually be brought in for FTP testing. I/M emissions from the 900 vehicles that were volunteered but not chosen would be compared with a randomly selected subset of the 9,000 vehicles that were not volunteered by their owners for testing.

Other methods can be used to obtain emissions measurements of relatively unbiased samples of vehicles. As discussed above, CARB has conducted roadside testing of vehicles randomly pulled over by law enforcement officers. Vehicle I/M programs are designed to measure the emissions of virtually all vehicles registered in an urban area. Finally, remote sensing instrumentation allows the unscheduled testing of nearly all vehicles that drive by the sensors. However, none of these methods is entirely free from sample bias. For legal reasons, roadside testing in California must be voluntary. Roadside and remote sensing studies only measure emissions of vehicles that happen to drive by the measurement sites, as noted previously, and each method has siting limitations. Finally, not all registered vehicles report for I/M testing.[5] In addition, these alternative methods for measuring vehicle emissions have other drawbacks, as discussed earlier. Researchers should consider the advantages and disadvantages of each measurement method when designing an emissions collection program.

The Effect of Repeated Testing

In I/M programs, vehicles that fail initial testing are supposed to be repaired and then retested until they pass the test. Vehicles are therefore characterized by the results of their initial test: vehicles with low emissions are passed and not retested, while vehicles with high emissions are retested, presumably after repairs, until they pass. However, as discussed above, emission levels of many vehicles, particularly those with intermittent malfunctions, can vary substantially from one test to the next, and, consequently, the average emissions of the fleet of failing vehicles may be lower in subsequent testing, even without any repairs, solely due to emissions variability. Likewise, the average emissions of the fleet of passing vehicles may be higher in subsequent testing due to emissions variability. For the same reason, the average emissions of the fleet of vehicles that failed their initial test will be higher if they were retested after their final passing test. Evaluations of I/M programs that do not account for the variability observed in repeated testing of only a portion of the vehicle fleet may overstate the emissions benefits of these programs.

Little effort has been made to determine what the effect of repeated testing has on the estimated effectiveness of an I/M program. The primary reason is insufficient data: it is not often the case that large numbers of vehicles are repeatedly tested under identical conditions. The easiest way to remedy this situation, of course, would be to conduct an experiment using multiple tests on the same vehicles under identical conditions. Many states have the capability of conducting full IM240 tests on a random sample of the vehicles that report for testing, and this capability allows states to collect the type of data needed. For example, Heirigs and Gordon (1996) report the results of an experimental program in Arizona in which back-to-back IM240 tests were conducted on a sample of vehicles that failed their initial I/M test after waiting at least 15 minutes for a test lane to open.

Testing Methodologies

The methodology used to measure vehicle emissions should be taken into account when analyzing emissions measurements. Although the type of test method used is not strictly a statistical issue, it is

[5] Older and newer vehicles often are exempted from I/M testing, and many eligible vehicles do not report for testing. For example, up to 26% of the vehicles in the Phoenix I/M program that failed their initial I/M test between January 1996 and July 1997 did not receive a final passing test within 3 to 15 months of their initial test. Of these vehicles, about one-third were still driving in the I/M area more than two years after their initial test (Wenzel 1999).

important to consider when analyzing emissions data, particularly when comparing measurements made with different instrumentation or under different methods.

Units of Measurement

Tailpipe emissions can be reported as exhaust concentrations (e.g., percent or ppm), normalized to the amount of fuel used (e.g., grams per gallon), or normalized to the distance traveled (e.g., grams per mile). The relationship between exhaust concentrations and fuel-normalized emissions factors is approximately linear except for extremely high CO and HC emitters (Singer 1998). In contrast, relating fuel-normalized to mileage-normalized emissions factors requires knowledge or assumptions about fuel efficiency. This issue is relevant when attempting to use non-FTP methodologies to evaluate in-use compliance with new car emissions standards expressed in grams per mile because driving mode directly affects fuel efficiency. At the extreme, mileage-normalized emissions are infinite under idle conditions when the vehicle is not moving; therefore, mileage-normalized emissions can be very high for driving cycles that include significant amounts of idle time. Likewise, the fuel economy measured during one set of driving conditions (e.g., the FTP) may differ from fuel economy under a different set of conditions (e.g., the fixed load conditions of ASM testing). Additional uncertainty thus results when using EPA-reported fuel efficiency values to convert remote sensing, ASM, or idle test results (in concentration or grams per gallon) to grams per mile.

Testing Methodology

The various test methodologies described earlier measure emissions during different vehicle driving modes and potentially widely varying environmental conditions. Even the IM240 and FTP, both dynamometer-based methodologies that include controlled, transient vehicle operation, involve different combinations of engine loads. Also, unlike the FTP, environmental conditions and vehicle preparation are generally not predefined or controlled for IM240 tests. In addition, since the purpose of I/M programs is to merely identify, and eventually repair, high-emitting vehicles and not necessarily to accurately measure every vehicle's emissions, EPA allows the IM240 test to be varied for exceptionally clean or dirty vehicles. For example, clean vehicles may pass their I/M test after only 30 seconds of driving, while exceptionally dirty vehicles may fail after 94 seconds of testing. Application of these "fast pass" and "fast fail" rules vary from state to state. The use of shorter test cycles complicates comparisons of fleet average emissions and emissions reductions because the driving patterns of the shortened tests differ from that of the full IM240, and the effect of uncontrolled environmental conditions and vehicle pretest conditioning are more pronounced. This suggests that great care should be taken when comparing emissions measured using different test methods, and/or under different test conditions. Some researchers have developed factors to convert emissions measured in I/M programs to projected emissions under FTP test conditions. These factors are developed by running regression models on the measured I/M and FTP emissions using a relatively small sample of vehicles tested under both test conditions (see Austin et al. 1997 and DeFries and Williamson 1997). However, such factors are only valid on a fleet-average basis, not for the emissions of individual vehicles (DeFries et al. 1999). Another approach is to compare instantaneous emissions measured during a specified engine load (Jimenez 1999a; McClintock 1999), which would allow remote sensing measurements, for example, to be compared with FTP, IM240, and even to ASM emissions test results.

Pollutant Measurement Equipment

The same basic physical principles and analytical equipment are used to measure CO and CO_2 concentrations in the FTP bags, from the tailpipe probes used in idle and IM240 I/M testing, and by roadside remote sensing. Thus, while the uncertainty of any CO measurement obtained by remote sensing may be higher than that of an FTP bag measurement (due to a lower signal and more interference in the remote measurement), results of the two tests may still be directly compared, all other factors being equal. This is not the case for HC and NOx. For HC, there is an important difference between the infrared (IR) technique used by remote sensors and for tailpipe I/M testing and the flame ionization detector (FID) systems used during FTP testing. FID systems essentially count carbon atoms and provide equivalent results on a per-carbon basis for individual hydrocarbon compounds with different structures. Infrared HC analyzers measure infrared light absorption at a wavelength specific to the carbon-hydrogen bond structure typical of n-alkanes (compounds like propane, butane, and hexane). FIDs and IR analyzers are both typically calibrated with propane standards. However, an infrared analyzer calibrated with propane will report only a fraction of the carbon atoms from hydrocarbon compounds that have different structures than propane (e.g., benzene, toluene, and ethene, all of which are major components of HC emissions in vehicle exhaust). The relationship between IR and FID measurements of exhaust HC depends on the relative amounts of each HC compound in a vehicle's exhaust, known as HC speciation, and the particular wavelength filter used in the infrared analyzer. On a fleet-average basis, infrared analyzers used for vehicle exhaust measurement (including remote sensors) report only about 50% of the HC emissions that would be reported by a FID measurement on the same exhaust sample (Singer et al. 1998). The disparity can vary from 20 to 80% for individual vehicles, depending on the distribution of HC species in the tailpipe exhuast, a function of the driving mode and the condition of the catalyst.

CONCLUSIONS

The applicability of several different methods in measuring real-world vehicle emissions has been described. In addition, several issues that complicate the statistical analysis of real-world vehicle emissions have been presented. For example, selection bias is often apparent in the recruitment of vehicles, but emissions professionals have few means to operationally or statistically remedy the situation. In addition, the data necessary to estimate the direction or degree of any such bias are often unavailable. Consequently, to meet the challenges inherent in analyzing vehicle emissions in such a way as to effectively have an impact on public policy and environmental quality, access to much more information is needed. In addition, statistical rigor and innovative approaches will be necessary in the design of experimental programs and the analysis of resulting emissions data. Better decisionmaking will likely only result through interdisciplinary cooperation among emissions professionals, engineers, scientists and members of the statistical community.

REFERENCES

Austin, T.C., L.S. Caretto, T.R. Carlson, and P.L. Heirigs. 1997. Development of a Proposed Procedure for Determining the Equivalency of Alternative Inspection and Maintenance Programs. Report prepared for U.S. EPA Office of Mobile Sources, Regional and State Programs Division, Sierra Research. Ann Arbor, MI: USEPA.

An, F., G. Scora, M. Barth, and J. Norbeck. Characterization and Modeling of Vehicle Lean-Burn Hydrocarbon Emissions. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Barth, M., F. An, T. Younglove, C. Levine, G. Scora, M. Ross, and T. Wenzel. 1999. The Development of a Comprehensive Modal Emissions Model. Final report, NCHRP Project 25–11. Washington, DC: Transportation Research Board, National Cooperative Highway Research Program.

Bishop, G.A., J.R. Starkey, A. Ihlenfeldt, W.J. Williams, and D.H. Stedman. 1989. IR Long-Path Photometry: A Remote Sensing Tool for Automobile Emissions. *Analytical Chemistry* 61, no. 10:671–7.

Bishop, G.A., D.H. Stedman, and L. Ashbaugh. 1996. Motor Vehicle Emissions Variability. *Journal of the Air & Waste Management Association* 46: 667–75.

California Air Resources Board (CARB). 1996. Comparison of the IM240 and ASM Tests in CARB's I&M Pilot Program. El Monte, CA.

_____. 1997. Test Report of the Light-Duty Vehicle Surveillance Program, Series 13 (LDVSP13). Mobile Source Operations Division, Report Number MSD-97–01. El Monte, CA.

DeFries, T.H. and H.J. Williamson. 1997. Methodology for Estimating California Fleet FTP Emissions. Report prepared for California Bureau of Automotive Repair by Radian International, Austin, TX.

DeFries, T.H., C.F. Palacios, S.D. Kishan, and H.J. Williamson. 1999. Models for Estimating California Fleet FTP Emissions from ASM Measurements. Draft report prepared for the California Bureau of Automotive Repair by Eastern Research Group, Austin, TX.

Drachand, K.D. of California Air Resources Board (CARB). 1994. Memorandum to Tom Cackette. June 23.

Frey, H.C., J. Zheng, and A. Unal. Quantitative Analysis of Variability and Uncertainty in Highway Vehicle Emissions Factors. Paper presented at the Ninth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1999.

Goodwin, R. and M. Ross. 1996. Off-Cycle Emissions from Modern Passenger Cars with Properly Functioning Emissions Controls. SAE Technical Paper Series, 960064. Warrendale, PA: Society of Automotive Engineers.

Heirigs, P.L. and J. Gordon. 1996. Preconditioning Effects on I/M Test Results Using IM240 and ASM Procedures. SAE Technical Paper Series, 962091. Warrendale, PA: Society of Automotive Engineers.

Jimenez, J.L., P.M. McClintock, G.J. McRae, D.D. Nelson, and M.S. Zahniser. 1999a. Vehicle Specific Power: A Useful Parameter for Remote Sensing and Emissions Studies. Paper presented at the Ninth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1999.

Jimenez, J.L., M.D. Koplow, D.D. Nelson, M.S. Zahniser, and S.E. Scmidt. 1999b. Characterization of On-Road Vehicle NO Emissions by a TILDAS Remote Sensor. *Journal of the Air & Waste Management Association* 49:463–70.

Knepper, J.C., W.J. Koehl, J.D. Benson, V.R. Burns, R.A. Gorse Jr., A.M. Hochhauser, W.R. Leppard, L.A. Rapp, and R.M. Reuter. 1993. Fuel Effects in Auto/Oil High Emitting Vehicles. SAE Technical Paper Series, 930137. Warrendale, PA: Society of Automotive Engineers.

Lawson, D.R., P. Groblicki, D.H. Stedman, G.A. Bishop, and P.L. Guenther. 1990. Emisions of In-Use Motor Vehicle in Los Angeles: A Pilot Study of Remote Sensing and the Inspection and Maintenance Program. *Journal of the Air & Waste Management Association* 40:1096–105.

McClintock, P. 1999. Remote Sensing Measurements of Real World High Exhaust Emitters, CRC Project No. E-23-Interim Report. Tuscon, AZ: Remote Sensing Technologies.

National Academy of Sciences (NAS). 2000. *Modeling Mobile Source Emissions*. Washington, DC: National Academy Press.

Pollack, A.K., P. Bhave, J. Heiken, K. Lee, S. Shepard, C. Tran, G. Yarwood, R.F. Sawyer, and B.A. Joy. 1999a. Investigation of Emissions Factors in the California EMFAC7G Model. Report prepared for Coordinating Research Council, Project E-39.

Pollack, A., T. Stoeckenius, and C. Tran. 1999b. Memorandum to Phil Enns, EPA OMS, regarding MOBILE6 deterioration rates. 18 August.

Popp, P.J., G.A. Bishop, and D.H. Stedman. Development of a High-Speed Ultraviolet Spectrophotometer Capable of Real-Time NO and Aromatic Hydrocarbon Detection in Vehicle Exhaust. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Ross, M. 1994. Automobile Fuel Consumption and Emissions: Effects of Vehicle and Driving Conditions. *Annual Review Energy Environment* 19:75–112.

Schwartz, J. 2000. Smog Check II Evaluation: Part II: Overview of Vehicle Emissions. Final Report. Sacramento, CA: California Inspection and Maintenance Review Committee.

Shih, R., S. Fable, and R.F. Sawyer. Effects of Driving Behavior on Automotive Emissions. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Singer, B.C. 1998. A fuel-based approach to estimating motor vehicle exhaust emissions. Ph.D. diss., University of California, Berkeley, CA.

Singer, B.C. and R.A. Harley. 1996. A Fuel-Based Motor Vehicle Emissions Inventory. *Journal of the Air & Waste Management Association* 46:581–93.

Singer, B.C., R.A. Harley, D. Littlejohn, J. Ho, and T. Vo. 1998. Scaling of Infrared Remote Sensor Hydrocarbon Measurements for Motor Vehicle Emissions Inventory Calculations. Environmental Science & Technology 53, no. 21:3241–48.

Singer, B.C. and R.A. Harley. 2000. A Fuel-Based Inventory of Motor Vehicle Exhaust Emissions in the Los Angeles Area during Summer 1997. *Atmospheric Environment* 34:1783–95.

Stedman, D.H., G.A. Bishop, S.P. Beaton, J.E. Peterson, P.L. Guenther, I.F. McVey, and Y. Zhang. 1994. On-Road Remote Sensing of CO and HC Emissions in California. California Air Resources Board, Contract No. A032–093.

Stedman, D.H., G.A. Bishop, P. Aldrete, and R.S. Slott. 1997. On-Road Evaluation of an Automobile Emission Test Program. *Environmental Science & Technology* 31: 927–31.

Stephens, R.D. 1994. Remote Sensing Data and a Potential Model of Vehicle Exhaust Emissions. *Journal of the Air & Waste Management Association* 44:1284–92.

Wenzel, T. 1997. I/M Failure Rates by Vehicle Model. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Wenzel, T. 1999. Evaluation of Arizona's Enhanced I/M Program. Paper presented at the Ninth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1999.

Wenzel, T. and M. Ross. 1998. Characterization of Recent-Model High-Emitting Automobiles. SAE Technical Paper Series, 981414. Warrendale, PA: Society of Automotive Engineers.

Wenzel, T., B. Singer, E. Gumerman, R. Sawyer, and R. Slott. 2000. Evaluation of the Enhanced Smog Check Program: A Report to the California Inspection and Maintenance Review Committee. Final Report. Berkeley, CA: LBNL.

Zhang, Y., D.H. Stedman, G.A. Bishop, P.L. Guenther, S.P. Beaton, and J.E. Peterson. 1993. On-Road Hydrocarbon Remote Sensing in the Denver Area. Environmental Science and Technology 27, no. 9:1885–91.

Zhang, Y., G.A. Bishop, and D.H. Stedman. 1994. Automobile Emissions are Statistically y-Distributed. Environmental Science and Technology 28:1370–4.

Zhang, Y., D.H. Stedman, G.A. Bishop, S.P. Beaton, P.L. Guenther, and I.F. McVey. 1996. Enhancement of Remote Sensing for Mobile Source Nitric Oxide. Journal of the Air & Waste Management Association 46:25–9.

Validity of Chase Car Data Used in Developing Emissions Cycles

JENNIFER E. MOREY

THIRAYOOT LIMANOND

Institute of Transportation Studies

University of California, Davis

DEBBIE A. NIEMEIER

Department of Civil and Environmental Engineering

University of California, Davis

ABSTRACT

In an effort to ensure vehicle compliance with U.S. air quality policies, driving cycles, profiles of average driving behavior, have been constructed to characterize the driving behavior of the overall fleet. The cycles are built from chase car data, speed-time profiles of in-use vehicles recorded using a chase car method. This study evaluates the acceptability of using chase car data as the foundation for driving cycle development and recommends changes in the current data collection protocol. Two data issues are closely examined: 1) the effectiveness of the current target vehicle selection procedure and 2) the validity of blending data collected from target vehicles with data collected from the chase car, a method used when target vehicles are unavailable. Although in the aggregate there do not appear to be significant discrepancies between these chase car and target vehicle data, when examined at disaggregate levels, significant differences appear that could affect the representativeness of existing driving cycles. Recommendations include increasing the proportion of target to chase car data in future databases by improving the existing protocol and considering the use of different recording technology.

Debbie A. Niemeier, Department of Civil and Environmental Engineering, University of California, Davis, One Shields Avenue, Davis, CA 95616. E-mail: dniemeier@ucdavis.edu

INTRODUCTION

Driving cycles, profiles of average driving behavior, are used to certify new vehicles, to verify vehicle compliance with inspection/maintenance (I/M) programs, and to create emissions factors for performing transportation conformity determinations. Although there may be no single representative driving cycle, characterizing average driving behavior is a very important element in describing overall fleet emissions. Numerous data have been collected to create these driving cycles. To date, two data collection methods have most often been employed: 1) the use of a chase car to mimic driving behavior while recording speed and acceleration data from "target" vehicles sampled from the population and 2) the use of onboard instrumentation in vehicles to record speed and acceleration data. Chase car data have primarily been used for developing driving cycles, while data from instrumented vehicles have been used only minimally. The use of other technologies, such as Global Positioning Systems (GPS), for collecting driving behavior data remains some time away from wide-scale implementation.

Given the importance of driving cycles to estimating mobile emissions, it is worth examining how the data are collected and the representativeness of the data for driving cycle development. Accordingly, this study has three objectives: 1) to assess the robustness of chase car data at a much finer resolution than previously examined, 2) to evaluate the appropriateness of mixing chase and target car data to develop the so-called composite driving cycles, and 3) to evaluate and recommend changes to minimize potential cycle construction biases that can arise as a result of chase car data collection procedures.

BACKGROUND

It is well established that the Federal Test Procedure (FTP), the foundation used for estimating mobile-source emissions inventories, does not adequately reflect normal driving patterns. Although still used in EPA-developed cycles, in 1990 the California Air Resources Board (CARB) initiated a project to develop new driving cycles to better represent actual driving behavior, thus improving mobile source emissions modeling (Gammariello and Long 1996). As part of this effort, driving data, specifically speed-time profiles, were collected on roadway networks in the Greater Metropolitan Los Angeles area in April and May of 1992. The resulting database is known as LA92, and data collection was accomplished using a chase car protocol.[1]

The LA92 chase car protocol was a refined version of procedures previously developed by General Motors (GM) and the U.S. Environmental Protection Agency (EPA). GM's approach involved a chase car following a vehicle from trip beginning to trip end and attempting to mirror the target vehicle's major speed changes, accelerations, and decelerations (Austin et al. 1993). The GM chase car was equipped with instrumentation for recording its own operations but not, however, with technology that allowed the recording of accurate estimates of target vehicle operations. Instead, the accuracy of the data hinged on the ability of GM chase car drivers to correctly match the speed and acceleration of target vehicles. In addition, the method was limited in that it did not account for the effects of changing road grades (Austin et al. 1993). The EPA protocol and equipment were similar to that of GM and also produced relatively imprecise speed-time profiles.

Two primary concerns arose with respect to this method. First, the potential for detection and resultant behavioral change by the target car driver was considered problematic. Second, the crudeness of acceleration and deceleration event measures likely resulted in inaccuracies in the data (Sierra Research 1997). To improve the resolution of previous measurements, the chase car used to collect data in the LA92 study was equipped with a range-finder laser designed to measure the relative distance between the chase car and target vehicles. A video camera, mounted inside the chase car, recorded the view through the windshield to provide a visual check for assessing data reliability. Presumably, the LA92 database more accurately captures the behavior of drivers in Los Angeles.

LA92 has been used by both CARB and EPA to create various driving cycles. The new EPA-constructed facility-specific cycles were developed

[1] A chase car protocol generally refers to the procedures used to identify a target vehicle and initiate speed-time data collection.

using the LA92 data in combination with the 1992 chase car data gathered in Spokane, Washington (S92) and in Baltimore, Maryland (B92), while CARB has used the LA92 data to develop the Unified Cycle (UC). The EPA speed-based cycles use speed, acceleration frequency, trip length, and level of service (LOS) variables to define cycles. The UC is a trip-based cycle that incorporates average speed, acceleration frequency, and trip length variables but does not include level of service. CARB has also constructed Unified Correction Cycles (UCCs) using LA92 to adjust for speeds between 10 and 50 mph. Since LA92 contains limited data at higher and lower speed ranges, CARB supplemented it with data from a separate 1992 EPA database that contained additional data collected in Baltimore, Spokane, and Atlanta using vehicles with onboard instrumentation. The EPA-instrumented vehicle data were used to supplement data for speeds between 0 and 10 mph and 55 and 75 mph.[2]

CURRENT STUDY METHODOLOGY

To evaluate the robustness of the chase car data and the appropriateness of mixing chase and target data, the present study considers three areas of concern: 1) potential inaccuracies in the data introduced by the current chase car protocol and equipment; 2) variation in the amount of data collected in Baltimore, Los Angeles, and Spokane and the data's representativeness of each region's traffic conditions; and 3) differences in driving behavior data recorded from target vehicles and from the chase vehicle when no targets are available, in other words, under "non-lock" conditions.

Potential Inaccuracies in the Data Due to the Current Chase Car Protocol and Equipment

Briefly, the current protocol, developed for the LA92 study, directs chase car drivers to collect second-by-second speed-time profiles from hundreds of target vehicles. The chase car follows predefined routes, "locking on" to target vehicles with the range-finder laser while simultaneously collecting data on such variables as road grade, type of vehicle targeted, road facility type, and level of service in addition to speed and acceleration (Austin et al. 1993). A full description of the chase car protocol can be found in Austin et al. 1993.

The way the chase car protocol is actually implemented during data collection can substantially affect the development of driving cycles. There are several potential problems that can cause the application of the protocol to vary. For the purposes of this study, there are two critically important instructions in the current chase car protocol: 1) the procedure for target car selection on busy surface streets and freeways and 2) the procedure for data collection under non-lock conditions.

The first procedure's objective is to ensure random vehicle selection. When chase car drivers enter a new roadway, the procedure instructs them to follow the first forward vehicle encountered in the same lane as the closest white vehicle. Specifically, the closest white vehicle is defined as "the closest white vehicle in front of an imaginary line passing through the center of the chase car and perpendicular to the direction of travel" (Austin et al. 1993, 53). If the chase car is in the same lane as a white vehicle and more than one white vehicle is present, "one car length is subtracted per 10 mph of speed before deciding which white vehicle is the closest" (Austin et al. 1993, 53).

The field application of this selection process can be complex under rapidly changing traffic conditions, making its execution very difficult. Review of videotapes recorded during the LA92 data collection indicated the procedure for selecting target vehicles was inconsistently applied, particularly when the chase car entered a new roadway. Many targets were not acquired even though the video suggested it was possible to do so. This appeared to be in part due to confusion about which vehicle should be chosen according to the target vehicle selection procedure.

With respect to the second procedure, chase car drivers are told to "drive in a fashion that approximately matches the general flow of through traffic," driving "faster than some vehicles and slower than a similar number of vehicles" in the absence of target vehicles (Austin et al. 1993, 54). In this

[2] It should be noted that although the LA92, B92, and S92 data sets were collected using the same protocol, to date only LA92 has been used on its own for driving cycle development.

case, the chase car records its own operating data with the range-finder laser disengaged. These data are then used to replace missing target data in the final "composite" database.

This use of chase car data in lieu of missing target car data is intended to increase the sample size available for building driving cycles. Target data and non-lock chase car data are joined together in series to create the composite data set. These composite data are then used to create Speed Acceleration Frequency Distributions (SAFDs), the cornerstone of driving cycle development. Since approximately 47% of LA92, 58% of S92, and 63% of B92 come from the chase car operations rather than the target vehicle, the driving behavior of chase car drivers and their ability to approximate the speed and acceleration of other vehicles become very influential. This is particularly true in light traffic conditions when there may be few vehicles to emulate. It should be noted that both the target and non-lock chase car data are recorded on a second-by-second basis, providing hundreds of profiles (realizations) of the sampling unit, that is, the driver-vehicle. When chase car data are used in place of missing target data, it does not increase the overall sample size of the data set. Instead, only one driver-vehicle profile, that of the chase driver, is added to the sample, increasing the sample size by one. Since this single profile will contain more speed-time data points than those of the target vehicles, there is great potential for the chase vehicle data to bias cycles developed from the data.

The choice of technology used for data collection can also have a significant impact on the successful application of the chase car protocol. The chase car has a built-in speed measurement system that records speed at every second with a precision of 0.38 mph (Austin et al. 1993). The range-finder laser installed behind the grill of the chase car emits 400 light pulses per second. When the laser beam bounces off the target vehicle, the time it takes for the signal to return to a receptor in the chase car's grill determines the distance between the two vehicles. The laser system was tested on static targets, yielding a distance accuracy within one foot. This presumably leads to a corresponding target vehicle speed error of 2 feet per second or 1.36 mph when the chase car is in motion (Austin et al. 1993). The potential errors, then, of the chase car speed measurement system and the range-finder laser together yield an error of ±1.74 mph in the estimated target vehicle speed.

This error is reasonable when estimating speed, but the impacts in terms of acceleration are less clear. If forward differencing is used to determine accelerations on a second-by-second basis (acceleration equals velocity at the second second minus velocity at the first second), the estimate of a target vehicle's acceleration could be off by as much as 3.48 mph/s.[3] While this represents the most extreme case, it illustrates how measurement errors could cause a target vehicle to appear to be accelerating when it is, in fact, at cruise.

Regional Differences in the Data

To examine data differences between cities, we analyzed the B92, S92, and LA92 composite data using two variables: level of service (LOS) and facility type. The B92 data include 191,119 seconds of data, representing 218 routes; S92 contain 175,137 seconds, encompassing 249 routes; and LA92 contain 102 chase car runs, resulting in 100,709 records (seconds) of data.

Variation in the Amount of Data Collected in Each Level of Service

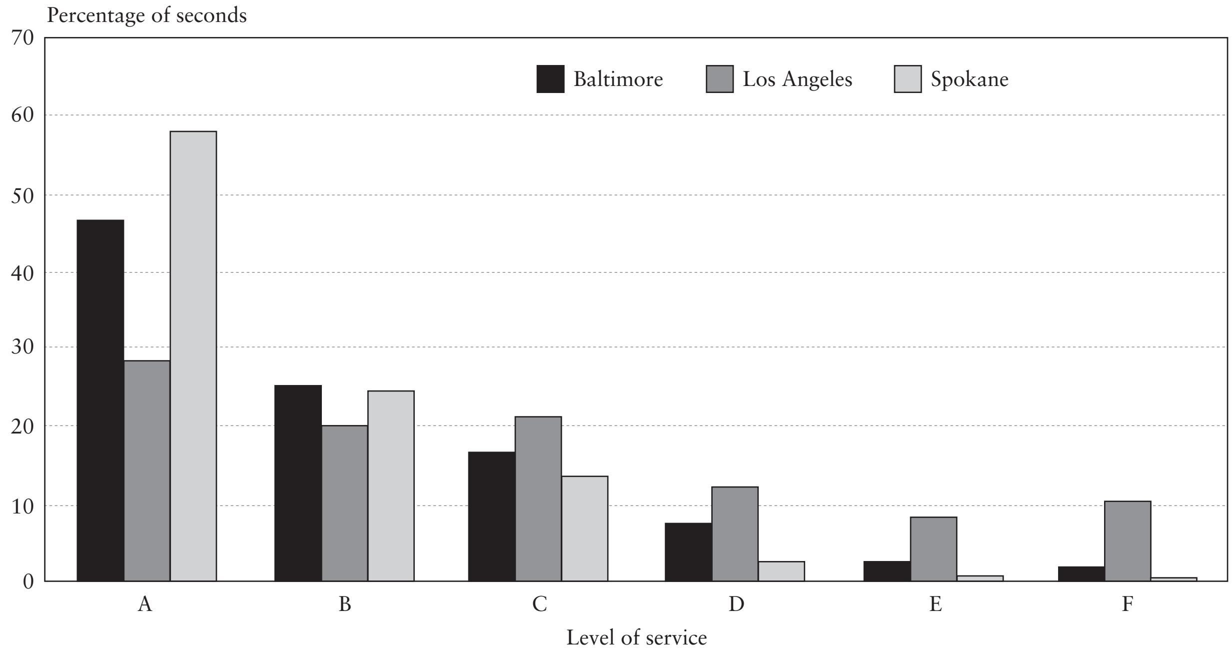
Level of service (LOS) refers to traffic density conditions observed on a specific facility at a specific time. In each of the chase car studies, the chase car observer visually assigned a level of service category (A, B, C, D, E, or F).[4] The observer used a switchbox mounted on the chase car's dashboard to manually record the level of service (Austin et al. 1993). Figure 1 presents the percentage of composite data by time collected at each LOS by city.

Despite differences in size, location, and availability of public transportation, Baltimore and Spokane exhibit similar amounts of time in levels

[3] Imagine that the velocity at the second second has an error of +1.74 mph, while the velocity at the first second has an error of –1.74 mph.

[4] These levels correspond to the levels of service given in the Highway Capacity Manual. Level A describes free-flow conditions, while Level F describes stop-and-go conditions. Levels B, C, D, and E represent levels of increasing traffic congestion.

FIGURE 1 Percentage of Seconds Spent in Each Level of Service (LOS) by City



of service B and C. Perhaps more interesting, however, is the difference in time spent at levels of service E and F between Baltimore and Los Angeles. The chase car recorded almost 3.5 times the amount of data in LOS E in Los Angeles as in Baltimore and approximately 6 times more data in LOS F. These numbers suggest substantial differences in traffic congestion level between cities.

One qualitative means to assess city to city difference in congestion is to use the Roadway Congestion Index (RCI) (Texas Transportation Institute 1998). The RCI is calculated as:

$$\mathrm{RCI} = \frac{\mathrm{FwyVMT}/\mathrm{LnM} \times \mathrm{FwyVMT} + \mathrm{ArtVMT}/\mathrm{LnM} \times \mathrm{ArtVMT}}{13{,}000 \times \mathrm{FwyVMT} + 5{,}000 \times \mathrm{ArtVMT}} \tag{1}$$

where

- FwyVMT is the estimated vehicle-miles traveled on the chosen area's freeways.
- LnM is the estimated lane-miles of roadway.
- ArtVMT is the estimated vehicle-miles traveled on principal arterial streets.

The constants 13,000 and 5,000 indicate the capacity of the facility type, in this case freeway and arterial. The RCI is an indicator of average congestion over an entire metropolitan region and is used extensively by public officials.

According to the 1992 RCI, Los Angeles had a score of 1.54, while Baltimore had a score of 1.04 (Texas Transportation Institute 1998). Since chase car data should ideally characterize the driving behavior of the population and relative congestion levels within a metropolitan region, it is possible to use the ratio of different RCI scores as a basis for comparing the relative levels of service represented by the chase car data. On a relative basis, if driving data on congested roadways were sampled more often than indicated by the RCI ratios, the data would not be representative of average driving conditions in the area. Using this logic, LA92 should only contain about 1.5 times more records in levels of service E and F than B92. This suggests that there may be some unevenness in how level of service between cities was assigned or, alternatively, that different traffic levels were not appropriately sampled. Other studies have also suggested that the visually determined LOS may not represent level of service target statistics assigned by the Highway Capacity Manual's methods (Niemeier et al. 1998).

TABLE 1 Descriptive Statistics on Composite Speeds by Facility Type and Level of Service (LOS)

Fac./LOS	Baltimore Min.	Baltimore Max.	Baltimore Mean	Baltimore St. dev.	Los Angeles Min.	Los Angeles Max.	Los Angeles Mean	Los Angeles St. dev.	Spokane Min.	Spokane Max.	Spokane Mean	Spokane St. dev.
Arterial												
All LOSs	0	74.50	25.41	18.04	0	63.10	21.48	15.56	0	74.90	26.53	17.01
LOS A	0	74.50	30.97	17.77	0	61.10	22.65	15.21	0	74.90	29.25	16.92
LOS B	0	65.00	23.60	17.49	0	63.10	22.90	15.85	0	69.50	26.39	16.99
LOS C	0	63.15	20.30	16.60	0	59.10	21.24	15.65	0	56.68	18.95	14.35
LOS D	0	71.29	17.51	15.15	0	53.00	18.35	15.12	0	53.40	15.14	13.13
LOS E	0	56.77	12.15	14.77	0	47.69	13.39	13.74	0	34.50	15.21	11.22
LOS F	0	43.20	11.79	12.02	0	44.20	11.15	12.54	0	36.90	5.51	8.45
Ramp												
All LOSs	0	72.69	40.35	17.06	0	76.00	29.83	20.25	0	79.10	37.23	17.87
LOS A	0	69.80	42.50	15.49	0	61.10	28.88	19.16	0	79.10	37.67	18.79
LOS B	0	71.15	39.62	15.32	0	76.00	35.44	20.17	0	61.80	38.60	13.93
LOS C	0	68.52	42.52	15.61	0	65.70	30.42	20.96	0	63.70	33.28	19.72
LOS D	0	72.69	44.04	19.69	0	65.70	24.31	21.33	—	—	—	—
LOS E	15.31	47.60	28.50	9.69	0	63.70	38.59	16.05	—	—	—	—
LOS F	0	42.40	4.31	8.69	0	51.50	20.90	15.58	—	—	—	—
Freeway												
All LOSs	0	80.91	56.12	13.21	0	80.30	44.75	20.31	0	83.15	59.10	8.85
LOS A	0	80.91	59.39	7.07	29.60	76.36	57.27	8.72	0	83.15	62.06	9.86
LOS B	0	77.30	59.13	8.57	28.00	75.60	62.72	6.14	0	73.83	58.20	7.12
LOS C	0	75.84	59.24	8.34	13.72	80.30	60.85	7.86	0	70.70	57.00	7.19
LOS D	0	75.80	57.13	10.49	6.10	73.97	56.42	8.83	31.57	60.04	54.92	6.79
LOS E	0	71.30	44.37	16.33	0	69.50	38.89	17.42	24.95	55.70	41.72	6.34
LOS F	0	69.47	23.81	16.46	0	66.00	23.03	13.82	—	—	—	—

— Missing data

Differences Across Cities within Each Recorded Level of Service

Table 1 contains estimated standard errors and other descriptive statistics for all levels of service represented in the three databases across three main facility types (arterial/collectors, freeways, and ramps). From the table, it can be seen that the mean speed for level of service F on arterial/collectors is very similar for Baltimore and Los Angeles but differs substantially from that recorded in Spokane; that is, the composite data for Baltimore and Los Angeles show approximately twice the mean speed as Spokane and about 1.5 times the standard deviation. The ramp facility type shows even more pronounced differences between mean speeds in Baltimore and Los Angeles at all levels of service, particularly under more congested conditions. For example, the mean speed in Los Angeles' level of service F on ramps is nearly five times that recorded during Baltimore's level of service F on ramps. Further, the associated standard deviation for Los Angeles is almost twice the magnitude of that for Baltimore.

The RCI and the differences in mean speeds and standard deviations, however, suggest that there may be some unevenness in how level of service between cities was assigned or, alternatively, that different traffic levels were not appropriately sampled. The accurate reflection of level of service is particularly important because the EPA facility-based cycle depends on level of service as one of its key variables.

Target versus Non-Lock Chase Car Data on Each Facility Type

We also investigated the relative "lock-on" rates of chase to target vehicles with respect to different facility types and different levels of service. Lock-on rates indicate how much of the data in each database actually comes from target vehicles as opposed to chase car operations. When viewed by facility type, it is clear that lock-on rates vary between cities and, as such, might be expected to vary dramatically between facility types. Table 2 compares the percentages of data recorded from target vehicles with those recorded from the non-lock chase car in each city by facility type, using each city's composite data.

As previously noted, approximately 47% of the data in LA92 originates as non-lock chase car records, while approximately 58% of the Spokane data and 63% of Baltimore data come from non-lock chase car records. The implication is that the facility types with low lock-on rates are extremely dependent on the ability of a chase car driver to accurately mimic prevailing traffic conditions and/or to drive like the "average driver" if there are no other vehicles around. Lock-on rates are relatively low on private roads, local roads, and, to some extent, ramps and arterials/collectors, suggesting few targets or difficult terrain conditions. With few target vehicles on the road, the chase car driver necessarily has difficulty gauging how to drive "with the general flow," as described in Austin et al. (1993, 53).

Equally important is the time spent on each roadway type. For example, although Los Angeles high occupancy vehicle (HOV) lanes show 100% lock-on, when the overall time spent in HOV lanes is examined, the limitations of the collected data become apparent. Table 3 shows the percentage lock-on for LA92 according to the share of total time recorded by facility type.

In table 3, the time-weighted lock-on rates are the actual amount of target data as a share of the aggregate data. It is apparent that the largest combined share of time and lock-on occurs on arterial/collectors, 29.4%. It also is apparent that target vehicle data on ramps are very limited. Although target vehicle data make up 37.8% of the data recorded on this facility type, ramps represent a mere 5.4% of the seconds in the overall data set.

TABLE 2 Lock-On Rates and Non-Lock Chase Car Data by Facility Type Using Composite Data Sets

City	Facility type	"Lock-on" rate (percentage of data from target vehicles)	"Non-lock" rate (percentage of data from chase vehicle)	Total
Baltimore	Private road	4.2	95.8	100.0
	Local road	3.7	96.3	100.0
	Arterial/collector	38.3	61.7	100.0
	Ramp	23.9	76.1	100.0
	Freeway	61.8	38.2	100.0
	HOV lane	66.7	33.3	100.0
	Aggregate[a]	**37.2[b]**	**67.8**	**100.0**
Los Angeles	Private road	0.0	100.0	100.0
	Local road	1.5	98.5	100.0
	Arterial/collector	45.2	54.8	100.0
	Ramp	37.8	62.2	100.0
	Freeway	76.9	23.1	100.0
	HOV lane	100.0	0.0	100.0
	Aggregate[a]	**53.1**	**46.9**	**100.0**
Spokane	Private road	5.0	95.0	100.0
	Local road	5.3	94.7	100.0
	Arterial/collector	41.9	58.1	100.0
	Ramp	27.9	72.1	100.0
	Freeway	74.9	25.1	100.0
	HOV lane	—	—	—
	Aggregate[a]	**42.3**	**57.7**	**100.0**

[a] "Aggregate" refers to each city's composite data set without distinction between facility type.
[b] 1,458 seconds of data in the Baltimore database were undefined and, therefore, left out of this analysis.
— Missing data.

TABLE 3 Percentage of Target Data in Los Angeles as a Share of Total Time on Each Facility Type and in the Aggregate

Facility type	Lock-on rate (percentage of data from target vehicles)	Time on each facility type (percentage of total seconds)	Percentage of time-weighted lock-on rate
Private road	0.0	0.2	0.0
Local road	1.5	1.4	0.02
Arterial/collector	45.2	65.0	29.4
Ramp	37.8	5.4	2.0
Freeway	76.9	28.1	21.6
HOV lane	100.0	0.0001	0.0001
Overall rate	53.1	100.0	53.1

Multiplying 37.8% by 5.4% reveals that only 2.0% of the composite data were recorded from target vehicles on ramps. Since accelerations necessarily occur at these locations, ramps are important in defining mobile emissions. Target car data from these facilities would be much more useful for emissions purposes than non-lock chase car data.

Differences in Driving Behavior Between Drivers of Target and Chase Vehicles

To conduct side by side comparisons of the driving behavior of chase and target car drivers, we examined speed-time traces in the LA92, B92, and S92 data. Some interesting results emerged from this analysis. Two of these results are illustrated in the speed-time trace constructed from the LA92 data, figures 2a and 2b.

The thin line in the figure represents the non-lock chase car's speed-time trace, while the bold line represents the target vehicle's trace. At the point marked "A" in figure 2a, a target car abruptly appears in the data and is abruptly cut off, forming a hook shape. At the point marked "B" in figure 2b, the chase car accelerates just before acquiring the target and then begins to slow to match the target's speed.

We hypothesized that these peculiarities were caused from loss of vehicle lock and/or a chase car attempting to catch up to a prospective target. Review of the LA92 videotape revealed the following possible explanations for these anomalies.

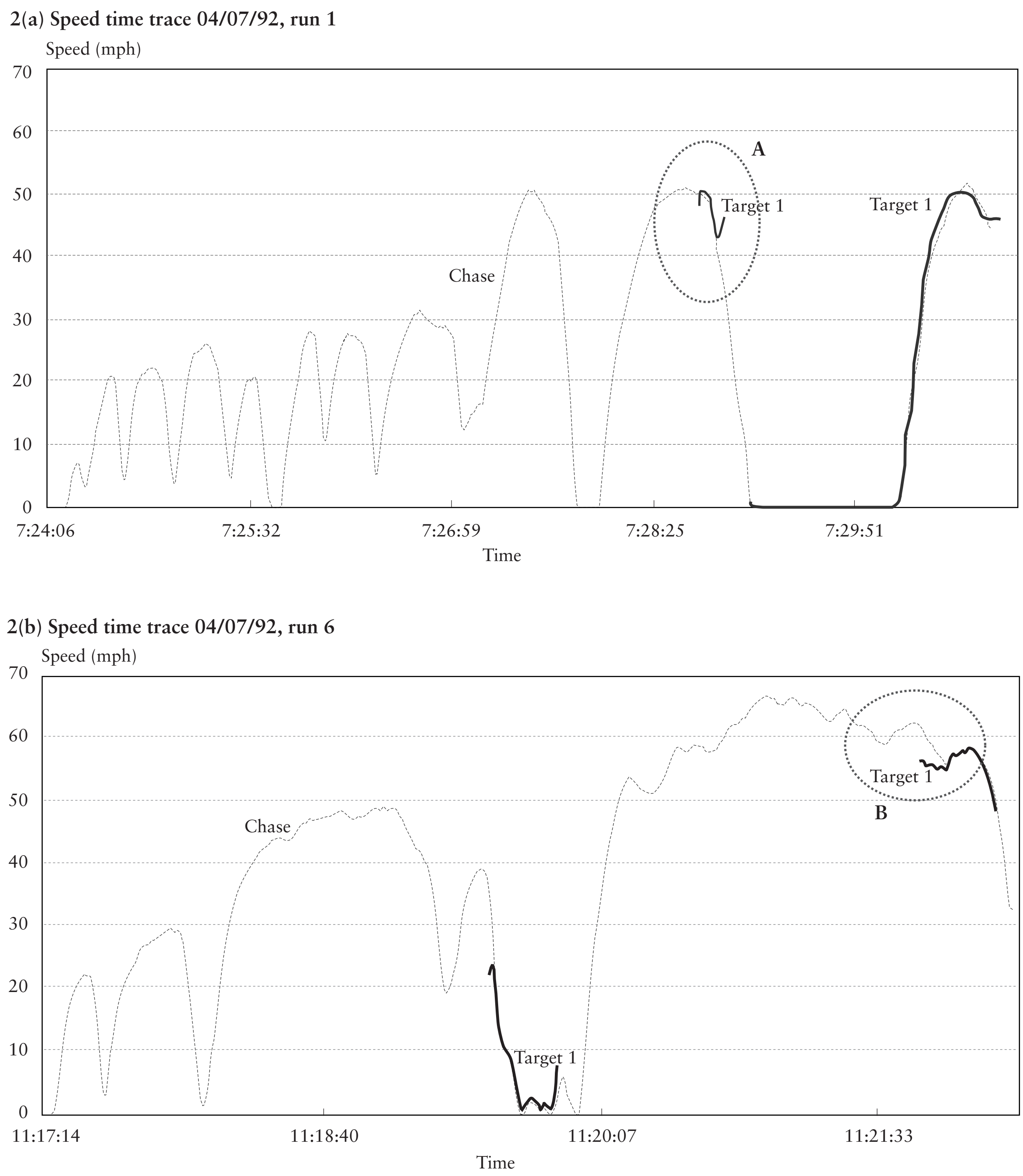
- As hypothesized, suspect points, such as that marked "A" in figure 2a, are due to loss of lock on the target.
- Other suspect points, such as that marked "B" in figure 2b, are not due to the chase car accelerating to acquire a target. Instead, they appear to be primarily due to turning events and lane changes by target vehicles. In some cases, the chase car turns off one facility onto another. In other cases, the target vehicle simply exits the route from which the chase car is recording data.

In addition to the LA92 videotapes, we reviewed videotapes pertaining to the 1997 Highway Performance Monitoring System (HPMS) data collection effort in the Sacramento, California region. The HPMS data were collected according to the LA92 chase car protocol and used the same chase vehicle as the LA92 project. The data exhibit similar anomalies, and a review of these additional videotapes corroborated our findings.

This analysis suggests that speed-time traces based on composite data are more representative of chase car operations than the target vehicle's. That is, chase car drivers do not seem to drive in a manner similar to the general public since they have different and specific motivations for their driving behavior. Consequently, the rationale for developing emissions cycles on the basis of these data may be invalid.

In order for non-lock chase car data to be representative of target vehicle data, the variation in speed under non-lock conditions should be equal to or less than the variation in target vehicle data. Although both lower and higher variation under non-lock conditions will weight the data disproportionately, smaller variation ensures a more conservative representation of driver behavior. A conservative perspective implies that chase car drivers tend toward the mean behavior of target vehicle drivers under similar conditions. The assumption underlying "conservative" is that the driving be-

FIGURE 2 Speed Time Traces of Chase and Target Vehicles from LA92



havior of the general population is represented by the mean speed of the target vehicles.

To assess if non-lock chase car data meet the conservative criterion, mean speeds and standard deviations of non-lock chase cars were compared to the mean speeds and standard deviations of target vehicles across all three cities. Since the percentages of lock-on vary widely by facility type and level of service, individual speed statistics were computed on the basis of these two factors. The HOV lane facility type has been removed from further consideration because there is so little data in this category. Table 4 contains mean speeds and standard deviations for both non-lock chase and

TABLE 4 Mean Speeds and Standard Deviations (mph) between Non-Lock Chase Car and Target Vehicle Data by Facility Type for All Three Cities

		Non-lock chase car (mph)		Target vehicle (mph)	
City	Facility type	Mean speed	St. dev.	Mean speed	St. dev.
Baltimore	All roadway types	28.2	19.7	35.7	22.0
	Private	11.4	10.9	14.3	12.0
	Local	18.2	13.0	17.1	15.5
	Arterial/collector	25.5	17.4	26.0	18.8
	Ramp	39.0	17.0	44.8	16.4
	Freeway	57.3	12.3	55.4	13.6
Los Angeles	All roadway types	26.6	18.5	30.0	21.4
	Private	5.6	4.7	—	—
	Local	16.4	11.1	16.4	11.8
	Arterial/collector	22.1	14.9	20.7	16.3
	Ramp	31.1	19.1	27.8	21.7
	Freeway	51.5	16.4	42.7	20.9
Spokane	All roadway types	27.7	18.2	32.7	20.8
	Private	5.3	6.8	1.3	2.6
	Local	21.1	17.1	37.4	15.5
	Arterial/collector	27.0	16.7	25.9	17.5
	Ramp	35.6	18.7	41.3	14.8
	Freeway	58.7	11.3	59.2	7.8

— Missing data.

target vehicles according to facility type and for all three cities.

In the aggregate data ("All roadway types"), B92 and S92 show larger differences in mean speeds between non-lock chase car and target vehicles than does LA92. In comparing mean speeds by facility type, we see that the most pronounced difference in Los Angeles occurs on freeways; in Baltimore, on ramps; and in Spokane, on local roads.

Table 5 shows the mean speeds and the results of a comparison of mean speeds by facility type using analysis of variance (ANOVA). The results of the ANOVA suggest that mean speeds of non-lock chase car and target vehicles, calculated on a second by second basis, are significantly different in all three cities and on nearly all facility types at a five percent significance level. The only facility type that did not show a significant difference in mean speeds was local roads in the LA92 database, but in this case the data were rather sparse.

An incremental difference in speed between non-lock chase cars and target vehicles is likely to be more important than identifying a large difference, particularly at higher speeds where coaxing the engine into an enrichment phase is likely to be accomplished with much smaller speed changes. To examine the speed variation from this perspective, the coefficient of variation (CV) of speeds was computed for target vehicles and for non-lock chase cars. Tables 6 and 7 contain these CV values for various facility types and levels of service represented in B92, S92, and LA92.

Table 6 contains the CV values for speeds computed on the basis of facility type. Observe that the target vehicle typically exhibits more variation in speed than does the non-lock chase car for LA92. As mentioned previously, to keep the estimates of driving behavior conservative, this should be the case. The coefficients of variation for LA92 suggest that the non-lock chase car data tend to reflect mean speeds, (i.e., the behavior of the "average driver,") rather than introducing additional variation into the data set. For B92 and S92, however, the results are less clear. For some facility types, such as local roads in Spokane and private roads in Baltimore, the variation in chase car speed far outweighs the variation in target vehicle speed.

TABLE 5 Mean Speeds and Comparison of Mean Speeds by Facility Type Using Analysis of Variance (ANOVA)

City	Facility type	Non-lock chase car mean speed (mph)	Target vehicle mean speed (mph)	Paired comparison of non-lock and target vehicle mean speeds (ANOVA) F-statistic	Significance
Baltimore	All roadway types	28.2	35.7	5,892.4	.000
	Private	11.4	14.3	6.1	.014
	Local	18.2	17.1	5.0	.026
	Arterial/collector	25.5	26.0	22.8	.000
	Ramp	39.0	44.8	157.1	.000
	Freeway	57.3	55.4	183.8	.000
Los Angeles	All roadway types	26.6	30.0	702.9	.000
	Private	5.6	—	—	—
	Local	16.4	16.4	.000	.998
	Arterial/collector	22.1	20.7	132.1	.000
	Ramp	31.1	27.8	32.9	.000
	Freeway	51.5	42.7	965.1	.000
Spokane	All roadway types	27.7	32.7	2,915.7	.000
	Private	5.3	1.3	26.4	.000
	Local	21.1	37.4	580.6	.000
	Arterial/collector	27.0	25.9	134.5	.000
	Ramp	35.6	41.3	62.957	.000
	Freeway	58.7	59.2	11.965	.001

— Missing data.

TABLE 6 Coefficients of Variation for Non-Lock Chase Car and Target Car Speeds in All Three Cities by Facility Type

	Non-lock chase CV	Target CV
Baltimore		
All roadway types	0.70	0.62
Private	0.95	0.84
Local	0.71	0.91
Arterial/collector	0.68	0.72
Ramp	0.44	0.37
Freeway	0.21	0.24
Los Angeles		
All roadway types	0.70	0.71
Private	0.80	—
Local	0.68	0.71
Arterial/collector	0.68	0.79
Ramp	0.62	0.78
Freeway	0.32	0.49
Spokane		
All roadway types	0.66	0.64
Private	1.30	2.05
Local	0.81	0.41
Arterial/collector	0.62	0.67
Ramp	0.52	0.36
Freeway	0.19	0.13

— Missing data.

In table 7, the CV values are disaggregated for the three predominant facility types (arterial/collector, freeway, and ramp) on the basis of level of service. As table 7 shows, speed is slightly more variable for non-lock chase cars than for target vehicles on freeways in Los Angeles under levels of service B, C, and D and on ramps under level of service C. However, table 7 indicates that in Baltimore there is significantly more variability in mean speeds when jointly considering level of service and facility type. The results in table 7 can be summarized by saying that chase car speeds exhibit greater variability under certain levels of service than do target vehicle speeds.

Accelerations and Decelerations

It is commonly known that acceleration events strongly affect emissions. However, Cernuschi et al. (1995) demonstrated that differences in relative deceleration rates have an effect on emissions as well. Consequently, in the present study acceleration and deceleration rates were compared for non-lock chase car and target vehicles represented in the LA92 database.

TABLE 7 **Coefficients of Variation of Non-Lock Chase Car and Target Vehicle Speeds by Facility Type and Level of Service**

	LOS A		LOS B		LOS C		LOS D		LOS E		LOS F	
	NL	TGT	NL	TGT	NL	TGT	NL	TGT	NL	TGT	NL	TGT
Baltimore												
Arterial	**0.57a**	0.54	**0.75**	0.68	0.78	0.81	**0.85**	0.84	1.08	1.28	**1.66**	0.74
Ramp	**0.38**	0.27	0.34	0.48	**0.43**	0.25	**0.51**	0.16	**0.34**	0.23	2.02	—
Freeway	**0.13**	0.11	0.14	0.14	**0.18**	0.11	**0.27**	0.14	**0.39**	0.36	**0.69**	0.68
Los Angeles												
Arterial	0.66	0.71	0.65	0.73	0.69	0.77	0.81	0.83	0.94	1.06	0.52	1.34
Ramp	0.66	0.70	0.55	0.59	**0.74**	0.60	0.60	1.24	0.34	0.51	0.60	0.87
Freeway	0.14	0.16	**0.11**	0.09	**0.14**	0.13	**0.19**	0.14	0.39	0.46	0.56	0.60
Spokane												
Arterial	**0.58**	0.57	**0.65**	0.64	0.76	0.76	0.77	0.89	**1.10**	0.69	**1.61**	1.22
Ramp	**0.51**	0.44	**0.46**	0.23	**0.66**	0.43	—	—	—	—	—	—
Freeway	**0.22**	0.13	**0.20**	0.08	0.12	0.13	**0.22**	0.12	0.12	0.15	—	—

NL = Non-lock chase car
TGT = Target vehicle
[a] **Bold numbers** indicate non-lock chase car CVs that are higher than the CVs of corresponding target vehicles.
— Missing data.

Three categories were created: cruise ($-0.0340 \leq a \leq 0.0340$ mph/s), normal acceleration ($0.0341 \leq a \leq 3.290$ mph/s), and hard acceleration ($a \geq 3.30$ mph/s), where a is the second-by-second acceleration of the vehicle. The cruise interval is based on work by Holmén and Niemeier (1998), and the hard acceleration interval is based on previous evaluations of chase car data (Austin et al. 1993). Deceleration rates were classified as mirror images of their acceleration counterparts: cruise ($-0.0340 \leq a \leq 0.0340$ mph/s), normal deceleration ($-0.0341 \leq a \leq -3.290$ mph/s), and hard deceleration ($a \leq -3.30$ mph/s). Table 8 indicates the percentage of time that chase cars and target vehicles spend in the various acceleration and deceleration intervals.

TABLE 8 **Percentages of Time Spent in Accelerations and Decelerations by Chase and Target Vehicles (LA92)**

	Chase (non-lock)	Target
Acceleration		
Cruise	27.9	21.4
Normal	66.3	75.2
Hard	5.9	3.4
Total	**100.0**	**100.0**
Deceleration		
Cruise	31.0	24.5
Normal	55.7	70.5
Hard	13.3	5.0
Total	**100.0**	**100.0**

Although the percentage of time that non-lock chase cars and target vehicles spend in the hard acceleration and hard deceleration intervals is small relative to the percentage of time spent in normal and cruise categories, the analysis indicates the differences in time spent in hard accelerations and hard decelerations are substantial. Non-lock chase cars recorded almost twice as many hard accelerations as did target vehicles and over 2.5 times as many normal acceleration events as target vehicles.

A relatively large number of the hard accelerations/decelerations may be explained by the chase car drivers' need to speed up or slow down to quickly acquire a target. However, since the non-lock chase car data is used to replace missing target vehicle data, the imputed values will be very influential and can contribute to an overprediction of modal frequency. If the overprediction is considerable, driving cycles will likewise tend to have too many of these modal events.

CONCLUSIONS AND RECOMMENDATIONS

Currently, composite chase car data provide the foundation from which driving cycles are developed, both to ensure vehicle compliance with air quality regulations and to characterize average emissions in the overall fleet. The robustness of the data can be judged by examining variation when the data are disaggregated by region, level of service, and facility type. The sources of this variation described and analyzed in this paper include difficult chase car protocol instructions, differences in the amount of data collected in each city, and differences between the driving behavior of target vehicle drivers and non-lock chase car drivers.

Two conclusions can be drawn.

- Instrument failures and misapplication of data collection procedures result in anomalies that significantly impact overall data representativeness.
- The composite data currently used for driving cycle development, which combine both non-lock chase car and target vehicle measurements, do not contain enough target vehicle information to adequately reflect the driving behavior of the general population.

These conclusions suggest that changes in the chase car protocol and in the technology used to measure target vehicle speed and acceleration could reduce biases in emissions cycles developed using the composite data. To address these problems in future chase car data collection efforts, we briefly elaborate on some recommendations.

Instrument Failures and Misapplication of Data Collection Procedures

Anomalies in speed-time traces from LA92 and HPMS data appear to be caused by protocol and/or instrument failure. Such anomalies may significantly influence cycles developed using the data sets and, therefore, the construction of driving cycles. The most notable deficiencies with respect to the existing technology are 1) the inability of the range-finder laser to maintain a lock on target vehicles when going over bumps, around slight curves, or on changing road grades and 2) the potentially large errors in measuring target vehicle accelerations. These can lead to both a marked lack of target data on ramps and inclines and a misrepresentation of target vehicle modal frequency. Currently, the most feasible and effective changes in technology would involve improvements to the laser or possibly the development of an appropriate scanning radar or scanning lidar system.

Problems with Mixing Non-Lock Chase Car and Target Vehicle Data

When examined at a fine scale of resolution by facility type and level of service, substantial variation, attributable to driving behavior, is observed in the composite data of B92, LA92, and S92. Significant differences in mean speeds between target and non-lock chase vehicles show that target drivers and chase car drivers represent separate populations. Similarly, an examination of accelerations and decelerations reveals disproportionate variation between non-lock chase car and target vehicle drivers. Given the uses of chase car data, it is important to note that combining the data may mask important differences between drivers in the resultant driving cycle. Therefore, we recommend that non-lock chase car data be minimized in, if not eliminated from, the driving cycle development process.

Proposed Changes in the Current Chase Car Protocol

Three changes in the chase car protocol can be made to create more robust databases at every level of aggregation. A primary element is the acquisition of additional target vehicle data. To acquire these data, the protocol should contain 1) simpler chase car routes and target car data collection procedures, 2) a simplified target vehicle selection procedure, and 3) the use of a traffic density measure rather than a visual assignment of level of service.

As evidenced by the LA92 videotapes, chase car routes appear to be too complicated for the drivers to concentrate on collecting target vehicle data. Various improvements to the route design could be made. One method would be to divide routes, predetermined from the top origin-destination pairs in a region, into segments by facility type and to have chase cars collect repeat data on the same segment in order to characterize target drivers' behaviors on that facility. The overall objective would be to allow chase car drivers to choose greater numbers of target vehicles and potentially stay with them for longer periods of time, thus increasing the records of target vehicles available in the databases. Simplifying the route design would also remove the chase car drivers' disincentive to engage target vehicles by leaving the chase car on a single facility for a longer period of time. The implicit assumption in this approach is that driving behavior on like facilities is similar.

Although the current vehicle selection method randomizes target vehicles and captures lane variation on multilane facilities, a revised lane sampling program based on predetermined lane choices would be less complex and would result in more reliable target car data. Any change in the lane-sampling program should be implemented in such a way as to guarantee all sources of variation are adequately represented. Two specific sources are within-lane variation and between-lane variation. Within-lane variation encompasses differences among potential target vehicles driving in the same lane, and sampling more vehicles and a wider range of vehicle types can adequately represent it. Capturing between-lane variation requires more extensive pre-run planning, especially on freeways where the sampling bias appears to be most extensive.

Finally, the collection of visually assigned level of service measurements appears to be of questionable use. As reported here, determination of level of service during chase car runs is very inconsistent. However, our analysis of data on the basis of level of service suggests it significantly affects mean speeds on certain facility types across all three cities. Poor level of service determinations may be the result of the data recording procedure rather than a reflection of actual driving behavior. Given its subjective basis, visually assigned level of service is not a reliable parameter for use in construction of driving cycles. Density measures, perhaps compiled from local travel management centers, would be more appropriate for use in regional driving cycle construction.

ACKNOWLEDGMENT

The authors gratefully acknowledge the insightful comments and suggestions of the reviewers and the issue editor. This study was funded under the CAMP Project by the California Department of Transportation. The views reflect those of the authors alone.

REFERENCES

Austin, T., F.J. DiGenova, T.R. Carlson, R.W. Joy, K.A. Gianolini, and J.M. Lee. 1993. *Characterization of Driving Patterns and Emissions from Light-Duty Vehicles in California, The Final Report.* Sacramento, California: Air Research Board, A932-185.

Cernuschi, S., M. Giugliano, A. Cemin, and I. Giovannini. 1995. Modal Analysis of Vehicle Emission Factors. *The Science of the Total Environment* 169, nos. 1–3: 175–83.

Gammariello, R.T. and J.R. Long. Development of Unified Correction Cycles. Paper presented at the sixth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, March 1996.

Highway Capacity Manual. 1994. Washington, DC: Transportation Research Board, National Research Council.

Holmén, B.A. and D.A. Niemeier. 1998. Characterizing the Effects of Driver Variability on Real-World Vehicle Emissions. *Transportation Research Part D* 3, no. 2: 117–28.

Niemeier, D.A., J.M. Utts, K. Nanzetta, and L. Ring 1998. *Speed Correction Curves: Scoping Study.* Institute of Transportation Studies, University of California, Davis, UCD-ITS-RR-98-12.

Sierra Research Inc. 1997. Average Speed Measurement in the SACOG Region: Technical Appendix. Report No. SR 97-09-01, California Department of Transportation.

Texas Transportation Institute. 1998. *Urban Roadway Congestion, Annual Report.* Available at http://mobility.tamu.edu/study/rci.stm.

Statistical Analysis and Model Validation of Automobile Emissions

DANIEL SCHULZ
THEODORE YOUNGLOVE
MATTHEW BARTH
University of California, Riverside

ABSTRACT

A comprehensive modal emissions model has been developed and is currently being integrated with a variety of transportation models as part of National Cooperative Highway Research Program Project 25–11. Second-by-second engine-out and tailpipe emissions data were collected on 340 light-duty vehicles, tested under "as is" conditions. Variability in emissions of CO_2, CO, HC, and NO_x were observed both between and within groups over various driving modes.

This paper summarizes initial statistical analysis and model validation using bootstrap validation methods. The bootstrap methodology was shown to be a valuable tool during model development. A significant positive bias (overprediction) in NO_x during higher speed driving was identified in CMEM v1.0 and eliminated in CMEM v1.2.

INTRODUCTION

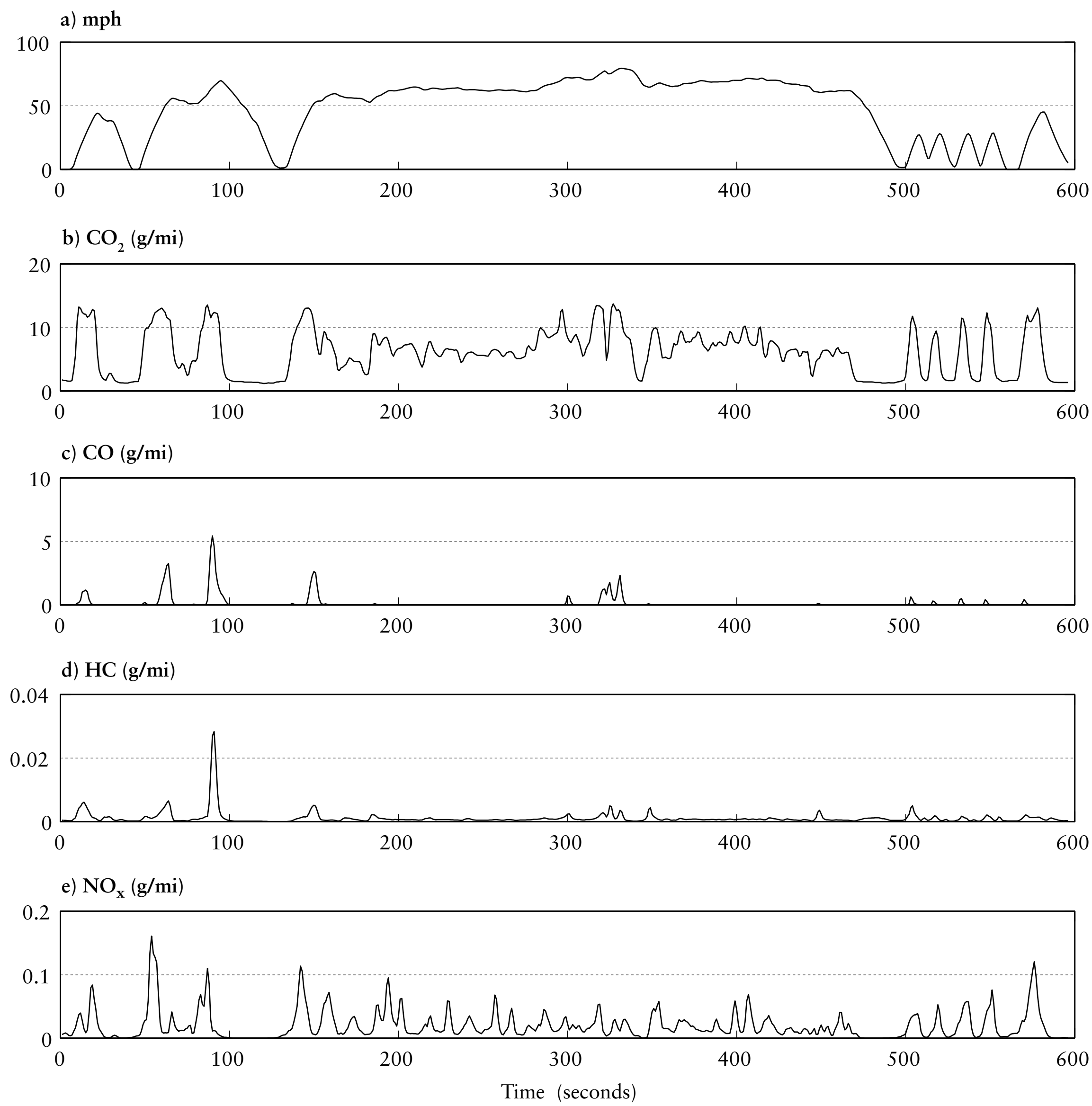
Measurements of automobile tailpipe emissions at second-by-second time resolution provide a statistically challenging data set for modeling and analysis. Emissions can vary by an order of magnitude within the space of a few seconds, with the response frequently nonlinear, due to enrichment

Theodore Younglove, Bourns College of Engineering, Center for Environmental Research and Technology, University of California, Riverside, CA 92521. Email: tyoung@cert.ucr.edu.

or enleanment of the air-fuel mixture. Figure 1 presents an emission trace from a representative, normally operating vehicle to illustrate the large differences in magnitude of tailpipe emissions over the driving schedule.

Enrichment occurs in modern computer-controlled vehicles based on proprietary engine control strategies. The computer enriches the air-fuel mixture at high power to protect the catalytic converter from heat damage, resulting in short-term spikes in emissions. The size and timing of the increases in emissions vary from vehicle to vehicle, even for identical models. Enleanment occurs in some modern computer-controlled vehicles during coastdown and braking events. In normal powered driving, the amount of condensed fuel on the walls of the intake manifold is in rough equilibrium with the addition of fresh condensate from fuel injection and with the loss by evaporation into the air moving into the cylinders. The amount of fuel on the walls depends to some extent on the recent history of fuel injection, that is, the recent power level. When engine power is negative, there is still significant air-flow but little or no fuel injection. The condensed fuel will be removed by evaporation over a period of seconds and will pass through

FIGURE 1 Second-by-Second Data for a 1986 Buick for a) Speed; b) CO_2; c) CO; d) HC; and e) NO_x



the cylinders. The critical fact is that during these events the fuel-air ratio is typically very lean, so lean that there is little or no combustion. In this case, hydrocarbon (HC) emissions can become quite high relative to normal operation. Second-by-second changes in emissions can occur during constant speed cruising, due in part to small changes in throttle position that, in turn, affect manifold air pressure without affecting vehicle speed.

In addition to these large differences in emissions for individual vehicles during driving, there are large differences in emissions from vehicle to vehicle. Changes in emissions behavior under different driving conditions occur because of changes in vehicle-emissions control technology. Large reductions in the emission of carbon dioxide (CO_2), carbon monoxide (CO), hydrocarbon (HC), and nitrogen oxides (NO_x) have been achieved over the past 25 years, resulting in great differences in emission rates between vehicle/technology groups (Calvert et al. 1993).

In late 1995, the Bourns College of Engineering, Center for Environmental Research and Technology (CE-CERT) at the University of California, Riverside undertook a cooperative investigation with the University of Michigan and Lawrence Berkeley National Laboratory in order to develop a comprehensive modal emissions model (CMEM). The overall objective of this research project was to develop and verify a modal emissions model that accurately reflects emissions from light-duty vehicle (LDV), cars and small trucks, produced as a function of the vehicle's operating mode. The model is comprehensive in the sense that it will be able to predict emissions for a wide variety of LDVs in various conditions (e.g., properly functioning, deteriorated, malfunctioning). The model is capable of predicting second-by-second tailpipe and engine-out emissions and fuel consumption for a wide range of vehicle/technology categories. The principal sponsor of this project is the National Cooperative Highway Research Program, NCHRP, Project 25–11 (see An et al. 1997). CMEM is a physical, parameter-based model requiring parameterization of many processes involving the vehicle, engine, emissions control system, and catalytic converter, and affecting how the vehicle is driven. Many of the relationships must be approximated within the model, and the parameters themselves are estimated from measurement data subject to error. This model differs from other conventional emissions models in that it is modal in nature: it predicts emissions for a wide variety of light-duty vehicles over a wide variety of driving modes, such as acceleration, deceleration, and steady-state cruise. The two primary models currently in use are MOBILE, developed by the U.S. Environmental Protection Agency, and EMFAC, developed by the California Air Resources Board. Both MOBILE and EMFAC predict vehicle emissions based in part on average trip speeds and depend on regression coefficients derived from a large number of trip average emission measurements for a driving schedule representative of "typical" driving. For more detail, see Barth et al. (1996), Barth et al. (1997), and An et al. (1997). Only emissions from light-duty vehicles are considered in this paper.

For model validation, the key question to answer is whether the model predicts emissions with reasonable accuracy and precision. Bornstein and Anderson (1979) have pointed out the need for communication between modelers and statisticians in air pollution research. Since then, Hanna has done considerable research into the development of statistical methods for air quality investigations (Hanna and Heinold 1985; Hanna 1988 and 1989). Of particular interest is his use of the normalized mean square error (NMSE) methods for estimating bias based on a percentile of observed and predicted differences, as well as his application of Efron's bootstrap resampling methods to compare different air pollution models (Efron 1982; Efron and Tibshirani 1986). Bootstrap bias plots, shape statistic plots, histograms of bias values, bootstrap confidence interval length plots, and maximum and minimum bias plots have also recently been used in the context of validating a complex modal emission model (Schulz et al. 1999).

In developing CMEM, several validation techniques were used: 1) validation of intermediate variables, such as modeled engine RPM against observed RPM, 2) composite vehicle schedule validation, and 3) second-by-second individual vehicle validation. Validation was undertaken on a second-by-second basis for individual vehicles to provide a robust data set on which to test the model and to ensure that a sufficient number of vehicles would be available for the bootstrap analysis. It should be noted that although this validation is accomplished at a second-by-second basis, the model was intended for use on driving modes lasting ten or more seconds. This difference was necessary for model development because of the need to identify situations in which problems were occurring. Practically speaking, many of the errors will "average out" over a driving schedule.

The focus of this paper is the validation methods employed on a second-by-second basis for use by the modeling team in model diagnostics and model improvements. The statistics used for model evaluation on a second-by-second basis must be valid under many possible distributions of emissions but must also be easily understood by nonstatisticians. In addition, while the initial validation presented in this paper was conducted on two large groups of vehicles, the methodology employed also needed to be valid for analysis of model performance on the individual vehicle/technology groups with 10 to 25 vehicles in each group. For these reasons, second-by-second validation methods inspired by Hanna's work are described and applied to two versions of the model.

METHODOLOGY

Vehicle Recruitment and Testing

The gasoline powered light-duty fleet was divided into 24 categories for vehicle recruitment, with divisions based on vehicle type (car or truck), emissions status (normal or high emitter), fuel control technology, emission control technology, power-to-weight ratio, and accumulated mileage. High-emitting vehicles were defined as those having CO, HC, or NO_x emissions 1.5 or more times higher than the certification standard for the vehicle. Vehicles ranged in age from a 1965 Ford Mustang to a 1997 Dodge Ram pickup and represented all major foreign and domestic auto manufacturers. The vehicle/technology groups were chosen to cover the range of vehicle technology types within the gasoline powered light-duty vehicle fleet. A total of 340 in-use vehicles were recruited and tested, primarily from the South Coast Air Basin, with a small subset brought in from other states. Particular care was given to target forty-nine state-certified vehicles, as well as California-certified vehicles, to ensure the model was representative of the national LDV population. Vehicles were selected at random from the Department of Motor Vehicles registration list for Southern California. Recruitment was conducted through a mailing to vehicle owners within the 24 categories, but category sample sizes were selected by model development needs rather than population proportions. Once recruited, the vehicles were tested on CE-CERT's forty-eight-inch electric chassis dynamometer using three driving schedules: the Federal Test Procedure (FTP), which the federal government uses to represent normal in-use driving; the US06 driving schedule, which the federal government uses to represent in-use hard driving; and the Modal Emission Cycle (MEC), developed as part of NCHRP Project 25–11 to measure emissions during specific driving modes (Barth et al. 1996). It should be noted that the third driving segment of the FTP driving schedule and the US06 driving schedule were not used in model development. For this reason, they were used as independent validation schedules. During testing, emissions of CO_2, CO, HC, and NO_x were measured on a second-by-second basis.

Time-Alignment of Data

To perform a meaningful second-by-second validation, the emissions test results first had to be time-aligned. The time delay between the start of data recording and the start of the vehicle is not automated and can vary by several seconds from one vehicle to the next. Prior to the application of bootstrap analysis to the vehicle data, all values were time-aligned to reflect acceleration from a common starting point. Small differences between the driving trace and the schedule speed trace are inevitable during the test schedules, so the time alignment is not perfect. Although some error can arise when time-aligning the files to the nearest second, it should be negligible when compared with the deviations from the driving trace resulting from human error.

Validation Statistics

A measure of closeness, called model bias but not the same as the statistical definition of bias (a property of an estimator of an unknown population parameter) is given by

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n} i^{\text{th}} \text{ modeled value} - i^{\text{th}} \text{ observed value} =$$

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}y_i = \overline{\hat{y}_i} - \overline{y_i}' \qquad (1)$$

where y_i is the i^{th} observed emission value, $\hat{y}_i$ is the corresponding i^{th} value predicted by the model, and there are n observations in the sample. This is consistent with the standard definition of bias historically used in environmental pollution studies (Zannetti 1990), but in the language of statistics it is referred to as mean prediction error. If this bias value is larger (smaller) than an acceptable predetermined cutoff value, then the model significantly overpredicts (underpredicts).

A point estimate of bias is useful, but statistics are random quantities that vary from sample to sample. Confidence intervals provide a better description of a reasonable range of values for the bias statistic. If the confidence interval (95% confidence intervals are used in this paper) contains the bias value of 0, then the model bias is not significantly different from 0, and the model is performing well. If the interval does not contain the bias value of zero, the model may have some prediction problems, thereby warranting further investigation.

In standard parametric statistical theory, confidence intervals are constructed assuming the statistic of interest follows a known distribution. The assumed distribution is frequently a normal distribution. These distributional assumptions are valid for simple statistics like the mean and variance. Here, for bias, a mean is calculated, but it is not the usual sample mean. Averaging involves emission values predicted from the model, which could have a strange, underlying distributional form. Therefore, it is undesirable to assume that bias follows a normal distribution since its true form is unknown. Also, there is no obvious calculation to estimate the standard error of the bias. For these reasons, the method of choice is the bootstrap method to determine confidence intervals (Efron and Tibshirani 1993).

The bootstrap algorithm can now be described in detail in this context. The bootstrap sampling is conducted at each time point in the driving schedule with new sequencing of the bootstrapped samples. First, assume a sample of n paired observations drawn from the population of interest. The first value in each pair is the observed emissions value, and the second value is the corresponding predicted emissions value. To construct the first bootstrap sample, a sample pair is chosen at random from the original sample. Its values are recorded, and the selected pair is returned to the original sample. A second pair is chosen at random from the original sample, its values recorded, and is then returned to the original sample. This is the second pair of values in the first bootstrap sample. Pairs of values are chosen from the original sample until the first bootstrap sample contains n pairs and thus is the same size as the original sample. In this fashion, a random group of vehicles the same size as the actual group is created. The first value of the bias statistic can be calculated from these paired values.

The second bootstrap sample is calculated in a similar way to the first with a new randomization of pairs chosen with replacement until there are n pairs in the bootstrap sample. The second value of the bias statistic is then computed. This procedure is repeated until B bootstrap samples, each of size n, have been drawn, and B bias estimates have been calculated. B must be quite large in order to obtain reasonably accurate results. For the present study, B = 1,000 is used. Of the 1,000 bias estimates calculated, the 25th smallest bias estimate, or 2.5 percentile, is determined, as well as the 25th largest bias estimate, or 97.5 percentile. The difference between these two numbers is an approximate 95% bootstrap confidence interval on the bias.

While there are other bootstrap methods for establishing confidence intervals (Efron and Tibshirani 1993), the percentile method is preferred for the present study due to its simplicity and because the intervals can be asymmetric, unlike traditional confidence intervals. Concerns about potential accuracy and underprediction are offset in this study by the number of vehicles considered, 340, as well as the number or replications of the procedure, 1,000. Consequently, for a given constituent emitted on a specific driving schedule, the 95% bootstrap confidence interval is calculated based on 1,000 replications for each second in time over the length of the driving schedule. The US06 driving schedule is about 589 seconds long, resulting in 589 intervals with different random sequences of vehicles. Formally speaking, these intervals are not to be used for strict statistical hypothesis testing. To do so could lead to overstated, erroneous conclusions. Informally speaking, the plots are quite useful for summarizing the available information in the data and for observing underlying patterns and trends through time. Plots of the length of the confidence intervals over time are used as a measure of variability of the bias statistic. Wider intervals indicate more variability. Narrower intervals indicate less variability.

In addition to the plots of bootstrap confidence intervals, called bias bootstrap plots, other potentially informative plots over time, such as plots of the shape statistic, can be constructed (Efron and Tibshirani 1993). The shape statistic is a measure of skewness, which numerically describes the shape of the distribution of the statistic of interest.

RESULTS

Due to the large differences in emissions and the possible differences in emissions behavior over driving modes, the normal-emitting (emissions less than 150% of the vehicle's certification standard) and high-emitting (emissions greater than or equal to 150% of the emissions standard) vehicles were analyzed separately. Second-by-second bias plots with bootstrap confidence limits were constructed for CO_2, CO, HC, and NO_x after calculation of model results. The US06 NO_x results are presented for CMEM v1.0 and CMEM v1.2. Differences in CMEM v1.0 and CMEM v1.2 are described below. The bias plots for CO_2, CO, and HC followed the same general pattern as those of NO_x but did not show large changes from CMEM v1.0 to CMEM v1.2 and are not presented here. NO_x results for CMEM v1.0 normal-emitting vehicles and high-emitting vehicles are shown in figures 2 and 3, respectively.

Figure 2 shows that the model overpredicts NO_x emissions to a small degree in normally operating vehicles during the high-speed cruise section of the US06 driving schedule. Figure 3 indicates that for the high-emitting vehicles there is no model over-

FIGURE 2 US06 Normal Emitter Second-by-Second Average Bias

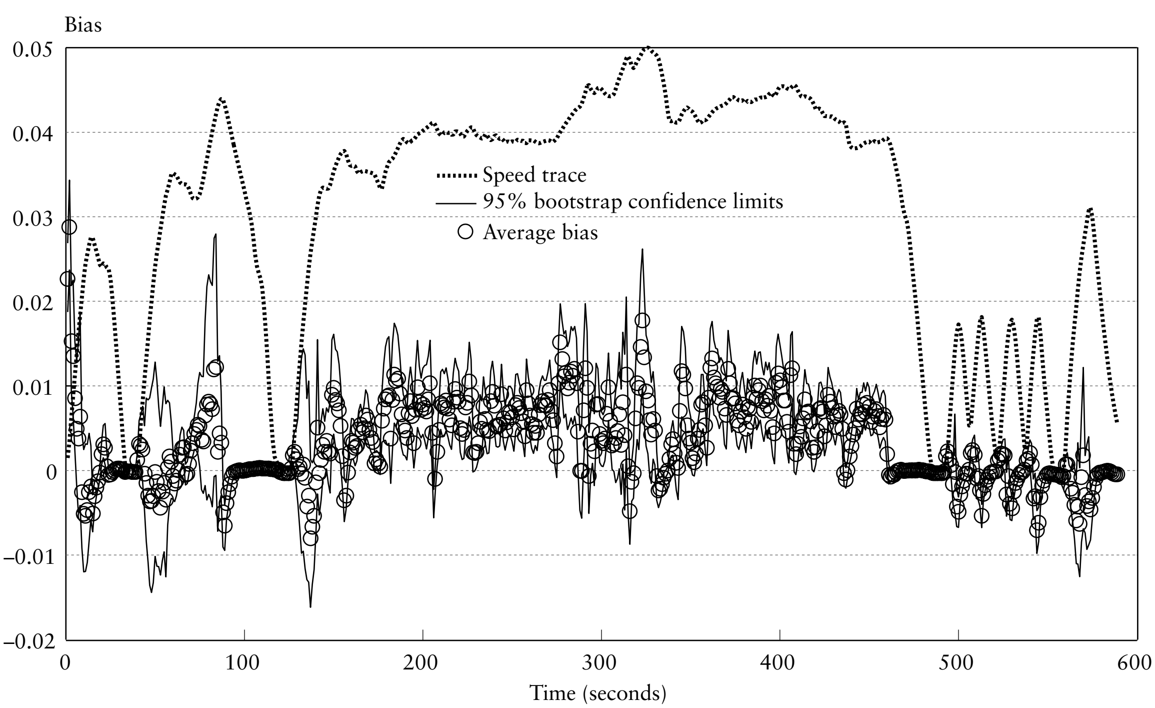
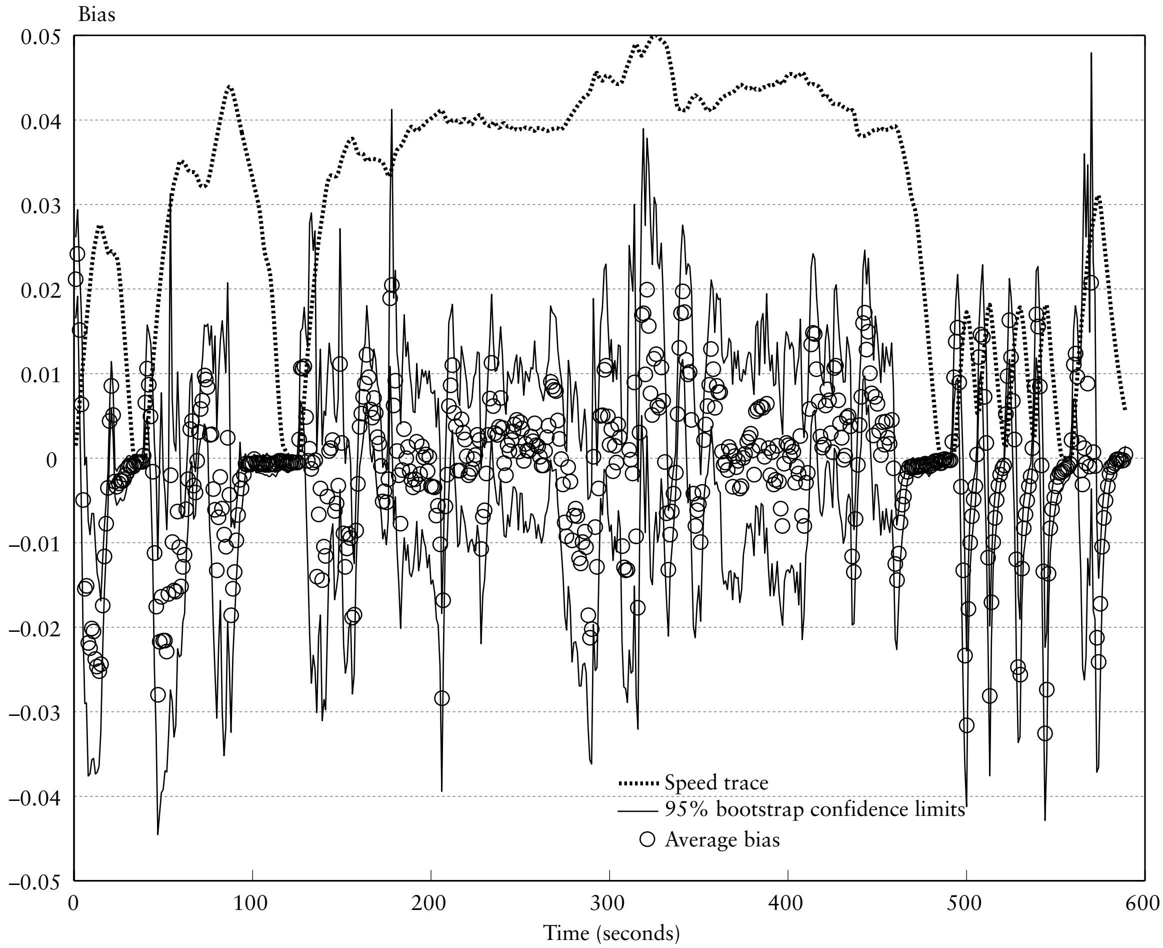


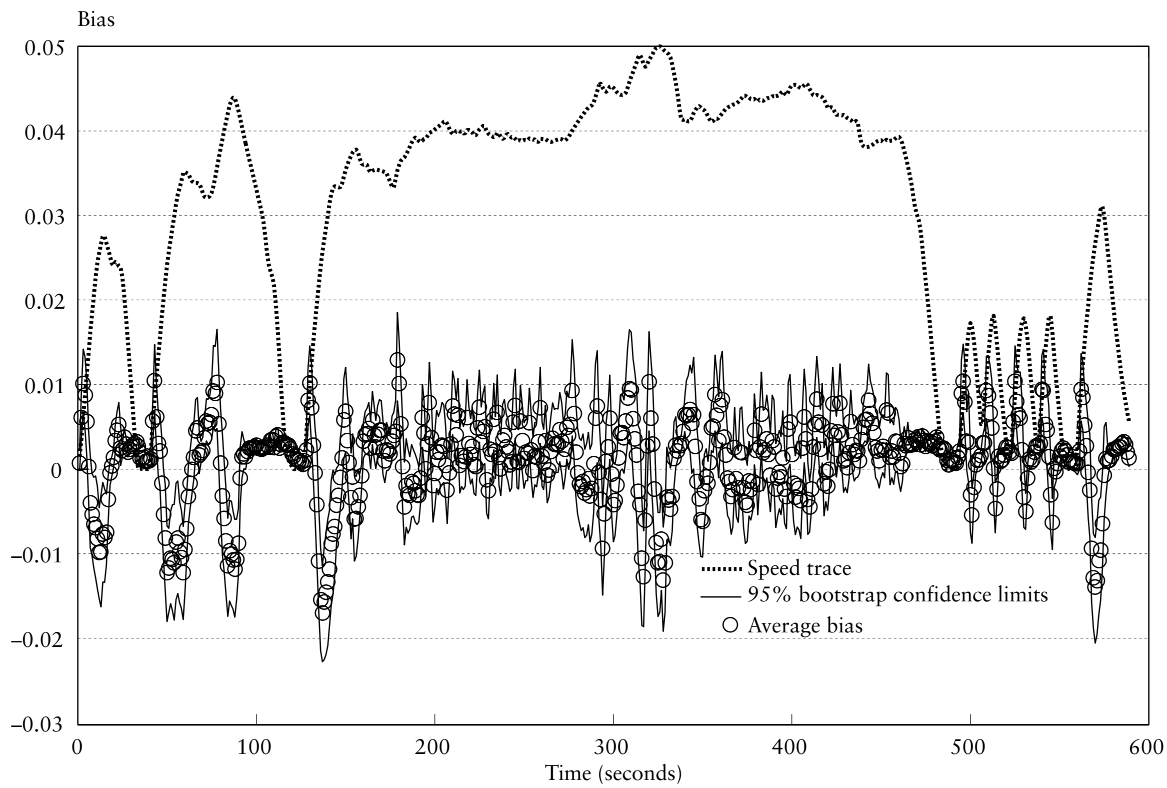
FIGURE 3 US06 High Emitter Second-by-Second Average Bias



prediction for the high-speed cruise section. Comparison of figure 2 and figure 3 also shows that bias is more variable for the high-emitting vehicles, apparent from the wider confidence limits and greater range of average bias values from second to second. This can be explained at least in part by the higher levels of emissions for the high-emitting vehicles and the higher variability in emissions of high-emitting vehicles. Additionally, both figures 2 and 3 suggest that the model overpredicts emissions at the start of an acceleration event and underpredicts them at the end of the acceleration event. Thus, the observed pattern in bias indicates that this version of the model may be inadequate for detailed second-by-second analysis while still appropriately capturing the intended range of emissions on the total driving trace and for driving modes. Driving modes are considered as individual events such as acceleration, deceleration, and steady-state cruising. For example, users of the model would be interested in the total emissions contribution of a vehicle accelerating onto the freeway and not in emissions at the start and end of the acceleration separately.

Due to the validation results discussed above, modifications were made to the NO_x components of the CMEM model, leading to the establishment of CMEM v1.2. NO_x emissions predictions for normal-emitting vehicles on the US06 using CMEM v1.2 are presented in figure 4. Similar results for the high-emitting vehicles are presented in figure 5. The bootstrap results show the resulting changes in the model bias. Note that the overprediction of NO_x in normal-operating vehicles at the high-speed portion of the US06 has been eliminated (figure 4). However, the deceleration events for which CMEM v1.0 exhibited no under- or overprediction now do exhibit overprediction of emissions, as seen in the positive values and narrow confidence bands around times 100 and 475.

FIGURE 4 US06 Normal Emitter Second-by-Second Average Bias



This indicates that CMEM v1.2 overpredicts NO_x on long deceleration modes for normal-operating vehicles. These changes, while not perfect, represent a substantial improvement in the model prediction accuracy for normal-operating vehicles because the high levels of NO_x in the high-speed portions are much more important than the low NO_x levels produced in the deceleration events.

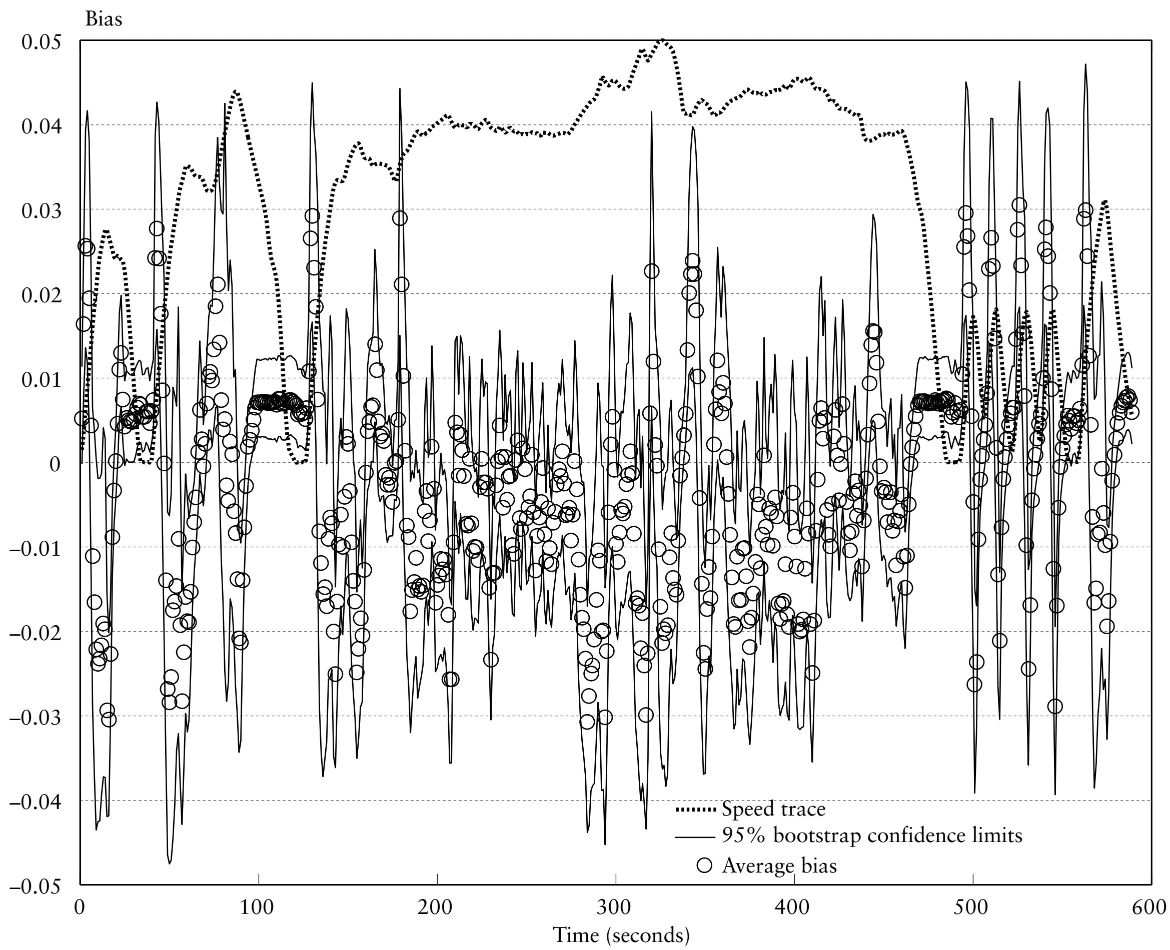
For the high-emitting vehicles, figure 5 suggests that the changes to the model have affected predictions of emissions at the high-speed portion of the schedule. The overprediction in the high-speed portions of the driving schedule is slightly lower for CMEM v1.2 than for CMEM v1.0 (figure 3 versus figure 5), with CMEM v1.2 tending towards underprediction of NO_x on the high-speed driving section. For CMEM v1.0, the confidence limits include zero indicating no under- or overprediction during the high-speed driving section, but for CMEM v1.2 some parts of the high-speed section do not include zero. The overprediction in NO_x for long deceleration events is also clearly visible on the high-emitting vehicles around times 100 and 475 (figure 5).

CONCLUSIONS

The bootstrap technique has been proven to be a useful method for graphically validating the predictions of CMEM on a second-by-second basis. This paper has also shown the bootstrap bias plot to be a useful tool for modelers during the model development process. It provides both detailed and summary information about the model's accuracy to facilitate model refinement. Using bootstrap bias plots, it can be determined if the model is predicting as well as desired, and if not, the bias plots identify where the bias is occurring in the driving schedule. Overall, the effects of model improvements can be observed directly in the plots, leading, in the particular case described, to the elimination of overprediction in NO_x under high-speed driving conditions for normal-operating vehicles. In the case presented here, the bias plots also identified unintended changes in model behavior resulting from the changes to the model.

The technique described here has been used on 340 vehicles split into 2 groups: normal emitters and high emitters. Differences in model bias were observed between the two groups. Further comparisons of these vehicles on the basis of the other

FIGURE 5 US06 High Emitter Second-by-Second Average Bias



classification criteria, such as carburetor versus fuel injection, could provide more valuable information for improving model bias. In addition, further research should be conducted to determine whether other statistics or other bootstrap methods of determining confidence intervals on emissions model predictions are more appropriate.

Finally, current efforts are focused on other ways to compare different versions of emissions models. Validation studies are targeting methods used to compare model results on the basis of an overall driving schedule in much the same way that the vehicles are expected to be used in practice, rather than on a second-by-second basis.

REFERENCES

An, F., M. Barth, M. Ross, and J. Norbeck. 1997. The Development of a Comprehensive Modal Emissions Model: Operating Under Hot-Stabilized Conditions. *Transportation Research Record* 1587:52–62.

Barth, M., F. An, J. Norbeck, and M. Ross. 1996. Modal Emissions Modeling: A Physical Approach. *Transportation Research Record* 1520:81–8.

Barth, M.J., T. Younglove, T. Wenzel, G. Scora, F. An, M. Ross, and J. Norbeck. 1997. Analysis of Modal Emissions from a Diverse In-Use Vehicle Fleet. *Transportation Research Record* 1587:73–84.

Bornstein, R.D. and S.F. Anderson. 1979. A Survey of Statistical Techniques Used in Validation Studies of Air Pollution Prediction Models. Technical Report 23, Stanford University. Stanford, CA.

Calvert, J.G., J.B. Heywood, R.F. Sawyer, and J.H. Seinfeld. 1993. Achieving Acceptable Air Quality: Some Reflections on Controlling Vehicle Emissions. *Science* 261:37–45.

Efron, B. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans.* Society of Industrial Applied Mathematics CBMS–National Science Foundation Monograph 38.

Efron, B. and R. Tibshirani. 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* 1, no. 1:54–77.

Efron, B. and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57. New York, NY: Chapman & Hall.

Hanna, S.R. and D.W. Heinold. 1985. *Development and Application of a Simple Method for Evaluating Air Quality Models*. Publication 4409. Washington, DC: American Petroleum Institute.

Hanna, S.R. 1988. Air Quality Model Evaluation and Uncertainty. *Journal of the Air Pollution Control* Association 38:406–12.

_______. 1989. Confidence Limits for Air Quality Model Evaluations, as Estimated by Bootstrap and Jackknife Resampling Methods. *Atmospheric Environment* 23, no. 6:1385–9.

Schulz, D., T. Younglove, M. Barth, C. Levine, and G. Scora. 1999. Development of a Model Validation Procedure for Use in Evaluating Changes in a Complex Modal Emissions Model, in *Proceedings of the Ninth CRC On-Road Vehicle Emissions Workshop*. Atlanta, GA: Coordinating Research Council, Inc.

Zannetti, P. 1990. *Air Pollution Modeling: Theories, Computational Methods and Available Software*. Boston, MA: Computational Mechanics Publications.

A Framework for Modeling In-Use Deterioration of Light-Duty Vehicle Emissions Using MOBILE6

PHIL ENNS

United States Environmental Protection Agency

ABSTRACT

The U.S. Environmental Protection Agency is currently revising its Mobile Source Emission Factor Model, used to estimate the inventory of exhaust and evaporative emissions from on-road motor vehicles. This paper describes the framework used in calculating basic exhaust emission rates as a function of accumulated vehicle mileage. In general, these rates increase with mileage. In version 6 of the model, MOBILE6, vehicle exhaust emissions are separated for the first time into "start" and "running" components. This enables more precise descriptions in the model for specific types of driving. Basic rates for start and running emissions are estimated from laboratory test data and from state inspection and maintenance program data. The data suffer from various limitations, and considerable engineering judgment must be used to augment traditional statistical methods to arrive at practical results.

INTRODUCTION

The U.S. Environmental Protection Agency's (EPA) Mobile Source Emission Factor Model (MOBILE) is used by various groups in government and industry to obtain estimates of emissions from on-road

Phil Enns, United States Environmental Protection Agency, National Vehicle and Fuel Emissions Laboratory, 2000 Traverwood Drive, Ann Arbor, MI 48105. Email: enns.phil@epa.gov

vehicles. State, regional, and local governments combine output from the model with estimates of pollution from other sources to help develop air quality management plans. Vehicle and fuel manufacturers are concerned with the impact of the model's predictions on government policies that affect activities in their industries. By the same token, as part of their mission members of the environmental community monitor local developments with MOBILE.

This paper reports on the development of basic exhaust emission rates expressed as a function of vehicle mileage accumulation. In general, these rates increase with mileage. At the request of many of the model's users, vehicle exhaust emissions will be separated for the first time into "start" and "running" components in version 6 of MOBILE (MOBILE6). This will enable more precise descriptions of specific types of driving in the model. Basic rates for start and running emissions are estimated from laboratory test data and from state inspection and maintenance (I/M) program data. In most cases, while the quantity of data is large, they have been collected for some other primary purpose and are not ideally suited to the problem of estimating emissions deterioration. Therefore, considerable engineering judgment has been used to augment traditional statistical methods in arriving at practical results.

The following account is intended as a broad overview of the steps taken to arrive at model equations. It represents a synthesis of work described in more detail in a series of reports prepared by EPA as part of the current MOBILE revision project. Readers interested in these details are referred to supporting documents cited here and available at the EPA web site: www.epa.gov/OMSWWW/M6. In particular, statistical measures of uncertainty are largely omitted from this paper.

Most of the work discussed here deals with 1981 to 1993 model year light-duty cars and trucks. At the time of this writing, substantial new data were available only for this portion of the vehicle fleet. Treatment in MOBILE6 of other model years and vehicle classes is described briefly in the results section, with more complete details found in additional EPA reports.

MODELING BASIC EMISSION RATE DETERIORATION

EPA's study of in-use deterioration of exhaust hydrocarbon (HC), carbon monoxide (CO), and oxides of nitrogen (NO_x) emissions began as a broad analysis of how rates of emission change as a vehicle ages and accumulates mileage. Much of the early work involved problem definition and identification of useful data sources.[1] Over time, the in-use deterioration study merged with the ongoing MOBILE model revision in order to supply the latter with required input. This shift gave the analysis greater focus, while perhaps limiting its generality.

In MOBILE6, vehicle exhaust emissions will be allocated between engine start (start emissions) and travel (running emissions). This split enables the separate characterization of start and running emissions for correction factors such as fuel effects and ambient temperature. It also allows a more precise weighting of these two aspects of exhaust emissions for particular driving situations, such as those associated with morning commutes, parking lots, and freeways.

Traditional emissions testing does not directly reflect the start/running emission division, and this creates difficulties in the development of models using actual data. The accepted unit of emission measurement is a vehicle's recorded emissions, in grams per mile, on the Federal Test Procedure (FTP), a laboratory test designed to reflect real-world driving.[2] An FTP score is computed from the values of three "bags" of emissions. Bags one and three capture a combination of start and running emissions, while bag two measures running emissions alone. A large body of FTP data has been collected since the inception of the FTP protocol and is available from various sources. In order to utilize these data in the study of start and running emissions deterioration, it is necessary to develop a method of segregating the start and running components associated with given FTP test results.

---

[1]For a more complete review, see Mobile Sources Technical Review Subcomittee (1997).

[2]In recognition that the FTP does not adequately represent more extreme levels of speed and acceleration, a "supplemental" FTP component will be included in future test programs. Data from this cycle were not available for the current study. See USEPA (1993) for complete details.

A second challenge involves addressing concerns over possible bias in FTP test results. Since the estimates of running emissions deterioration are based on FTP tests obtained from public vehicle recruitment programs, there is a concern that low vehicle recruitment rates in these programs may sustain sampling bias; typically less than 25% of drivers/owners asked to participate actually do so (Mobile Sources Technical Review Subcomittee 1997). Whether such a bias, if it exists, results in overestimation or underestimation of the true emissions deterioration is a matter of debate. Nonetheless, a method for overcoming this situation was included in the analysis described below. It utilizes data from state and inspection maintenance (I/M) programs based on the IM 240 test. This test is designed to produce emissions similar to the FTP over a shorter cycle more appropriate for the high volume of testing required in an I/M program (USEPA 1992).

DATA

Several data sets were needed ultimately to model deterioration of light-duty vehicle exhaust emissions.

FTP Data

The 1,876-second FTP has long served as the standard for exhaust emissions testing. Among other features, it contains elements of driving that produce start emissions as well as running emissions. In the MOBILE model revision, three FTP data sources were employed: 1) tests conducted or sponsored by EPA, most of which were performed at the EPA laboratory in Ann Arbor, Michigan; 2) data received from the American Automobile Manufacturers Association (AAMA) based on testing conducted in Michigan and Arizona; and 3) American Petroleum Institute (API) data collected in Arizona (USEPA 1999a). Vehicle model years range from 1981 through 1993, and both cars and trucks are included. Table 1 contains a cross tabulation of all the vehicles by type, model year, and technology for the three data sets combined.

Most of the data from 1990 and later model year vehicles were supplied by AAMA, while most of the pre-1990 data came from EPA laboratory testing. The API sample, 99 cars and trucks, is relatively small. Its chief appeal is that the vehicles' mileage readings, all over 100,000 miles, are generally higher than the rest of the sample. There has been a general transition from carbureted and open loop technologies in early model years to fuel injection in more recent years. Port fuel injected vehicles have represented the dominant technology since the 1990 model year. Although not directly apparent in table 1, new catalyst technology has been slowly phased into the U.S. fleet since the mid 1980s.

TABLE 1 Numbers of Vehicles by Model Year and Technology in the Combined FTP Data Set

	Cars					Trucks					
Model year	Open loop	Carbureted	TBI	PFI	Subtotal	Open loop	Carbureted	TBI	PFI	Subtotal	Total
1981	657	367	15	29	1,068	0	124	0	0	124	1,192
1982	71	71	74	8	224	0	45	0	0	45	269
1983	57	63	127	62	309	3	8	0	0	11	320
1984	30	5	46	35	116	22	26	0	1	49	165
1985	74	24	56	66	220	30	33	13	6	82	302
1986	34	7	60	92	193	9	14	23	41	87	280
1987	17	1	76	106	200	0	0	6	4	10	210
1988	15	0	69	113	197	0	0	0	0	0	197
1989	22	0	38	103	163	0	0	0	0	0	163
1990	0	0	160	250	410	0	0	144	1	145	555
1991	0	0	91	426	517	0	0	141	144	285	802
1992	0	0	57	347	404	0	0	92	92	184	588
1993	0	0	29	366	395	0	0	90	93	183	578
All years	**977**	**538**	**898**	**2,003**	**4,416**	**64**	**250**	**509**	**382**	**1,205**	**5,621**

PFI = Port fuel injected
TBI = Throttle body injected

IM 240 Data

Because FTP data can be potentially tainted with vehicle sampling (recruitment) bias, a means must be devised to account for it. One possible approach is to supplement FTP data, or adjust it, with IM 240 data. The IM 240 test cycle was developed to provide a relatively short (240-second) test that captures the essential features of the FTP. While the IM 240 test cycle is considered less representative of real world driving than the FTP, it has one clear advantage: because it is required for all vehicles in every U.S. noncompliance region, the data sets are very large and essentially free of recruitment bias. Consequently, the results of IM 240 tests provide candidate data with which to supplement FTP results (USEPA 1999b).

HR505 Data

Unfortunately, the IM 240 test only collects running emissions. Therefore, IM 240 data can only be used to adjust the running portion of FTP emissions supplied to MOBILE6. This can be accomplished using results from FTP tests that involve collecting an extra bag of emissions known as the Hot Running 505 (HR505) component (USEPA 1999c). The HR505 is an extra exhaust emissions test cycle performed immediately following collection of the conventional third bag in the standard FTP. This additional bag is a duplicate in terms of speed and time of the first and third bags. The only difference between the bags is that the HR505 does not include an engine start. For the MOBILE model revision project, a special set of 77 FTP tests were used for which the HR505 was available.

Dayton IM 240 Fast-Pass Data

Data from the Ohio I/M program include IM 240 test results on all 1981 and older registered cars and light-duty trucks scheduled to be tested from April 1996 through March 1997. Since the testing frequency in the Ohio program is biennial, this collection, which contains more than one million vehicles from three separate Ohio cities (Cleveland, Akron/Canton, and Dayton/Springfield), represents approximately half the overall population. However, only the data from Dayton/Springfield, the "Dayton data," were used in the MOBILE model revision project, and these data were further restricted to the valid initial tests; no post-repair retests were used. Only the Dayton data were used because that city never implemented any I/M or anti-tampering program (ATP). Consequently, there was reason to believe that deterioration of measured emissions would be more "natural" than in other parts of the state. The resulting data set contains IM 240 test scores for more than 180,000 cars and light-duty trucks.

An important feature of the Ohio I/M program is that it employs a "fast-pass" algorithm to speed up the testing process. Under this protocol, a vehicle's emissions are monitored in real time, and the test is terminated before completion if the accumulated emissions are sufficiently low. As a result, in order for the Dayton data set to be useful, measurements for a full 240-second cycle must eventually be constructed.

Wisconsin Full 240-Second I/M Data

A set of full 240-second I/M data were collected in Wisconsin during December 1995, April 1996, and October 1996. This data set contains observations on 3,148 cars and 1,192 light-duty trucks, with model years ranging from 1981 through 1995. Data from Wisconsin are preferable to similar data from Arizona and Colorado, the other two IM 240 states reporting second-by-second data, due to the geographic, demographic, and meteorological similarities between Ohio and Wisconsin. Recall that the Dayton IM 240 data is of interest because of its more natural state. Furthermore, both states use the same testing contractor, so analyzers and specific test procedures are likely to be similar.

FTP and IM 240 Correlation Data

Additional data are ultimately required to determine the correlation between IM 240 and FTP emissions measurements. The data available for this purpose consist of 938 FTP and IM 240 paired tests conducted on vehicles chosen from I/M lanes in Hammond, Indiana and Phoenix, Arizona. Vehicles were randomly selected at the inspection lanes to be included in this program, and IM 240 tests were conducted using the fuel resident in their tanks. The vehicles were then moved to the lab, and each vehicle's fuel was replaced with Indolene, in accordance with standard test protocol. An FTP test was then conducted, as was another in-lab IM 240. Only the IM 240 tests on tank fuel are of interest since the IM 240 data from Ohio derive from tank fuel tests. Again note that it is the Dayton IM 240 that is of specific interest to the MOBILE model revision project.

FTP ADJUSTMENT METHODOLOGY

Historically, the MOBILE model has portrayed vehicle exhaust emissions as a piecewise linear function of accumulated mileage. Different functions are used for the three pollutants, hydrocarbons (HC), carbon monoxide (CO), and oxides of nitrogen (NO_x), and for different vehicle categories defined by model year and engine technology. This approach has been retained in MOBILE6 for the modeling of running emissions. A somewhat different concept underlies the modeling of start emissions, where vehicles are assumed to fall into discrete categories of "normal" and "high" emitters. The adjustment for possible bias in FTP data employs a series of regression analyses that enable the use of IM 240 tests to model the deterioration of FTP values. Figure 1 contains a flowchart depicting how the various data sources are combined and how the FTP results are eventually adjusted.

Model Year and Vehicle Categories

Individual deterioration functions were developed for subsets of vehicles classified by vehicle type, model year, and fuel metering technology. The categories in table 2 were determined largely on the basis of engineering judgement. To a great extent, model year serves as a proxy for technology advances. The choice of model year groupings also approximately reflects the continued tightening of regulatory emission standards.

For a particular model year/technology class, MOBILE5 included an upward turning "kink" in the basic emission rate (BER) function at 50,000 miles; that is, emissions are expected to deteriorate at a faster rate beyond 50,000 accumulated miles (USEPA 1994). As already described, a great deal of additional data were available from tests on newer model vehicles for use in the development of MOBILE6. This permitted a closer examination of the appropriateness of the 50K kink and led to its eventual elimination as an explanatory factor.

Separating Start and Running Emissions

Emissions on the FTP are computed by weighting the gram per mile measurements from the three bags of the test cycle. The formula for this calculation is:

$$\text{FTP} = 0.21 \times (\text{bag } 1) + 0.52 \times (\text{bag } 2) + 0.27 \times (\text{bag } 3) \qquad (1)$$

The weights equal the fractions of vehicle-miles traveled in the three modes of driving captured by the cycle. Bags one and two constitute the "LA4" cycle, which refers to the underlying driving data collected in Los Angeles. This leads to formulas for the calculation of running and start emissions using the bag measurements from an FTP.

For a given FTP test, running emissions are determined by a linear function of bag two emissions and HR505 emissions. The result is labeled "Running LA4" since it captures emissions for the LA4 cycle with start emissions removed. The general form of this function is given by:

$$\text{Running LA4 emissions (grams/mile)} = 0.48 \times (\text{FTP bag } 2) + 0.52 \times \text{HR505} \qquad (2)$$

The HR505 values are themselves estimated from emissions measurements associated with bags one to three using a regression model developed from the special 77-car HR505 data set. In this way, it is possible to compute running emissions for each of the FTP tests in the EPA/Industry database.

Likewise, start emissions for each FTP test are determined by a linear function of HR505 and the start emissions components of bags one and three. The general form of this expression is:

$$\text{Start emissions (grams)} = 0.21 \times (\text{FTP bag } 1) + 0.27 \times (\text{FTP bag } 3) - 0.48 \times \text{HR505} \qquad (3)$$

HR505 values are the same ones used in determining running emissions.

Modeling Running Emissions

For running emissions, several functional forms were studied before selecting a piecewise linear model in which emissions are constant at lower mileages and increase after about 20,000 miles. The 20,000-mile point was selected subjectively following graphical

FIGURE 1 Overview of Methodology to Estimate Running and Start Basic Emission Rates

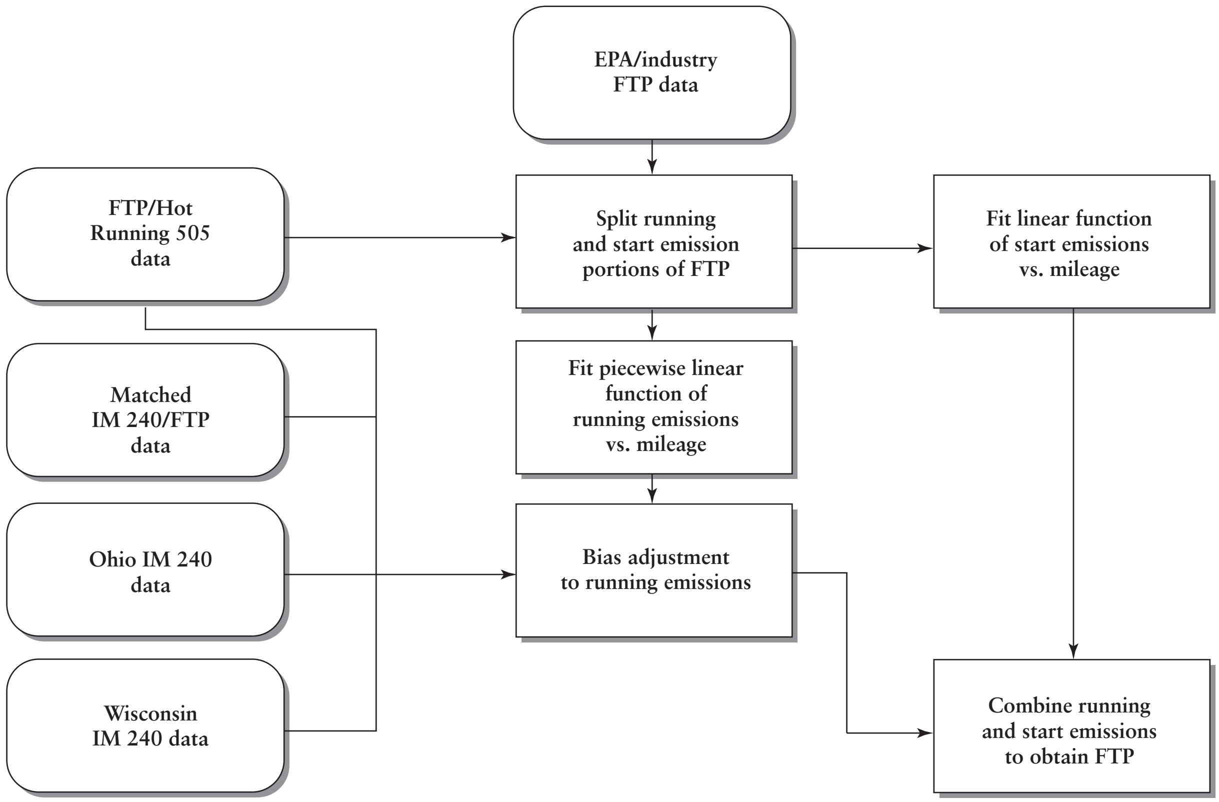


TABLE 2 Model Year and Technology Categories Used in MOBILE6 to Model 1981 to 1993 Light-Duty Vehicles

Model years	Technology
Cars	
1988–93	Port fuel injected (PFI)
1988–93	Throttle body injected (TBI)
1983–87	Fuel injected (PFI and TBI)
1986–93	Closed loop carbureted/open loop
1983–85	Closed loop carbureted/open loop
1981–82	Fuel injected (PFI and TBI)
1981–82	Closed loop carbureted/open loop
Trucks	
1988–93	Port fuel injected (PFI)
1988–93	Throttle body injected (TBI)
1984–93	Closed loop carbureted/open loop
1981–87	Fuel injected (PFI and TBI)
1981–83	Closed loop carbureted/open loop

"Open loop" refers to vehicles which do not use electronic feedback systems to control the delivery of fuel to the engine cylinders. Most current light-duty vehicles make use of feedback, or "closed loop," systems.

inspection of regression lines estimated from the available data. This approach was adopted after observing that simple linear regressions produced unrealistic fits of emissions at low mileage. This is attributed to a shortage of low-mileage FTP observations.

Modeling Start Emissions

With the start component of emissions, it was assumed that there are two categories of vehicles: "normal" and "high" emitters. This distinction facilitates the treatment of inspection and maintenance credits in MOBILE. A vehicle was assigned to its emitter class for a given pollutant depending on whether or not its FTP emissions exceeded an arbitrary multiple of its regulated emissions standard. For HC and NO_x, this multiple was chosen as two, while for CO it was three.

Following this classification, HC and CO deterioration in normal emitters was modeled as a simple linear function of mileage using least squares regression. For these pollutants, the high emitters were found to be uncorrelated with mileage, so they were taken to have constant emissions equal to the sample mean. At a given mileage, the normal and high values were combined in a weighted average, with the weights equal to the fractions of normal and high emitters associated with that mileage level. For NO_x emissions, the emitter class distinction was not used for modeling deterioration. A simple linear function of mileage was fitted using least squares regression.

Adjustment for Possible Bias

The adjustment for possible bias was based on the Dayton I/M data. It only applies to the Running LA4 equation since the IM 240 cycle does not contain an engine start. The adjustment was achieved through a series of steps that involved transforming fast-pass IM 240 scores into running FTP emissions. In the first step, a regression model was fitted to the Wisconsin full 240-second I/M data to predict full 240-second values from fast-pass scores. Separate models were determined for each pollutant.

Next, the matched IM 240 and FTP measurements from Indiana and Arizona were used to construct a full suite of running emissions from IM 240 measurements. This first required the use of the equation for Running LA4 emissions to compute a running emissions value for each FTP test. A separate regression equation derived from the matched pairs was then used to predict running emissions from each IM 240 measurement.

Finally, this model was applied to the Ohio IM 240 values to obtain running emissions estimates for each of the points in that large database. Associated with each of these values is a vehicle odometer reading and model year. However, the Dayton IM 240 data suffered from a problem with unreliable odometer readings. After an attempt to correct the data, it was decided to circumvent the problem by replacing the recorded mileage with average accumulated mileages from national vehicle travel surveys (Oak Ridge National Lab 1995). For each vehicle in the Ohio IM 240 data set, the model year and technology were identified and the corresponding average mileage was assigned. Then, average running emissions were computed for each model year and technology class. The emissions averages were then compared to the running emissions estimated from the piecewise linear functions at the appropriate mileages. This produced several differences in each of the model year/technology groups shown in table 2. Within each of these groups, the differences were smoothed using simple regression analysis, yielding additive adjustment factors equaling zero at mileage zero and changing linearly with mileage. Most of the adjustments are in the direction of increased emissions.

RESULTS

Figure 2 (a-c) illustrates the effect on running emissions of the bias adjustment for 1988-93 port fuel injected (PFI) cars. Running and start emissions can be reconstituted as a complete FTP estimate. Similar graphs for earlier model years are found in USEPA (1999a). The equation coefficients underlying these graphs, a part of the MOBILE6 computer code, also appear in that report.

Start emission estimates are reported in USEPA (1999d). The split between normal and high emitting vehicles described earlier enables the calculation of a fraction of high emitters at any mileage. This fraction increases with accumulated mileage. For the purpose of comparing MOBILE6 to MOBILE5, a composite of running and start emissions can be calculated to produce FTP estimates for the new model. The equation for this calculation is the simple linear function:

$$\text{FTP} = (7.5 \times \text{running} + 0.521 \times \text{start}) \div 7.5 \quad (4)$$

The constant 7.5 derives from the fact that in the LA4 cycle the total distance driven during the collection of bags 1 and 2 is 7.5 miles. Figure 3 (a–c) depicts MOBILE FTP HC deterioration functions for model years 1992, 1987, and 1981 fuel injected cars and compares the MOBILE5 kinked line to the proposed MOBILE6 line. This pattern is typical. For recent model years, MOBILE6 predicts emissions with considerably less deterioration than the earlier version whereas in older model years these differences are less pronounced.

Other Model Years

The equations used in MOBILE5 provided the basis for the modeling of pre-1981 open loop vehicles in MOBILE6. In the new version of the model, it is necessary to estimate the start and running

FIGURE 2 Running LA4 Emissions for 1998–1993 PFI Cars

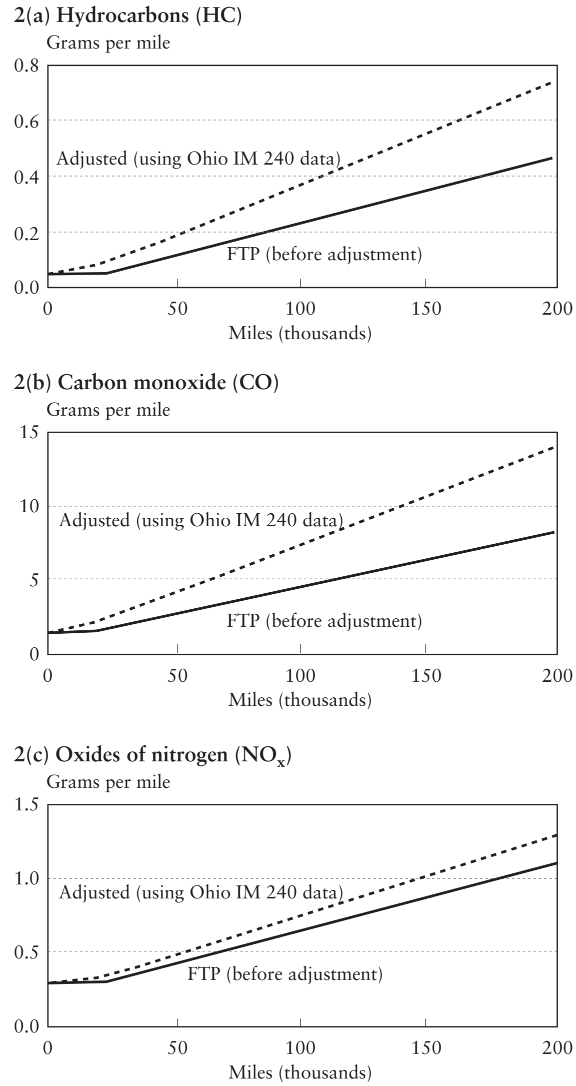
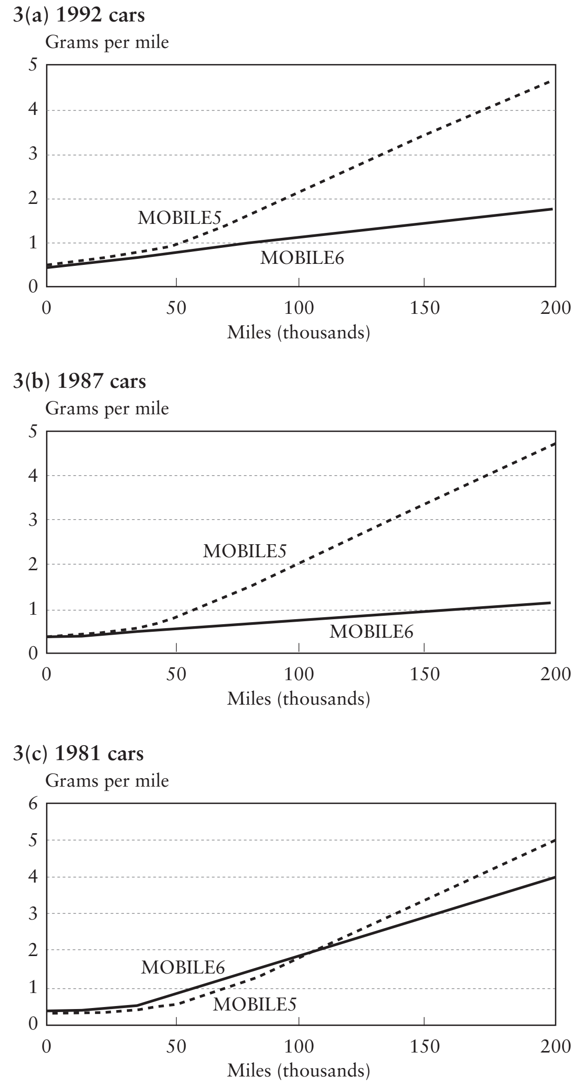


FIGURE 3 FTP Comparison of MOBILE5 and Proposed MOBILE6 Hydrocarbon (HC) Emission Factors



components of the FTP. This was accomplished by computing start and running emission fractions using the FTP data underlying the MOBILE5 deterioration functions together with the results of the 77-test HR505 data analysis (USEPA 1999e).

For model years 1994 and later, exhaust emissions deterioration is influenced by new regulations, including the introduction of On-Board Diagnostics (OBD) and enhanced I/M programs. Data with which to model deterioration in these vehicles were not available in sufficient quantity for use in the model revision project. In MOBILE6, these changes are modeled by assuming that emissions are reduced from earlier model-year levels in proportion to the tightening of standards for these newer vehicles (USEPA 1999f; USEPA 1999g).

Statistical Uncertainty

For the various regression analyses described above, goodness-of-fit, as measured by R^2, is superficially encouraging. For example, in the key regression models of running emissions and start emissions, R^2 values range from 0.922 to 0.953 for the three pollutants. Due to the natural skewness found in emissions data, log transformations were sometimes applied to emissions test values prior to fitting regression equations, a process which tends to produce higher R^2 values than with untransformed data. The EPA reports cited here include values of R^2 and other standard measures of goodness-of-fit for the various estimated models. When the equations are combined using the steps summarized in figure 1, overall confidence in the model coefficients is undoubtedly reduced.

CONCLUSIONS AND RECOMMENDATIONS

The MOBILE model is an important tool for planning and implementing air quality management. The changes to the model described in this paper have significant implications for decisionmakers responsible for developing programs of emissions control. In particular, the reduced rates of emissions factor deterioration in newer vehicles could lead to a re-evaluation of control strategies in areas not in attainment with federal standards.

The MOBILE6 emissions inventory model includes some important modifications to earlier versions of the model. When separated into running and start components, the estimation of deterioration in basic emission rates poses a difficult challenge. The work described in this paper represents EPA's current approach to addressing that challenge.

Given the nature of available data, modeling deterioration of vehicle exhaust emissions requires considerable judgement and experience. For the purpose of the MOBILE model, there is also a strong incentive to apply simple, easily understood statistical methodologies. These principles guided the model construction described in this paper, despite the apparent complexity of the overall scheme. As noted earlier, the more complete measurement of confidence in the final basic emission rate (BER) equations would be a worthwhile undertaking.

The shortcomings of the data used in this work underscore the need for better test program design and data measurement. Emissions testing is expensive, and frequently it is not possible to obtain data according to the requirements of good experimental design. A large dataset, like that obtained from a state I/M program, does not guarantee satisfactory results in the absence of other desired statistical criteria. Many of these concerns would be reduced with closer collaboration between practitioners in the fields of emissions testing and emissions modeling, and EPA is currently instituting programs toward that end.

The state of California maintains a parallel emissions modeling program designed to support its somewhat more stringent air quality regulations. The On-Road Emissions Inventory Estimation Model (EMFAC) is similar in many ways to MOBILE, and there is a high degree of coordination between the California and EPA modeling efforts. Nevertheless, there are substantial differences between the two models in terms of underlying data, assumptions, and methodology. Comparing predictions generated by these models would be a challenging but useful exercise.

REFERENCES

Mobile Sources Technical Review Subcommittee (MSTRS). 1997. *Report of the In-Use Deterioration Workgroup*. Washington, DC.

U.S. Department of Transportation, Federal Highway Administration. 1995. Nationwide Personal Transportation Survey data. Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1992. *I/M Costs, Benefits, and Impacts*. Washington, DC.

___. 1993. *Federal Test Procedure Review Project: Preliminary Technical Report*. EPA 420-R-93-007. Washington, DC.

___. 1994. *User's Guide to MOBILE5 (Mobile Source Emission Factor Model)*. EPA-AA-AQAB-94-01. Washington, DC.

___. 1999a. *Determination of Running Emissions as a Function of Mileage for 1981–1993 Model Year Light-Duty Cars and Trucks*. EPA420-P-99-010. Washington, DC.

___. 1999b. *Analysis of Emissions Deterioration Using Ohio and Wisconsin IM 240 Data*. EPA420-P-99-013. Washington, DC.

___. 1999c. *Determination of Hot Running Emissions from FTP Bag Emissions*. EPA420-P-99-014. Washington, DC.

___. 1999d. *Determination of Start Emissions as a Function of Mileage and Soak Time for 1981–1993 Model Year Light-Duty Vehicles*. EPA420-P-99-015. Washington, DC.

___. 1999e. *Basic Exhaust Emission Rates of Open Loop Vehicles for MOBILE6: Exhaust Emissions at High and Low Altitudes for Engine Starts and Running Emissions for Motorcycles, Light-Duty Diesel Vehicles and Trucks and Pre-1981 Model Year Light-Duty Gasoline Vehicles and Trucks*. EPA420-P-99-020. Washington, DC.

___. 1999f. *Determination of NO_x and HC Basic Emission Rates, OBD and I/M Effects for Tier 1 and Later LDVs and LDTs*. EPA420-P-99-009. Washington, DC.

___. 1999g. *Determination of CO Basic Emission Rates, OBD and I/M Effects for Tier 1 and Later LDVs and LDTs*. EPA420-P-99-017. Washington, DC.

The Use of Mixed Effects ANCOVA to Characterize Vehicle Emissions Profiles

MICHAEL L. DEATON

JAMES J. WINEBRAKE

James Madison University

ABSTRACT

This paper uses a mixed-effects analysis of covariance model (with both fixed and random effects) to characterize mileage-dependent emissions profiles for any given group of vehicles having a common model design. Such profiles are useful for evaluating, for example, how emissions will change over time within a new line of vehicles. The U.S. Environmental Protection Agency uses these types of evaluations to certify whether or not new models conform to existing emissions standards. Given such a group of vehicles, the statistical model introduced in this paper describes both the average emissions profile for that group while also accounting for individual vehicle variability among vehicles within the group. The model can be used to provide realistic confidence bounds for the average emissions deterioration profile within a given group, therefore allowing accurate emissions comparisons of multiple groups. The approach is illustrated with a sample of emissions data from two types of vehicles: natural gas Dodge Ram vans and gasoline Dodge Ram vans (all from the 1992–94 model years). The population profile for nonmethane hydrocarbons is explored. The results indicate the presence of vehicle-to-vehicle variation within each

Michael L. Deaton, James Madison University, Integrated Science and Technology Program, MSC 4102, Harrisonburg, VA 22807. Phone: 540–568–2725. Fax: 540–568–2955. Email: deatonml@jmu.edu.

vehicle type. This variation leads to confidence profiles that can be markedly different (but more appropriate) than what would be obtained from a simple fixed-effects regression model. The results highlight the potential for incorrectly characterizing emissions profiles whenever decisionmakers rely on standard regression techniques.

INTRODUCTION

Policymakers who establish emissions standards for new vehicles often focus on both the *baseline* emissions when the automobile is new, as well as on the rate at which those emissions *deteriorate* with vehicle age and use. Unfortunately, the emissions (as well as the emissions deterioration rate) from any individual vehicle after a specified amount of use can vary significantly from the average emissions of all similar vehicles under the same conditions. Hence, when evaluating average emissions across a population of vehicles that are nominally identical (same make, model, design) but utilize new technologies (such as alternative fuels), it is necessary to characterize the emissions profiles (average emissions as a function of mileage traveled) for the population, while also accounting for variation among vehicles within the population.

Over the past several years, many studies have attempted to collect and analyze emissions from in-use alternative fuel vehicles (AFVs) (i.e., AFVs operating in normal, daily driving conditions). Examples of these studies for light-duty vehicles include the work of Gabele (1990, 1995), Kelly et al. (1996a, 1996b, 1996c), Kirchstetter et al. (1996), Norbeck et al. (1998), Durbin et al. (1999), and Whalen et al. (1999). Examples from the heavy-duty literature include Clark et al. (1998), Chandler et al. (1999), and McCormick et al. (1999).

One significant data-collection effort has been funded by the U.S. Department of Energy and managed by the National Renewable Energy Laboratory (NREL). This program has collected emissions data from over 400 AFVs and gasoline control vehicles operating in federal government fleets. These vehicles operate on a variety of fuels, including methanol blends, ethanol blends, compressed natural gas, and propane. Vehicles are operated in various federal agency fleets and represent a variety of driving conditions and operations. The National Alternative Fuels Data Center (AFDC), located in Golden, Colorado, collects and publishes data from these emissions tests.

Policymakers are interested in the results of such studies in order to evaluate the potential impact of AFVs on air pollution. This necessarily requires that researchers develop models for the emissions generated by these vehicles over their useful lifetime. These emissions profiles may then be used to characterize lifetime emissions for those vehicles and to help establish standards for acceptable emissions levels at various points in a vehicle's lifetime.

The goal of this paper is to illustrate one approach that evaluates an assumed functional relationship between emissions and mileage, but also attempts to properly incorporate and account for variation in emissions from one vehicle to another. In doing so, a more complete understanding of the average deterioration in a group of vehicles and of the variation among vehicles and between fuel types is possible.

The statistical model described in this paper is a generalization of the classic analysis of covariance (ANCOVA) model. This approach is more precise than conventional regression models because it accounts for both engine age (as measured indirectly by odometer readings) and variations between vehicles of the same make and model.[1] Furthermore, the generalized ANCOVA allows more realistic estimates of the variation inherent in comparisons between vehicles operating on different fuels and allows more realistic estimates of the size of confidence bands for the average emissions across all vehicles and also for individual vehicle emissions.

The second section illustrates the impact that variations among vehicles can have on estimated emissions profiles and on the width of confidence bands for the average emissions profile. We use a simple example to illustrate the key concepts. We demon-

[1] The statistical model presented in this paper can be generalized to describe emissions profiles in populations containing a variety of vehicle designs, model years, etc. This more generalized model would be useful for characterizing the emissions of a highly diversified population (i.e., a fleet owned by a large corporation or government agency). However, this paper focuses on the more restrictive problem of characterizing the emissions profile in a group of vehicles that are nominally identical with respect to model design, engine type, etc.

strate that evaluations of emissions profiles that fail to properly account for vehicle-to-vehicle variation can lead to confidence bands that give overly optimistic estimates of the precision with which the average emissions profile (averaged across all vehicles in the group of interest) can be determined.

The third section describes a general mixed-effects ANCOVA model that may be used to: 1) estimate emissions profiles in one or more groups of vehicles, and 2) compare emissions profiles among those groups. This model accounts for random variations between vehicles, thereby avoiding the pitfalls illustrated in the second section.

The final section demonstrates the use of the general ANCOVA model described earlier by analyzing nonmethane hydrocarbon (NMHC) emissions from 58 in-use vehicles selected from the AFDC database. All 58 vehicles are Dodge Ram vans with the same engine size, and all from model years 1992–94. Twenty-seven of these vehicles ran exclusively on compressed natural gas, while the other 31 vehicles were dedicated to the exclusive use of California Phase II reformulated gasoline (RFG).

THE IMPACT OF VEHICLE-TO-VEHICLE VARIATION ON ESTIMATED EMISSIONS PROFILES

One seemingly common-sense approach to evaluating emissions profiles over vehicle lifetimes is to express emissions as a simple linear function of mileage (thereby indirectly accounting for deterioration effects). That is, one can fit the simple linear regression model

$$Y_{ij} = \alpha + \beta m_{ij} + \varepsilon_{ij} \tag{1}$$

where Y_{ij} is the j^{th} emissions reading on the i^{th} car taken at odometer reading m_{ij}. This model assumes that emissions are a linear function of mileage. This model is also based on the important assumption that the only random variation in emissions comes from the error term ε_{ij}.

Such an approach, however, does not adequately account for the inherent variation among individual vehicles within a group.[2] Hence, the resulting confidence bands for the *average group-wide average emissions profile,* as well as the tolerance bands giving estimates of the expected *range* of emissions from individual vehicles, are often too narrow. This failure to account for vehicle-to-vehicle emissions variability may also lead to incorrect statistical testing and estimation procedures, thereby making it difficult to reliably detect differences between groups of vehicles and fuel types.

In order to illustrate these concepts, imagine the case in which one randomly selected new car is used to evaluate the population-wide average emissions profile for all similar vehicles.[3] This vehicle is driven for 100,000 miles on a test track and its NMHC emissions are measured every 10,000 miles. This imaginary study would provide 10 ordered pairs of data (miles driven, NMHC emissions). The common-sense approach described above would use these 10 observations to fit a simple model of the form given in equation (1), where Y_{ij} is the measured NMHC emissions of the i^{th} car after m_{ij} miles of driving; m_{ij} is the miles driven by car i on the j^{th} measurement, and ε_{ij} is the random variation due to unexplained factors.[4] It is typically assumed that the ε_{ij}'s are independently distributed from a normal distribution with a mean of zero and a standard deviation of σ_ε. Under this traditional regression model (which does not account for vehicle-to-vehicle variation), the population-wide average emissions $E(Y)$ after m miles of driving is given by

$$E(Y) = \alpha + \beta m. \tag{2}$$

Conventional least-squares estimation of the above model leads to estimates of α and β, which are designated as $\hat{\alpha}$ and $\hat{\beta}$. Using these well-known results, along with the simplifying assumption that

[2] A *group* of vehicles is defined here as all vehicles that are nominally *identical* with respect to make, model, engine size, year, and fuel type. The analysis reported herein assumes that a random sample of vehicles from this group has been taken and the emissions monitored over an extended mileage range.

[3] It is clear that the use of one vehicle to characterize the emissions profile for an entire group of similar vehicles is not a very sound practice. However, this simple case will be used here in order to simplify the mathematical presentation. Moreover, the Environmental Protection Agency's emissions certification program requires manufacturers to test only one vehicle in order to estimate emissions profiles for an entire population of similar vehicles (Hormes 2000).

[4] Note that the subscript i is not necessary here, but is included to emphasize the fact that the i^{th} car in the population has been selected. The reasons for including this notation will be evident later.

the error standard deviation σ_ε is known, the conventional regression approach will lead to the following quantities of interest (Graybill 1996).

1. The estimated emissions profile:

$$\hat{E}(Y) = \hat{\alpha} + \hat{\beta}m. \quad (3)$$

2. A 95% confidence band for the average emissions (averaged across all vehicles in the population) at mileage m:

$$\hat{E}(Y) \pm 1.96 \cdot \sigma_\varepsilon \cdot \sqrt{\frac{1}{n} + \frac{n(m-\bar{m})^2}{n\Sigma m_{ij}^2 - (\Sigma m_{ij})^2}} \quad (4)$$

3. A 95% prediction band for the emissions of an *individual vehicle* at mileage m:

$$\hat{E}(Y) \pm 1.96 \cdot \sigma_\varepsilon \cdot \sqrt{1 + \frac{1}{n} + \frac{n(m-\bar{m})^2}{n\Sigma m_{ij}^2 - (\Sigma m_{ij})^2}} \quad (5)$$

Note that the quantity $\bar{m}$ in equations (4) and (5) stands for the average mileage odometer reading in the data, and n is the total number of observations in the study (n = 10 in this example).

Now suppose that there is a sizeable difference in emissions levels between vehicles in the population. For simplicity, assume that all the vehicles in the population exhibit the same deterioration rate of NMHC emissions (i.e., the value of β is the same for all vehicles in the population), but that the baseline emissions value is different from one vehicle to another (i.e., the intercept varies between vehicles). In this case, we can generalize the model in (1) to be

$$Y_{ij} = \alpha + \upsilon_i + \beta m_{ij} + \varepsilon_{ij}. \quad (6)$$

Notice that the only difference between (6) and the traditional model in (1) is that quantity υ_i has been added to the intercept. With this model, α is the average value of the intercept (averaged across all vehicles in the population), and the quantity υ_i is the amount that the intercept for vehicle i deviates from the population-wide average (α). Here, all vehicles in the population exhibit emissions profiles that follow the same slope, but these profiles of are offset from one vehicle to the next.

Assuming the vehicle in the study was randomly selected, the value of υ_i is random. Moreover, if the value of α is unknown, the value of α and υ_i cannot be uniquely determined from the data. It is typically assumed that the values of the υ_i in the population are independent and follow a normal distribution with a mean of zero (i.e., the average intercept across all vehicles in the population is α) and a standard deviation of σ_υ (i.e., the intercepts vary randomly from vehicle-to-vehicle, and the standard deviation of intercepts from all vehicles is σ_υ).

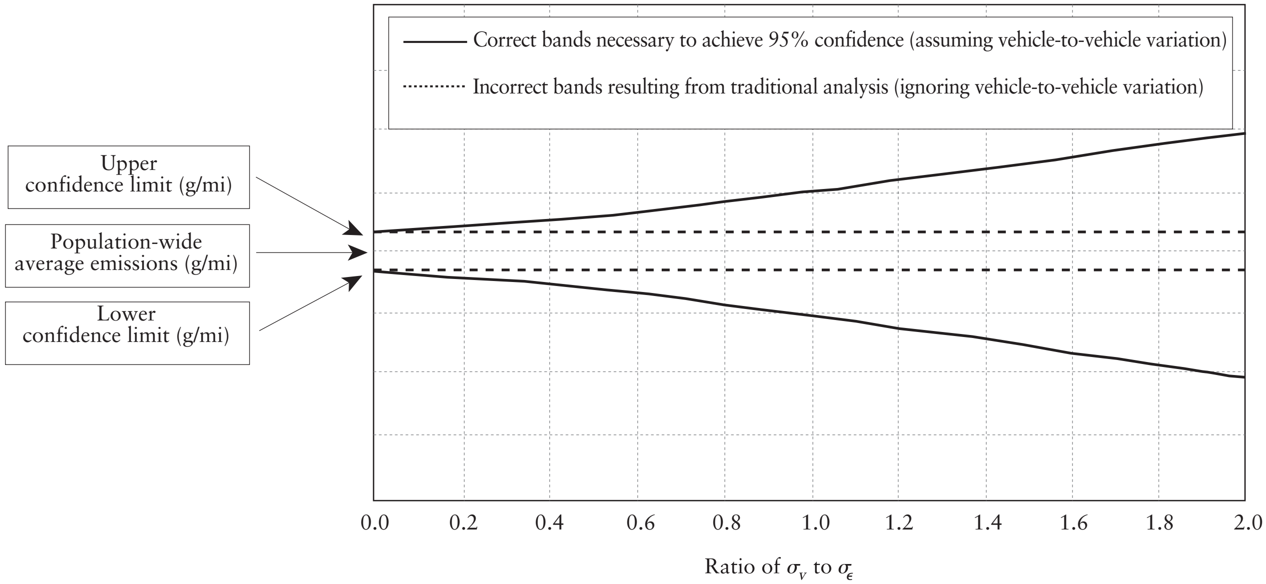
Now suppose that the researcher fails to recognize the structure in (6), and fits the model in (1) using standard least squares techniques; that is, he fits a model that fails to account for the random variation between vehicles.[5] Given these assumptions, Appendix A shows that the following are true:

1. The estimated average profile given in (3) still gives an unbiased estimate of the population-wide average emissions; and
2. A 95% confidence band in (4) for the population-wide average emissions and a 95% prediction band in (5) for predicting the emissions of an individual vehicle after m miles of use are too narrow. The discussion below elaborates on this point.

Statement 2 above is supported by figure 1, which illustrates the width of a 95% confidence band for population-wide average emissions at 55,000 miles in the hypothetical example. The theoretically correct 95% bandwidth is spanned by the outside, solid-line curves. Confidence intervals that have a 95% probability of including the actual population-wide average emissions have an expected bandwidth that corresponds to the solid-line curves. The bandwidth of the traditional interval, as determined from equation (4) above, is spanned by the inside, dashed-line curves. Confidence intervals based on this bandwidth will have less than 95% probability of including the true population-wide average emissions. The x-axis displays the ratio of the vehicle-to-vehicle

[5] The study design in our example would be inadequate for detecting vehicle-to-vehicle variation. If vehicle-to-vehicle variation was believed to be present, care would be taken to collect data from several randomly selected vehicles from the fleet. Using the techniques described later in this paper, the value of the vehicle-to-vehicle standard deviation could then be estimated.

FIGURE 1 Comparison of Confidence Bands on Population-Wide Average Emissions When Vehicle-to-Vehicle Variation is Present



standard deviation (σ_v) to the error standard deviation (σ_ϵ). Hence, when this ratio is zero, there is no vehicle-to-vehicle variation and the traditional approach is appropriate. Notice that when the ratio on the x-axis is zero, the "correct" confidence band and the band from traditional regression are identical.

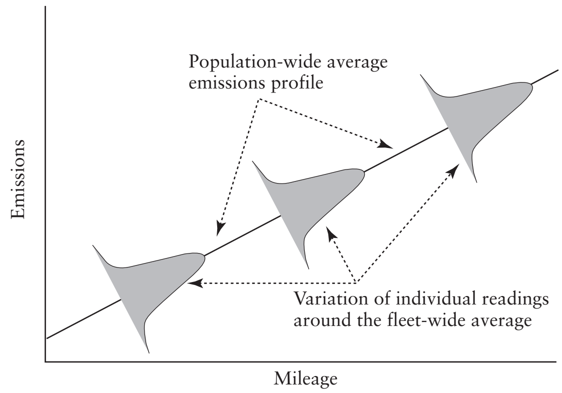
On the other hand, when the ratio on the x-axis is large, the vehicle-to-vehicle variation is also large. In such a case, the traditional regression model fails to account for the additional source of variation between vehicles. For example, consider the case when the vehicle-to-vehicle variation is the same size as the error variation (i.e., the ratio on the x-axis is equal to 1). It is clear from figure 1 that the traditional confidence band is too narrow by a factor of 3 or more. Hence, in this case, the traditional approach leads to a grossly over-optimistic picture of how precisely the population-wide average emissions profile may be estimated. In fact, even if the size of vehicle-to-vehicle variation is small (as when the ratio on the x-axis is 0.4 to 0.6), the error in the confidence bandwidth can be large. In such a case, the use of the conventional simple linear regression model in (1) will lead to confidence bands that are advertised to have a 95% confidence level, but that have a much lower confidence level in reality.[6]

Figure 1 illustrates the practical implications of vehicle-to-vehicle variation. Figures 2, 3, and 4 illustrate this in a slightly different way. Figure 2 illustrates the case in which there is no vehicle-to-vehicle variation. In this case, all the vehicles in the population have an assumed common emissions profile, indicated by the solid line. However, because of random variations from one measurement of emissions to the next (due to imprecision in laboratory methods, etc.), a given vehicle's emissions measurement at a particular mileage will vary randomly around the population-wide profile. This variation is represented by the bell-shaped curves spaced along the line. Each bell-shaped curve represents the distribution of emissions measurements that one could expect to see at the specified mileage reading.

Figure 3 illustrates the case in which each vehicle in the population has its own emissions profile. More specifically, figure 3 represents the case in which all of the profiles are parallel (i.e., the rate of emissions deterioration is constant for all vehicles),

[6] The error in the confidence band will not be as great if multiple cars are included in the sample. Nonetheless, even if multiple cars are sampled, the error in the confidence bandwidth can still be sizeable, provided that the vehicle-to-vehicle variation is large.

FIGURE 2 Illustration of the Case in Which All Vehicles Have the Same Emissions Profile



while the intercept of the emissions profile varies from one vehicle to the next. This corresponds to the model in (6). Notice that each individual line in figure 3 also displays several bell-shaped curves that represent the distribution of actual emissions measurements from each individual car at a given mileage. Figure 4 superimposes on figure 3 the population-wide average emissions profile, along with a corresponding set of bell-shaped curves along that profile. Notice that the bell-shaped curves in figure 4 are much wider than in figure 2 where no vehicle-to-vehicle variation is present. This is because the collection of emissions readings from a randomly selected car at a fixed mileage will vary from the population-wide average due to random error variation (σ) and because of variations between vehicles (σ_v).

Hence, if vehicle-to-vehicle variation is present in the form indicated in equation (6), then regression analysis that is based on the simple linear model in (1) will lead to confidence bands and prediction intervals that can be highly inefficient and possibly even deceptive. Policymakers who rely on such estimates to make comparisons between different groups of vehicles (e.g., vehicles operating on different fuels) run a sizeable risk of making decisions that do not realistically reflect the actual capabilities of those populations.

FIGURE 3 Illustration of the Case in Which Individual Vehicles' Emissions Profiles Have Different Intercepts from One Vehicle to the Next

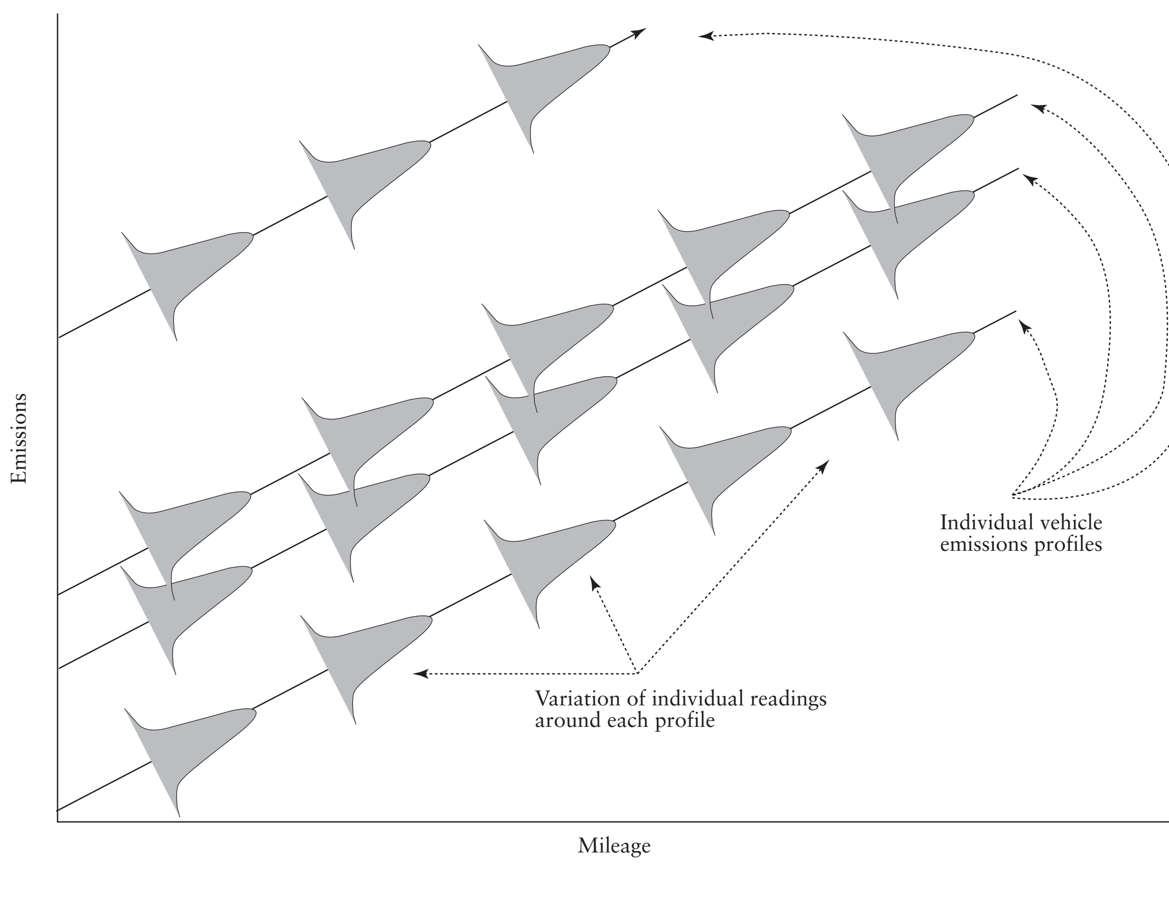
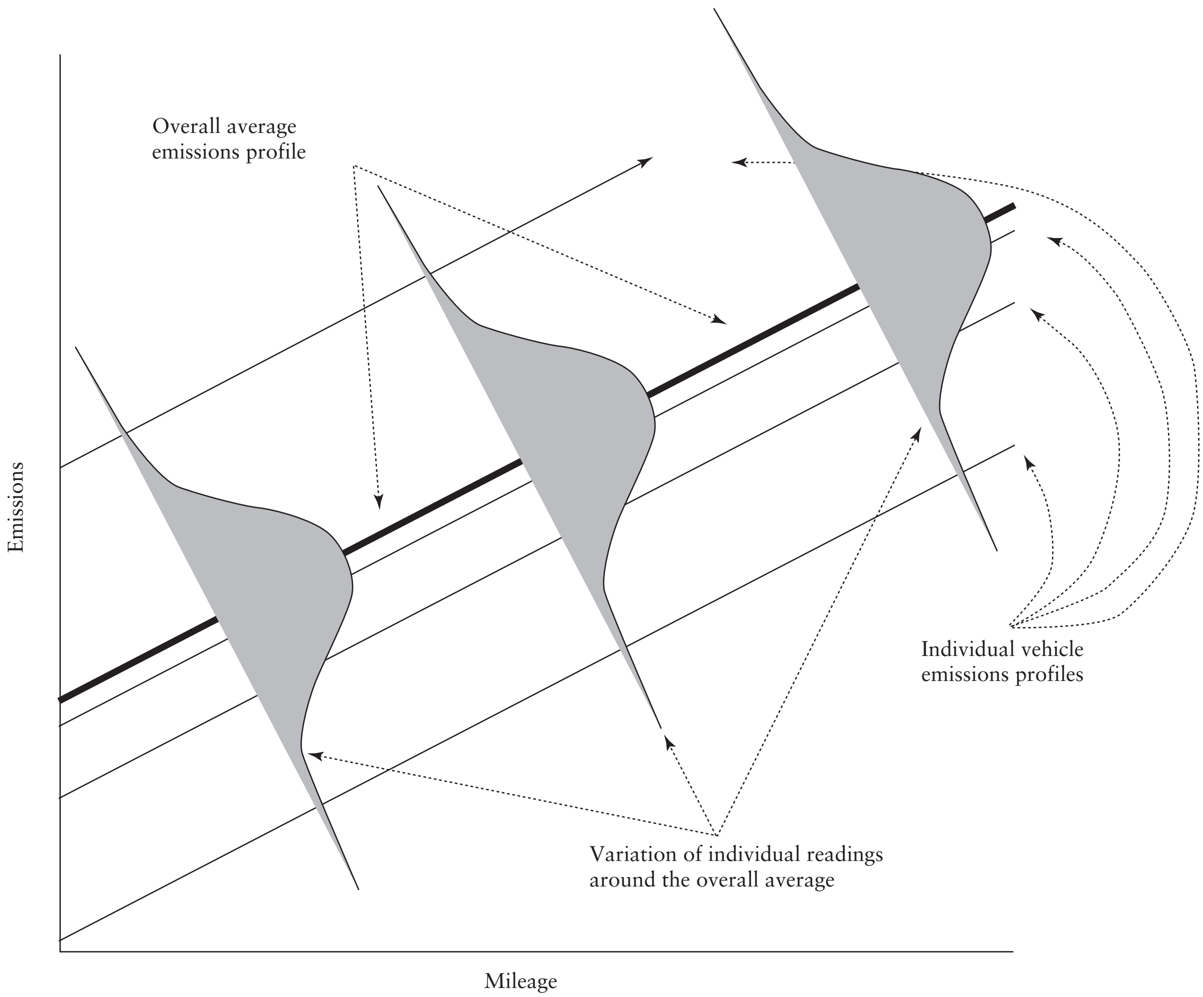


FIGURE 4 Average Emissions Profile in the Presence of Vehicle-to-Vehicle Variation



Consider, for example, the case in which the vehicle-to-vehicle standard deviation is the same size as the error standard deviation. Hence, for this case, the ratio on the horizontal axis of figure 1 is 1.0. In such a case, if model (1) is used to characterize the population-wide emissions profile, then the resulting 95% confidence bands for average emissions at 55,000 miles will be too narrow by approximately 70%, and the confidence level for those bands will in fact be much lower than 95%. Such an error can lead policymakers to have an overly optimistic picture of how variable emissions will be from vehicles in this population. This misunderstanding can lead to emissions standards that are unreasonably tight.

Figure 1 also suggests that the type I error rate (i.e., the α level) associated with traditional hypothesis testing procedures can be much greater than the advertised level whenever vehicle-to-vehicle variability is present and is not properly accounted for in the analysis. This means that chances of spurious statistically significant results can be much greater than the advertised α-level when vehicle-to-vehicle variability is ignored. For example, suppose that the vehicle-to-vehicle standard deviation is the same size as the error standard deviation and a two-sample t-test with an α-level of 0.05 is used to compare a group of alternative fuel vehicles to a corresponding group of conventional fuel vehicles. Further suppose that the analysis did not properly account for the vehicle-to-vehicle variation. Based on the earlier argument, the resulting hypothesis test may in fact have an α-level that is much greater than the advertised α-level of 5%. This means that the researcher has much more than a 5% risk of incorrectly finding a difference between the two groups of vehicles when no such difference really exists.

Model (6) is more realistic than model (1) because it allows for potential variations between vehicles in a population. However, (6) can be further improved by allowing for different deterioration rates as well as different baseline emissions from one vehicle to another. In such a case, one would expect that the problems with the confidence bands and prediction bands from a traditional regression model would be even more acute than illustrated here. The next section introduces this more general model and also incorporates terms that allow for statistical comparison of different populations or fuel types.

DESCRIPTION OF A GENERAL MODEL

The statistical model used in this study relies on the general methodology of *analysis of covariance* (ANCOVA) discussed in Searle (1971). This model can be used to compare two or more "treatments" that have been applied to a group of individuals. In the present study, the "group" consists of individual vehicles assumed to be nominally "identical" with respect to make, model, engine size, fuel type, etc. The treatments are the different fuels under which these vehicles are operated. The response of interest is the emissions of a given pollutant. The simplest ANCOVA model accounts for the fact that the response (i.e., emissions) depends on a "covariate" (i.e., mileage driven), which can change from one observation to the next. In this sense, the ANCOVA model is a general application of the standard analysis of variance (ANOVA) in which one or more treatments are compared, but in which there is no covariate.

The model illustrated here generalizes the simplest ANCOVA model to also account for the random variation between vehicles within the population. By doing so, the analyst is afforded accurate tests for comparing emissions profiles among fuel types and for comparing emissions at any specified mileage. The approach is well established in the statistical literature (see, e.g., Searle 1971 and Federer and Meredith 1992), but it has received little attention in the field of emissions modeling (one exception is a study conducted by Battelle Memorial Institute (1995)).

Let Y_{ijk} represent a specific emissions constituent as observed on the k^{th} test on the j^{th} vehicle that is operating on fuel type i. Let $m_{k(i,j)}$ stand for the k^{th} mileage reading on car j operating on fuel type i. It is assumed that only one emissions result is obtained at each mileage reading on each vehicle (but the model can be generalized to handle multiple measurements). The model has the form:

$$Y_{ijk} = [\alpha + \beta \bullet m_{k(i,j)}] + [\phi_i + \delta_i \bullet m_{k(i,j)}] + [\upsilon_{j(i)} + \overline{\omega}_{j(i)} \bullet \mathrm{m}_{\kappa(i,j)}] + \varepsilon_{ijk} \quad (7)$$

The first two terms $[\alpha + \beta \bullet (m_{k(i,j)})]$ represent the *average* dependence of the emissions on vehicle mileage, regardless of which fuel type is used or the variation that is inherent among individual vehicles. The next two terms $[\phi_i + \delta_i (m_{k(i,j)})]$ represent how this average dependence is affected by fuel type i. The next two terms $[\upsilon_{j(i)} + \overline{\omega}_{j(i)}(m_{\kappa(i,j)})]$ represent how the average dependence is affected by the unique characteristics of vehicle j that operates on fuel type i.

This model allows for the realistic situation in which there is an overall population-wide deterioration curve that describes the average emissions for all vehicles in the group of interest that are using fuel type i. The group-wide emissions curve when operating on fuel type i is defined by the expression $\alpha + \beta \bullet (m_{k(i,j)}) + \phi_i + \delta_i (m_{k(i,j)})$. However, the model also accounts for the fact that each vehicle in the group may have an emissions curve that differs slightly from the average curve for all similar vehicles. This variation from the average curve can occur in either the intercept (through $\upsilon_{j(i)}$), the slope (through $\overline{\omega}_{j(i)}$), or through both the intercept and slope. The final term (ε_{ijk}) represents the random variation in emissions that are not accounted for in the model. This variation may be attributed to such things as variation from the test method used, differences between laboratories (if each car is tested at multiple labs), or any number of other factors.

The assumptions behind this model are stated as follows:

- Assumption 1: At a fixed mileage, emissions follow a normal distribution.
- Assumption 2: The quantities α, β, ϕ_i, and δ_i in the model in equation (7) are fixed, but unknown parameters. Moreover, since the ϕ_i and δ_i represent deviations from the mean intercept and slope, respectively, it assumed that $\sum \phi_i = \sum \delta_i = 0$. If the study is aimed at characterizing the emissions profile of a fixed or specified group of vehicles and for a fixed set of fuel types, then this fixed-effects assumption is

reasonable. However, if the study's goals are to characterize emissions across a wide variety of vehicles and fuel types, but data have been collected on only a random sample of vehicles and a random sample of fuel types, this assumption must be relaxed. The present study (and many other studies of practical interest) are consistent with this fixed effect's assumption.

- Assumption 3: $\upsilon_{j(i)}$, $\overline{\omega}_{j(i)}$, and ε_{ijk} are all random quantities. Each of these terms is assumed to follow a normal distribution having a mean of zero. The standard deviations of these distributions are σ_{υ} , $\sigma_{\overline{\omega}}$, and σ_{ε}, respectively. The standard deviations σ_{υ} and $\sigma_{\overline{\omega}}$ measure how much individual vehicle emissions profiles will vary around the population average emissions profile; that is, the larger σ_{υ} and $\sigma_{\overline{\omega}}$ are, the more individual vehicle emissions profiles may vary from the population average profile. It is also assumed that $\upsilon_{j(i)}$, $\overline{\omega}_{j(i)}$, and ε_{ijk} are mutually independent.

The reader should note that this model does not explicitly account for variation between the laboratories conducting the tests. The AFDC data analyzed in this paper were collected across three different laboratories, one of which was located at a high altitude. Lab-to-lab variation can be a dominant source of variation in these types of measurements. However, the model will provide a reliable test for comparing emissions from the two fuel types provided that (i) each car was tested at only one lab, and (ii) within each lab, vehicles from both fuel types were tested. Both requirements were satisfied by the data analyzed in this paper. Furthermore, under these assumptions, the lab-to-lab variation will be accounted for in the model, but will be indistinguishable from vehicle-to-vehicle variability. Hence, if the analysis suggests a large variation between vehicles within each group of interest, we cannot conclude that this source of variation is found only in differences between vehicles. It may partly be caused by variations between testing labs.

EXAMPLE APPLICATION: 58 DODGE RAM VANS FROM THE AFDC DATABASE

The ANCOVA model presented here was applied to emissions values from the AFDC database for 27 compressed natural gas (CNG) Dodge Ram vans and 31 gasoline counterparts (henceforth referred to as "RFG" for "reformulated gasoline"). Data was extracted on August 11, 1998. Several pollutants were measured on each car. Results for nonmethane hydrocarbons are analyzed and reported here.

Emissions tests on these vehicles were conducted at three commercial laboratories in various locations in the United States. A competitive bidding process was used to select the labs. A panel of experts (including U.S. Environmental Protection Agency—EPA—personnel) conducted site visits to ensure that standardized testing methods were used across all three labs and that appropriate quality assurance procedures were in place. Each vehicle was tested using the EPA's Federal Test Procedure (FTP) protocol at accumulated mileage readings of approximately 4,000 miles, 10,000 miles, and every 10,000 miles thereafter. Because of obvious logistical reasons, it is not the case that all the vehicles were tested at these exact mileage specifications. The general test procedures, emissions test driving profiles, and hydrocarbon specification procedures, along with other facts about the AFDC testing program and vehicles are reported elsewhere (Kelly et al. 1996a, 1996b, and 1996c).

Table 1 provides information about the vehicles, their fuels, and the number of vehicles per fuel (sample sizes). Note that all the CNG vehicles were original-equipment-manufactured Dodge Ram vans (i.e., none of the vehicles was an aftermarket conversion). Although no data are available on exactly how each vehicle was used, it is assumed that all the vehicles experienced similar driving conditions. This assumption may not be valid, and thus should be considered when interpreting the results of this analysis.

As shown in table 1, the alternative fuel vehicles come mostly from model year (MY) 1992, with fewer coming from MY 1994. The reverse is true for the RFG vehicles in the study. This discrepancy could jeopardize the ability to make comparisons of the CNG and RFG emissions if different emissions control systems had been installed on the

TABLE 1 Information on Vehicle Types and Fuels

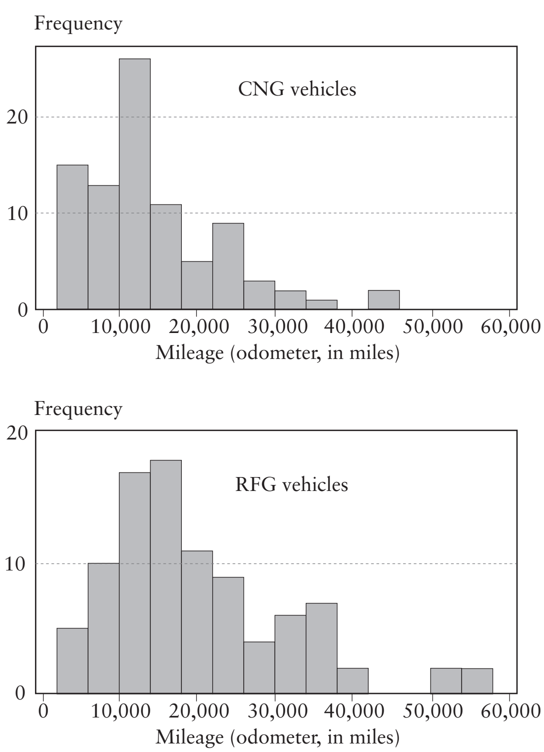
Vehicle type	N (by model year)
Dedicated original equipment manufactured CNG Dodge Ram B250 Van (CNG/Ram)	22 (1992) 5 (1994)
■ 5.2 liter V-8 engine configuration ■ Multi-point fuel injection ■ 4-speed automatic ■ 11.1–15.7 equivalent gallon fuel capacity ■ 6,400 lbs gross vehicle weight ■ LEV-certified	
RFG Dodge Ram B250 Van (RFG/Ram)	11 (1992) 20 (1994)
■ 5.2 liter V-8 engine configuration ■ Multi-point fuel injection ■ 4-speed automatic ■ 35 gallon fuel capacity ■ 6,400 lbs gross vehicle weight	

1992 vehicles as compared with the 1994 vehicles. This, however, is not the case: emissions control systems in MY 1992 and MY 1994 vehicles are identical for the Dodge Ram vans in this study.[7] It is also important to recognize that these vehicles are now 6 to 8 years old and that they represent emissions control technologies that may have been modified or even replaced. The reader is encouraged to keep in mind the fast pace at which emissions control technologies may change (especially for new AFVs), and to take the potential for new technological advancement into account when interpreting the results reported here. Beyond this issue, MY is given no further consideration in the modeling and analysis.

These NREL-tracked vehicles were FTP tested several times at each of several different mileages. The AFDC database contains weighted FTP (WT) test results for each vehicle at each mileage, which were used in the present study. The original AFDC database included data on over 450 vehicles and 13 different models. In order to provide a sample of vehicles that represented a uniform population with respect to model (body design and engine) and model year, only the data for Dodge Ram vans was used. This original sample included 108 such vehicles. Vehicles were eliminated that were tested at only one mileage reading or if the difference in mileage between the first test and last test was less than 4,000 miles. In addition, emissions tests at mileages less than 3,000 miles were eliminated due to the possibility of a "green catalyst" effect. These selection criteria left the final sample of 58 vehicles (27 CNG and 31 RFG vehicles).

A comparative frequency distribution of the collective mileages with all tests on all 58 vehicles is shown in figure 5. The average mileage for all tests on all CNG vehicles is 14,159 miles, with a median of 11,397 and a maximum of 45,159. The average mileage for all tests on all RFG vehicles is 20,217 miles, with a median of 17,206 and a maximum of 57,099. It is impossible to determine from the available data whether these differences are due to variations in trip duration, trip frequency, or both. It should be noted that the original experimental design specified that all vehicles be tested at the same mileage readings through the course of the study. This allows emissions profiles to be equitably monitored across all vehicles, thereby simpli-

FIGURE 5 Mileage Frequency Distribution for Natural Gas (CNG) and Reformulated Gasoline (RFG) Vehicles



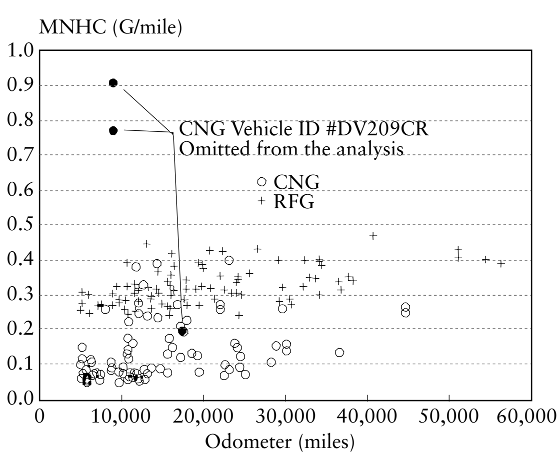
[7] Note, however, that the emissions control equipment for the CNG vehicles is designed for operation on CNG and is different from the equipment used in RFG vehicles.

fying the interpretation of the analysis. Unfortunately, due to the logistical limitations and the large scope of the study, this ideal was not strictly achieved (as illustrated by the non-uniform distribution of mileages in figure 5). While this departure from the intended design complicates the analysis somewhat, it does not invalidate the approach described here. Furthermore, the statistical model discussed above characterizes emissions deterioration only for the specific range of mileages covered in the data. At the outer limits of this range, the precision of the estimated profile is less than at the center of the range where more data are available. This is reflected in wider confidence bands around predicted emissions at high mileages.

Figure 6 visually displays the raw data for all 58 vehicles. A difference in NMHC emissions between the fuel types is suggested in this plot. In addition, the rate of increase in NMHC emissions does not exhibit any sizeable difference between the two fuel types. Both of these features are formally addressed and tested in the analysis.

Figure 6 also exhibits two outliers. These both came from one CNG vehicle that yielded much higher NMHC emissions in its first readings than in subsequent readings. That vehicle's data were omitted from the analysis.

FIGURE 6 Plot of NMHC Emissions vs Odometer Reading for 58 Dodge Ram Vans



RESULTS

As previously noted, the ANCOVA model presented in equation (7) is used to determine whether statistically significant differences exist in the average emissions profile between vehicles operating on different fuels (CNG and RFG), while also accounting for the variations that are inherent from one vehicle to another. The emissions profiles generated by this model estimate the average emissions values that can be expected for a group of vehicles operating on each particular fuel type at any given mileage.

Average emissions values for each fuel type were determined by fitting the complete model in equation (7) using the PROC MIXED procedure in SAS, version 6.12. A listing of the appropriate SAS code is provided in Appendix B. Parameter estimates and their variances were found, allowing the generation of predicted values and confidence bands for the average population-wide emissions component of the model when operating on a particular fuel type. In other words, values and confidence bands were determined for E_i, where E_i is the average emissions from vehicles when operating on fuel type i at a specific mileage m, as follows

$$E_i = \alpha + \beta \cdot m + \phi_i + \delta_i \cdot m \qquad (8)$$

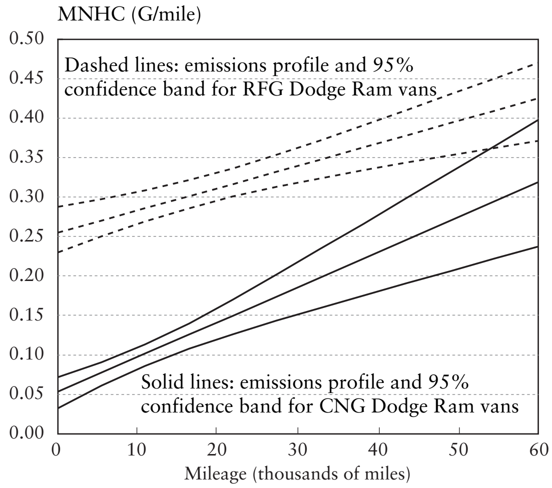
The NMHC emissions profiles in equation (8), along with their 95% confidence intervals, are plotted in figure 7.

The analysis also provides estimates of the error variance (σ^2_ε) and the two variances associated with vehicle-to-vehicle variation (σ^2_ν and σ^2_ω). Table 2 displays these estimates for NMHC. Recall that figure 1 demonstrates that when the vehicle-to-vehicle variation is large relative to the error variation, a model that fails to account for such variation will lead to confidence intervals that are too narrow for the stated level of confidence. Figure 1 shows that the larger the *ratio* of vehicle-

TABLE 2 Estimated Vehicle-to-Vehicle Variation and Error Variation for NMHC Among 58 Dodge Ram Vans

Variance component estimate		
σ^2_ε	σ^2_ν	σ^2_ω
0.0033	0.0000	0.0012

FIGURE 7 NMHC Emissions Profiles with 95% Confidence Bands for Dodge Ram Vans (gasoline vs. natural gas models)



to-vehicle standard deviation to error standard deviation, the more misleading are the confidence intervals (or hypothesis tests) derived from using an incorrect model. From table 2, the ratio in figure 1 is calculated by

$$\sqrt{\frac{\text{vehicle-to-vehicle variance}}{\text{error variance}}} = \sqrt{\frac{\sigma^2_{\nu} + \sigma^2_{\varpi}}{\sigma^2_{\varepsilon}}} = 0.61$$

Using this value on the x-axis in figure 1 suggests that a traditional analysis that fails to account for this large variation between vehicles can lead to confidence bands that are too narrow by a factor of approximately 50%. If emissions standards for in-use vehicles are based on such analyses, those standards may in fact provide an unrealistic picture of the range of emissions to be expected over the lifetime of any group of vehicles.

Table 3 summarizes the results of standard ANCOVA F-tests used to compare the average emissions profiles between the two fuel types. The *F-test for different slopes* in table 3 indicates whether the *rates* of emissions deterioration are the same for both fuel types. If this test is significant, there is strong evidence that the slopes of the NMHC emissions profiles differ between the two fuel types. If the first F-test is not significant, the second F-test (*F-test for a common nonzero slope*) and third F-test (*F-test for a common intercept*) should be examined. If the second test is significant (and the first F-test is not significant), it is safe to conclude that NMHC emissions do change with mileage and that the two groups of vehicles exhibit parallel (and possibly identical) profiles. If the third F-test is significant (and the first F-test is not significant), it is safe to conclude that the two groups of vehicles exhibit parallel, but distinct emissions profiles. Those profiles may be "flat" (unchanging with mileage) or they may exhibit a common nonzero trend, depending on whether or not the second F-test is significant.

TABLE 3 ANCOVA F-Test Results for Comparing NMHC Emissions Profiles Between CNG and RFG Vehicles

F-test for different slopes (p-value)	F-test for a common nonzero slope (p-value)	F-test for a common intercept (p-value)
Not significant	Significant	Significant
(0.1394)	(0.0001)	(0.0001)

Figure 7 displays the estimated emissions profiles for NMHC in both types of vehicles. With respect to NMHC, the CNG vehicles in the study appear to be cleaner than their RFG counterparts across all mileages. This is supported by the F-tests in table 3. The *F-test for slope* and the *F-test for a common nonzero slope* jointly indicate that there is a common nonzero slope in the NMHC emissions profiles for both groups of vehicles. The *F-test for a common intercept* in table 3 indicates that, while the two profiles appear to have a common slope, they are distinct. Combining these results with figure 7, it can be seen that the CNG Rams represented in this study indeed have lower average NMHC emissions than the RFG Rams throughout the mileage range covered and that this difference is statistically significant.

CONCLUSIONS

This paper motivates and describes a generalized analysis of covariance (ANCOVA) model for characterizing emissions profiles among populations of vehicles operating on different fuel types. The approach is illustrated on a data set comprised of 27 CNG and 31 RFG Dodge Ram vans operating in the U.S. federal fleet. The analysis and discussion emphasizes that a proper analysis of emissions must consider: 1) the emissions deterioration that occurs over the lifetime of a vehicle; 2) the emissions variability that is prevalent for individual vehicles; and 3) the emissions variability from one vehicle to another. Conventional regression analyses fail to properly account for 2 and 3. The ANCOVA model used in this study explicitly accounts for all of these factors and can be readily applied to more precisely characterize the emissions of any alternative or conventional fuel vehicles.

Moreover, by properly accounting for variation between vehicles, one can develop a more realistic understanding of the *range* of emissions values that are possible from any randomly chosen vehicle in the population. This range may, in fact, be considerably different from what would be obtained from more classical regression models that fail to account for variations between individual vehicles. This type of understanding can be critical to policymakers and researchers.

The confidence bands displayed in figure 7 are based on the model in equation (7) that accounts for variation among vehicles in the same population. While common sense suggests that such variation does exist, its impact on analyses aimed at characterizing emissions profiles has not generally been appreciated. Whenever the vehicle-to-vehicle variation is large (compared with the error variation), then any analysis that fails to account for variation between vehicles can lead to confidence bands around the emissions profile that are misleading (and may even be seriously misleading). In such a case, comparisons of emissions profiles from different populations or different fuel types are suspect.

APPENDIX A: FORMULAS USED FOR GENERATING FIGURE 1

This section outlines the statistical theory behind the confidence bandwidths displayed in figure 1. It is assumed that the reader is familiar with probability theory and the theory of general linear statistical models as described in Graybill (1976).

Recall that the context for interpreting figure 1 is as follows. Data is collected on some emissions constituent (e.g., NMHC) from a single vehicle after 10,000, 20,000,….,100,000 miles of use. Least squares analysis is then used to fit the model given in equation (1) and to calculate traditional confidence bands for the average emissions after 50,000 miles (using equation (4)). Now suppose that there is some unknown vehicle-to-vehicle variation among the cars in the population of interest. In particular, the intercept of equation (1) varies randomly from one vehicle to the next, so that the correct model for these data is actually equation (6). The question to be answered is this: how misleading is the confidence interval calculated from equation (4)? Figure 1 attempts to provide one way of answering that question.

Note that figure 1 displays the *95% confidence bandwidth* for the traditional confidence interval (from equation (4)), along with the corresponding bandwidth that would be necessary to achieve 95% confidence (assuming that the model in equation (1) is correct). Given the relative size of the vehicle-to-vehicle variation (σ_v) with respect to the error variation (σ), the expected width of the traditional confidence interval can be compared with the width that would be necessary to achieve true 95% confidence (i.e., in order that the probability that the interval covers the true average emissions is truly equal to 95%). The x-axis specifies the ratio $\sigma_v \div \sigma$ and the y-axis displays the expected size of the ± bounds of the traditional interval and the theoretically correct interval. Figure 1 clearly illustrates that as $\sigma_v \div \sigma$ increases, the disparity between the confidence intervals increases.

In order to demonstrate how the bandwidths in figure 1 are calculated, a matrix representation of the general regression model will be used (Graybill 1976). Suppose one new vehicle is randomly selected from the population of interest. This vehicle will be operated for a fixed number of miles (e.g., 100,000 miles), and one or more emissions constituents (e.g., NMHC) will be measured at fixed mileages along the way. Suppose n emissions values are obtained from the vehicle during the life of the study. Further suppose that the relationship between emissions and mileage for each car is correctly represented by equation (6); that is,

$$Y_{ij} = \alpha + \nu_i + \beta m_{ij} + \varepsilon_{ij}, \tag{9}$$

where $i = 1$, and $j = 1,\ldots, n$. Assume that the error terms (ε_{ij}) are independent and identically distributed according to a $N(0,\sigma^2)$ distribution, and that the ν_i terms are independent and identically distributed according to a $N(0,\sigma_\nu^2)$ distribution.

Now suppose that vehicle-to-vehicle variation (as represented by ν_i in (9)) is mistakenly assumed to be absent, and traditional regression methods are used to fit the model in (1), i.e.,

$$Y_{ij} = \alpha + \beta m_{ij} + \varepsilon_{ij}, \tag{10}$$

Using the traditional ordinary least squares estimate of the model in (10), the goal is to calculate the average bandwidth of the 95% confidence interval for the model in (10) (which is based on the assumption of no vehicle-to-vehicle variation) and compare its bandwidth with the correct bandwidth that would be required in order to assure 95% confidence (when vehicle-to-vehicle variation is correctly incorporated).

Following Graybill (1976), matrix notation can be used to represent the model in (10). Define the following matrices.

$$X_{n\times 2}=\begin{bmatrix} 1 & m_{1,1} \\ 1 & m_{1,2} \\ \cdots & \cdots \\ 1 & m_{1,n} \end{bmatrix} \quad {}_{n\times 1}=\begin{bmatrix} Y_{1,1} \\ Y_{1,2} \\ \cdots \\ Y_{1,n} \end{bmatrix} \quad B_{2\times 1}=\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad E_{n\times 1}=\begin{bmatrix} \varepsilon_{1,1} \\ \varepsilon_{1,2} \\ \cdots \\ \varepsilon_{1,n} \end{bmatrix}$$

$$\Sigma_{n\times n} = var\,(Y_{n\times 1}|X).$$

The model in (10) can then be written in matrix notation as follows:

$$= XB + E.$$

Using ordinary least squares, the estimate of the regression coefficients, $\hat{B}$, is given by

$$\hat{B} = (X'X)^{-1}\,X'Y,$$

and the estimated population-wide average emissions at mileage m is given by

$$M'\hat{B} = M'(X'X)^{-1}\,X'Y,$$

where $M' = (1, m)$. It is easily shown (Graybill 1976) that this estimate is normally distributed with a mean of $M'B$ (i.e., the estimate is unbiased) and a standard deviation equal to

$$\sqrt{M'(X'X)^{-1}\,X'\Sigma X(X'X)^{-1}\,M}.$$

Hence, assuming that the covariance matrix Σ is known, the theoretically correct 95% confidence interval for the estimated emissions at mileage m is given by

$$M'\hat{B} \pm 1.96 \cdot \sqrt{M'(X'X)^{-1}\,X'\Sigma X(X'X)^{-1}\,M} \tag{11}$$

Whenever there is no vehicle-to-vehicle variation, then $\Sigma = \sigma^2 I$, where I is the identity matrix, and expression (11) simplifies to

$$M'\hat{B} \pm 1.96\sigma \cdot \sqrt{M'(X'X)^{-1}\,X'IX(X'X)^{-1}\,M} =$$

$$M'\hat{B} \pm 1.96 \cdot \sqrt{M'(X'X)^{-1}\,M}. \tag{12}$$

This last expression is the matrix representation of the confidence band in equation (4).

On the other hand, if vehicle-to-vehicle variation is present, then $\Sigma = \sigma^2 I + \sigma^2_{\nu} J$ where J is a matrix of all 1s. Under these conditions, expression (11) does not simplify to the form in (12). Hence, if it is mistakenly assumed that there is no vehicle-to-vehicle variation and expression (12) (or, equivalently, expression (4)) is used to calculate confidence intervals, the resulting confidence bands will be based on incorrect error terms, and the confidence interval will be less than 95%. The correct 95% bounds are instead given by (11).

The error term in expression (12) (applied to the hypothetical example discussed in section 2) corresponds to the traditional confidence bandwidth displayed in figure 1. The error term in expression (11) corresponds to the correct 95% bandwidth displayed in figure 1.

APPENDIX B: SAS CODE FOR FITTING THE ANCOVA MODEL AND OBTAINING 95% CONFIDENCE BANDS FOR THE POPULATION AVERAGE EMISSIONS DISPLAYED IN FIGURE 6

```
/*
SAS code to get "best" variance component estimates and predicted emissions separately within each fuel type. These predictions and standard errors correctly account for the covariance structure imposed by the random effects.

A separate call to PROC MIXED is required for each response.

Variables are:
      VID =  vehicle ID code (unique for each
             vehicle)
     FUEL =  type of fuel used by the vehicle
```

```
              (model assumes only one fuel type
              is used on each vehicle)
    ODOM =    odometer reading
    NMHC =    nonmethane hydrocarbon reading
              on the vehicle at the specified
              mileage
*/
PROC MIXED DATA = SASUSER.FINALRAM
METHOD=ML;
CLASSES VID FUEL_TYP;
/*
The MODEL statement specifies only the "fixed
terms" in the model (i.e., the fuel type and odome-
ter reading). The FUEL*ODOM crossproduct
term instructs SAS to fit a separate slope for each
FUEL type.
*/
MODEL NMHC = FUEL ODOM ODOM*FUEL
/ SOLUTION DDFM=SATTERTH;
/*
The RANDOM statement identifies those terms in
the model that are random. Any terms identified in
the RANDOM statement are automatically includ-
ed in the model and are therefore not explicitly
named in the MODEL statement.
*/
RANDOM VID(FUEL) ODOM*VID(FUEL);
/*
The LSMEANS statements instruct SAS to calcu-
late the predicted mean emissions for each fuel type
at the specified mileage reading. This corresponds
to the quantity given in equation (8) of the paper.
The LSMEANS statement also provides the stan-
dard error that can be used to calculate the 95%
confidence interval for the mean emissions at the
specified odometer reading.
*/
LSMEANS FUEL/AT ODOM = 5000 PDIFF;
LSMEANS FUEL/AT ODOM = 10000 PDIFF;
LSMEANS FUEL/AT ODOM = 15000 PDIFF;
LSMEANS FUEL/AT ODOM = 20000 PDIFF;
LSMEANS FUEL/AT ODOM = 25000 PDIFF;
LSMEANS FUEL/AT ODOM = 30000 PDIFF;
LSMEANS FUEL/AT ODOM = 35000 PDIFF;
LSMEANS FUEL/AT ODOM = 40000 PDIFF;
LSMEANS FUEL/AT ODOM = 45000 PDIFF;
LSMEANS FUEL/AT ODOM = 50000 PDIFF;
LSMEANS FUEL/AT ODOM = 55000 PDIFF;
LSMEANS FUEL/AT ODOM = 60000 PDIFF;
RUN;
```

REFERENCES

Battelle Memorial Institute. 1995. *Clean Fleet: Vehicle Emissions, Statistical Analysis Report No. 6*. Columbus, OH.

Chandler, K.L., P. Norton, and N. Clark. 1999. *Interim Results from Alternative Fuel Truck Evaluation Project,* SAE Report 1999–01–1505. Warrendale, PA: SAE International.

Clark, N.N., B.L. Rapp, M. Gautam, W. Wang, and D. Lyons. 1998. *A Long Term Field Emissions Study of Natural Gas Fueled Refuse Haulers in New York City,* SAE Report 982456. Warrendale, PA: SAE International.

Durbin, T., J.R. Collins, J.M. Norbeck, and M.R. Smith. 1999. *Evaluation of the Effects of Alternative Diesel Fuel Formulations on Exhaust Emission Rates and Reactivity,* Report #99-VE-RT2P-001-FR, South Coast Air Quality Management District Contract #98102. Sacramento, CA: College of Engineering–Center for Environmental Research and Technology, University of California–Riverside.

Federer, W.T. and M.P. Meredith. 1992. *The American Statistician* 46:155–62.

Gabele, P.A. 1990. Characterization of Emissions from a Variable Gasoline/Methanol Fueled Car. *Journal of the Air and Waste Management Association* 40:296–304.

Gabele, P.A. 1995. Exhaust Emissions from In-Use Alternative Vehicles. *Journal of the Air and Waste Management Association* 45:770–77.

Graybill, F.A. 1996. *Theory and Application of the Linear Model.* North Scituate, MA: Duxbury Press.

Hormes, L., U.S. Environmental Protection Agency, Ann Arbor, MI. 2000. Personal communication.

Kelly, K.J., B.K. Bailey, T.C. Coburn, W. Clark, L. Eudy, and P. Lissiuk. 1996a. *FTP Emissions Test Results from Flexible-Fuel Methanol Dodge Spirits and Ford Econoline Vans,* SAE Technical Paper 961090. Warrendale, PA: SAE International.

_____. 1996b. *Round 1 Emissions Test Results from Compressed Natural Gas Vans and Gasoline Controls Operating in the U.S. Federal Fleet,* SAE Technical Paper 961091. Warrendale, PA: SAE International.

_____. 1996c. *Federal Test Procedure Emissions Test Results from Ethanol Variable-Fuel Vehicle Chevrolet Luminas,* SAE Technical Paper 961092. Warrendale, PA: SAE International.

Kirchstetter, T.W., B.C. Singer, R.A. Harvey, G.R. Kendall, and W. Chan. 1996. Impact of Oxygenated Gasoline Use in California Light Duty Vehicle Emissions. *Environmental Science and Technology* 30:661–70.

McCormick, R.L., M.S. Graboski, T. Allemean, A.M. Herring, and P. Nelson. 1999. *In-Use Emissions from Natural Gas Fueled Heavy-Duty Vehicles*, SAE Report 1999–01–1507. Warrendale, PA: SAE International.

Norbeck, J.M., T.J. Truex, M.R. Smith, and T. Durbin. 1998. *Inventory of AFVs and AFV Comparison: OEM vs. Retrofits,* Report #98-VE-RT2W-008-FR, South Coast Air Quality Management District Contract #97029. Sacramento, CA: College of Engineering–Center for Environmental Research and Technology, University of California–Riverside.

Searle, S.R. 1971. *Linear Models*. New York, NY: John Wiley & Sons Publishing.

Whalen, P., K. Kelly, and M. John. 1999. *Barwood CNG Cab Fleet Study: Final Results,* NREL/TP-540–26035. Golden, CO: National Renewable Energy Laboratory.

Validation of the MEASURE Automobile Emissions Model: A Statistical Analysis

IGNATIUS FOMUNUNG

Clark Atlanta University

SIMON WASHINGTON

RANDALL GUENSLER

WILLIAM BACHMAN

Georgia Institute of Technology

ABSTRACT

This paper details the results of an external validation effort for the hot stabilized option currently included in the Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE). The MEASURE model is one of several new modal emissions models designed to improve predictions of CO, HC, and NO_x for the on-road vehicle fleet. Mathematical algorithms within MEASURE predict hot stabilized emission rates for various motor vehicle technology groups as a function of the conditions under which the vehicles are operating, specifically various aggregate measures of their speed and acceleration profiles. Validation of these algorithms is performed on an independent data set using three statistical criteria. Statistical comparisons of the predictive performance of the MEASURE and MOBILE5a models indicate that the MEASURE algorithms provide significant improvements in both average emission estimates and explanatory power over MOBILE5a for all three pollutants across almost every operating cycle tested. In addition, the MEASURE model appears to be less biased, the most critical model performance measure for point-estimate forecasts, than MOBILE5a.

Ignatius W. Fomunung, Department of Engineering, Clark Atlanta University, Atlanta, GA 30314. E-mail: ifomunun@cau.edu.

INTRODUCTION

Emission rate model uncertainties in currently employed regional emissions models arise in part because emission rates rely primarily on average speed as the dominant, continuous, independent variable in the regression analysis. However, many factors, both continuous and discrete, in addition to average speed, affect the net load demanded of an engine, which in turn affects a vehicle's resultant emissions. These factors include roadway grade, rolling resistance, aerodynamic drag, engine speed, engine friction losses, transmission losses, vehicle mass, power consumption of accessories, and so forth. Numerous references identify these factors as influential in the formulation of various pollutants; however, they are largely omitted in currently employed emission prediction algorithms (Guensler 1993).

Cicero-Fernandez et al. (1997a; 1997b) demonstrated that emissions from an individual vehicle may increase by a factor of two when driven on an uphill grade, yet current inventory models do not account for grade. In addition, real-world driving conditions, in terms of speed/acceleration distributions and/or traces, are not well represented in the current models. The Federal Test Procedure (FTP), appropriately used to develop baseline emissions factors, does not capture the extremes of emission-producing activities associated with aggressive driving. Jimenez-Palacio (1999), using a new definition of specific power, calculated the maximum specific power of the FTP to be approximately 22 kilowatts per metric ton. More telling, the research indicates that the onset of commanded enrichment for many vehicles occurs at this maximum. Commanded enrichment is responsible for elevated or "super" emissions, which can be one to several orders of magnitude higher than emissions obtained under stoichiometric engine operation. As a result, a large proportion of commanded enrichment is not likely to appear under the FTP.

Driver behavior may also be an important source of uncertainty and variability in motor vehicle emissions (Bishop et al. 1996). A study of repeated measurements on the IM240 driving cycle indicates that driver behavior may be responsible for potentially order-of-magnitude differences in emissions for clean low-emitting vehicles (Webster and Shih 1996). Despite this recognition, few advances have been made in quantifying the effect of driving behavior on emissions, except for Shih et al. (1997), who used throttle position distributions to represent driver behavior, albeit with mixed results. Their research provides evidence that throttle position distributions might be used to reflect differences in driving behavior, but such models still need refinement. The forecasting of throttle position distributions, which interact with specific driver types, facility types, and trip purposes, may prove too difficult.

Emerging Models

Efforts at improving motor vehicle emissions have occupied researchers for quite some time. Cadle et al. (1997) recently summarized advances in real-world motor vehicle emissions modeling. The U.S. Environmental Protection Agency (EPA) is currently revising the MOBILE5a emissions rate model. MOBILE6 promises to provide significant improvements in terms of representing modal impacts on emissions rates because supplemental driving cycles that mimic on-road conditions under various levels of congestion are being used to develop cycle-based speed correction factors. New certification testing cycles also promise to reduce the frequency of on-road enrichment. The USO6 cycle represents emissions in aggressive driving, and the SCO3 cycle reflects the effects of accessory loads like air conditioning usage, power steering, and so forth.

Modal modeling approaches are also currently under development. A modal emissions model being developed at the University of California, Riverside by An et al. (1997) is based on 300 vehicles tested under a variety of laboratory driving cycles. Two modal approaches developed at the Georgia Institute of Technology are included in the GIS-based modal emission model: an aggregate modal model based on statistical analysis of historic laboratory data (Guensler et al. 1998) and a load-based prediction module based on analysis of instrumented vehicle data (Rodgers 1995).

For the past six years, the Georgia Tech Research Partnership has been developing a research-grade motor vehicle emissions model within a geographic information system (GIS) framework. Once validation and peer review efforts are complete, MEASURE may serve as an alternative or supplement to the current MOBILE5a model. The aggregate modal model within MEASURE predicts emissions from light-duty vehicles. The Georgia Tech aggregate modal model predicts emissions as a function of vehicle operating mode, representing a spectrum of vehicle operating conditions including cruise, acceleration, deceleration, idle, and the power demand conditions that lead to enrichment, that is, high fuel to air ratios. The model accounts for interactions between specific vehicle fleet characteristics and vehicle operating modes. For each technology group within a light-duty motor vehicle fleet, the relationships between modal activity and emissions can differ significantly. The framework allows for facility-level aggregations of microscopic traffic simulation or disaggregation of traditional macroscopic four-step travel demand forecasting models to develop emission-specific vehicle activity data.

The aggregate modal model within MEASURE employs emission rates based on theoretical engine-emissions relationships that have been modeled using various statistical techniques (Fomunung et al. 1999). The emissions rate models have been estimated through a process that utilizes the best aspects of hierarchical tree-based regression (HTBR) (Breiman et al. 1984) and ordinary least squares (OLS) regression. The relationships are dependent on both modal and vehicle technology variables, and they are "aggregate" in the sense that they rely on bag data to derive their modal activities (Washington 1994). Thus, they are suitable for existing aggregate approaches contained within the travel demand modeling (TDM) framework.

Although much effort has been conducted and reported in the literature on the emission algorithms within MEASURE, little has been done toward the external validation of the MEASURE emissions predictions components or to compare the performance of MEASURE with that of MOBILE5a. Model validation, the use of a sample of external data to assess model predictive abilities, is perhaps the single most important measure of a model's ability to capture relationships across space and time. In addition, it is the only way to compare two competing models fairly. This paper details the results of an external validation effort for the hot stabilized exhaust option currently included in MEASURE. The performances of MEASURE and MOBILE5a are compared using mean absolute prediction errors, linear correlation coefficients between observed and predicted emissions, and mean prediction errors. Results are provided for each driving cycle and for vehicle technology classes.

MEASURE AGGREGATE MODAL MODELS

In the context of this paper, the term "model" refers to a mathematical algorithm or expression that relates emissions measurements to various explanatory variables. The model estimation data consisted of more than 13,000 laboratory tests conducted by EPA and the California Air Resources Board (CARB) using standardized test cycle conditions, as well as alternative driving cycles (Fomunung et al. 1999). The aggregate modal model algorithms presented below were estimated using the logarithm of the emission rate ratio for each pollutant as a response variable (Fomunung et al. 1999). The ratio is the emission rate (in grams per second) (g/sec) for a vehicle driven on a given cycle (or equivalently across a specified speed/acceleration matrix), divided by that vehicle's emissions rate (g/sec) obtained from the FTP bag 2 testing cycle. MEASURE's Aggregate Modal Model predicts the ratio of g/sec emission rates for several vehicle technology groups. The following sequence of equations shows the method for calculating the predicted emissions rates for each pollutant in units of either g/sec (ψ) or g/mile ($\tilde{\psi}$):

$$\psi_i = \tilde{\psi}_i \times \text{DIST} / \text{DUR} \quad (1),$$

$$\psi_{i\text{bag2}} = \tilde{\psi}_{i\text{bag2}} \times 3.91/866 \quad (2),$$

and

$$R_i = P_i \; / \; \psi_{i\text{bag2}} \quad (3).$$

In these equations, ψ_i and $\tilde{\psi}_i$ are the observed or measured pollutant (i is the index for CO, HC, or NO_x); P_i is the predicted value of pollutant i; $\psi_{i\text{bag2}}$ and $\tilde{\psi}_{i\text{bag2}}$ are the observed FTP bag 2 rates for pollutant i in a given vehicle; DIST is the driving cycle distance in miles; DUR is the cycle duration in seconds; 3.91 is the hot stabilized FTP bag 2 subcycle distance in miles; and 866 is the FTP bag 2 subcycle duration in seconds.

The emissions models in MEASURE are presented in two formats: in the form in which they were estimated (suitable for making statistical inferences) and in original variable units, often more intuitive for use in emissions rate prediction and for interpretation of results. The statistical details of the models are provided in tables 1, 2, and 3 for CO, HC, and NO_x, respectively. Details of the model development process including goodness-of-fit, analysis of residuals, and interpretation of coefficients are published elsewhere (Fomunung et al. 1999; Fomunung 2000).

Model Estimation Forms

Equation (4) shows the estimation form for CO. Equation (5) shows the estimation form for HC, and equation (6) shows the estimation form of NO_x.

For CO,

$$\text{Log}R_{CO} = 0.0809 + 0.002 \times \text{AVGSPD} + 0.0461 \times \text{ACC.3} + 0.0165 \times \text{IPS.60} - 0.0283 \times \text{ips45sar2} + 0.3778 \times \text{ips90tran1} - 0.0055 \times \text{tran3idle} + 0.1345 \times \text{tran5km1} + 0.3966 \times \text{finj3sar3} - 0.0887 \times \text{cat3tran1} - 0.2636 \times \text{sar3tran4} - 0.481 \times \text{flagco} \quad (4)$$

where

AVGSPD is the average speed of the driving cycle in mph,

ACC.3 is the proportion of the driving cycle on acceleration greater than three mph per second,

IPS.X is the proportion of the driving cycle on inertial power surrogate (IPS) (speed times acceleration) greater than X mph^2/sec (Washington 1994) (thus, IPS.60 implies IPS greater than 60 mph^2/sec),

ips45sar2 is an interaction between IPS.45 and a vehicle with no air injection,

ips90tran1 is an interaction variable for a vehicle with automatic transmission on IPS.90,

cat3idle is an interaction variable for a three-speed manual transmission at idle,

tran5km1 is an interaction variable for a five-speed manual transmission vehicle with mileage ≤ 25,000 miles,

finj3sar3 is an interaction variable for a vehicle that has throttle body fuel injection and pump air injection,

cat3tran1 is an interaction variable for a vehicle with automatic transmission and three-way catalyst (TWC),

TABLE 1 CO Model Details

Variable	Estimated coefficient	Standard error	*t*-value	Pr(>|*t*|)
Intercept	0.0809	0.0154	5.2382	<0.001
AVGSPD	0.0020	0.0004	5.0514	<0.001
ACC.3	0.0461	0.0026	17.8998	<0.001
IPS.60	0.0165	0.0066	2.4909	<0.013
ips45sar2	–0.0283	0.0067	–4.2136	<0.001
ips90tran1	0.3778	0.0265	14.2899	<0.001
cat3idle	–0.0055	0.0004	–13.8299	<0.001
tran5km1	0.1345	0.0134	10.0067	<0.001
finj3sar3	0.3966	0.0314	12.6305	<0.001
cat3tran1	–0.0887	0.0145	–6.1218	<0.001
sar3tran4	–0.2636	0.1177	–2.2401	<0.025
flagco	–0.4810	0.0290	–16.5777	<0.001

Residual standard error: 0.9177 on 12,965 degrees of freedom
R^2 (adjusted): 0.1726*
F-statistic: 245.9 on 11 and 12,965 degrees of freedom, the p-value is 0
*The low R^2 for the CO model is an indication that the model doesn't fit very well. It is low relative to the values for the HC and NO_x models because the production mechanism for CO emissions in the engine and exhaust manifold is more complex than for HC and NO_x emissions. The EPA testing protocol that generated the current database did not include important variables such as catalyst efficiency under varying load conditions and various transient oxygen effects, which research has shown account for much of the variability in CO emissions. It is expected that a CO model estimated using a data set with these additional variables would result in a much improved R^2.

TABLE 2 HC Model Details

Variable	Estimated coefficient	Standard error	*t*-value	Pr(>|*t*|)
intercept	0.1685	0.0098	17.1164	<0.001
my79	0.3601	0.0098	36.5975	<0.001
finj2tran4	–0.0732	0.0196	–3.7260	<0.001
cat2sar1	0.3324	0.0206	16.1707	<0.001
cat3sar1	–0.4201	0.0247	–17.004	<0.001
cat3sar2	–0.1188	0.0123	–9.6257	<0.001
sar3tran1	–0.3602	0.0194	–18.5248	<0.001
cyl8	–0.2349	0.0115	–20.4826	<0.001
sar3km1	–0.2175	0.0152	–14.3368	<0.001
finj2km3	–0.0290	0.0034	–8.4548	<0.001
acc1finj2	–0.0550	0.0030	–18.3900	<0.001
acc3cat2	–1.3528	0.0234	–57.7883	<0.001
ips90sar3	–0.9201	0.0566	–16.2530	<0.001
dps8finj2	0.0391	0.0007	54.0156	<0.001

Residual standard error: 9.414 on 12,350 degrees of freedom
R^2 (adjusted): 0.6094
F-statistic: 1,482 on 13 and 12,350 degrees of freedom, the *p*-value is 0

TABLE 3 NO_x Model Details

Variable	Estimated coefficient	Standard error	*t*-value	Pr(>|*t*|)
(intercept)	–0.5864	0.0068	–85.9273	<0.001
AVGSPD	0.0225	0.0002	131.6271	<0.001
IPS.120	0.3424	0.0452	7.5684	<0.001
ACC.6	0.6329	0.1683	3.7595	<0.002
DEC.2	0.0247	0.0007	34.8026	<0.001
finj2km1	0.0083	0.0008	10.4205	<0.001
finj2km2	0.0028	0.0004	6.8670	<0.001
cat2km3	–0.0021	0.0004	–5.9243	<0.001
cat3km2	0.0026	0.0002	13.5707	<0.001
cat3km3	0.0003	0.0001	2.9355	<0.001
finj1km3flagnox	–0.0085	0.0015	–5.7854	<0.001
finj3km3flagnox	–0.0068	0.0009	–7.4491	<0.001

Residual standard error: 0.3479 on 12,962 degrees of freedom
R^2 (adjusted): 0.623
F-statistic: 1,947 on 13 and 12,962 degrees of freedom, the *p*-value is 0 under varying load conditions and various transient oxygen effects, which research has shown account for much of the variability in CO emissions. It is expected that a CO model estimated using a data set with these additional variables would result in a much improved R^2.

sar3tran4 is an interaction variable for a vehicle with four-speed manual transmission and pump air injection, and

flagco is a flag used to tag a vehicle emitting high CO emissions (Wolf et al 1998).

For HC,

$$\text{Log}R_{HC} = 0.1685 + 0.3601(\text{my79}) - 0.0732\,(\text{finj2tran4}) + 0.3324\,(\text{cat2sar1}) - 0.4201\,(\text{cat3sar1}) - 0.1188\,(\text{cat3sar2}) - 0.3602\,(\text{sar3tran1}) - 0.2349\,(\text{cyl8}) - 0.2175\,(\text{sar3km1}) - 0.0290\,(\text{finj2km3}) - 0.055\,(\text{ACC.1finj2}) - 1.3528\,(\text{ACC.3cat2}) - 0.9201\,(\text{IPS.90sar3}) + 0.0391\,(\text{DPS.800finj2}) \quad (5)$$

where

my79 is model year < 79

finj2tran4 is an interaction variable for a four-speed manual transmission vehicle with a carburetor,

cat2sar1 is a variable for a pre-1981 model year vehicle with an "oxidation only" catalyst and of unknown air injection type,

cat3sar1 is a variable for a pre-1981 model year vehicle with a TWC and of unknown air injection type,

cat3sar2 is a variable for a vehicle with TWC and no air injection,

sar3tran1 is an automatic transmission vehicle with pump air injection,

cyl8 is a vehicle with an eight-cylinder engine,

sar3km1 is a vehicle with pump air injection and mileage ≤ 25,000 miles,

finj2km3 is a vehicle with pump air injection and 50,000 < mileage ≤ 100,000 miles,

acc1finj2 is a carburetor-equipped vehicle operating with acceleration greater than one mph per second,

acc3cat2 is an "oxidation only" catalyst vehicle on ACC.3,

ips90sar3 is a vehicle with air pump on IPS.90, and

dps800finj2 is the proportion of drag power surrogate (DPS) (speed times speed times acceleration) greater than 800 mph^3/sec.

For NO_x,

$$\text{Log}R_{NO_x} = -0.5864 + 0.0225 \times \text{AVGSPD} + 0.3424 \times \text{IPS.120} + 0.6329 \times \text{ACC.6} + 0.0247 \times \text{DEC.2} + 0.0083 \times \text{finj2km1} + 0.0028 \times \text{finj2km2} - 0.0021 \times \text{cat2km3} + 0.0026 \times \text{cat3km2} + 0.0003 \times \text{cat3km3} - 0.0085 \times \text{finj1km3flagnox} - 0.0068 \times \text{finj3km3flagnox} \quad (6)$$

where

IPS.120 is IPS greater than 120 mph^2/sec,

ACC.6 is the proportion of acceleration greater than six mph per second,

DEC.2 is the proportion of deceleration greater than two mph per second,

finj2km1 is a carburetor-equipped vehicle with mileage less than 25,000 miles,

finj2km2 is a carburetor-equipped vehicle with 25,000 < mileage ≤ 50,000 miles,

cat2km3 is an "oxidation only" catalyst vehicle with 50,000 < mileage ≤ 100,000miles,

cat3km2 is a TWC vehicle with 25,000 < mileage ≤ 50,000 miles,

cat3km3 is a TWC vehicle with 50,000 < mileage ≤ 100,000 miles,

finj1km3flagnox is a second-order interaction variable for a high-emitting vehicle with port fuel injection and 50,000 < mileage ≤ 100,000 miles, and

finj3km3flagnox is a second-order interaction variable for a high-emitting vehicle with throttle body fuel injection and 50,000 < mileage ≤ 100,000 miles.

This implies that on a vehicle-by-vehicle basis after calculating R_i from the response variable, the predicted rate P_i in g/sec is

$$P_i = R_i \times \psi_{i\text{bag2}} \quad (7).$$

Note that equation (7) is similar in form to the embedded algorithm in MOBILE5a, which gives emission rates as Correction Factors times Base Emission Rate (BER). BER is similar to $\psi_{i\text{bag2}}$; R_i represents all the variables which figure into the models for each pollutant and can be thought of as speed, load, and technology correction factors. The conversion to g/mile is straightforward and given by

$$\tilde{P}_i = R_i \times \psi_{i\text{bag2}} \times 1/\text{AVGSPD} \quad (8).$$

Model Prediction Forms

The prediction forms for CO, HC, and NO_x are shown in equations (9), (10), and (11), respectively, and the variables are as previously described.

The prediction equations are no more than the antilogs of the estimation equations.

For CO, in g/sec,

$$P_{CO} = 1.205 \times \text{FTP bag2} \times \text{antilog}\,[0.002 \times \text{AVGSPD} + 0.0461 \times \text{ACC.3} + 0.0165 \times \text{IPS.60} - 0.0283 \times \text{ips45sar2} + 0.3778 \times \text{ips90tr1} - 0.0055 \times \text{tran3idle} + 0.1345 \times \text{tran51} + 0.3966 \times \text{finj3sar3} - 0.0887 \times \text{cat3tran1} - 0.2636 \times \text{sar3tran4} - 0.481 \times \text{flagco}] \quad (9).$$

For HC,

$$P_{HC} = 1.474 \times \text{FTP bag2hc} \times \text{antilog}[0.3601(\text{myhc81}) - 0.0732\ (\text{finj2tran4}) + 0.3324\ (\text{cat2sar1}) - 0.4201\ (\text{cat3sar1}) - 0.1188\ (\text{cat3sar2}) - 0.3602\ (\text{sar3tran1}) - 0.2349\ (\text{cyl8}) - 0.2175\ (\text{sar3km1}) - 0.0290\ (\text{finj2km3}) - 0.055\ (\text{ACC.1finj2}) - 1.3528\ (\text{ACC.3cat2}) - 0.9201\ (\text{IPS.90sar3}) + 0.0391\ (\text{DPS.800finj2})] \quad (10).$$

For NO_x,

$$P_{NOx} = 0.259 \times \text{FTP bag2} \times \text{antilog}\,[0.0225(\text{AVGSPD}) + 0.3424(\text{IPS.120}) + 0.6329(\text{ACC.6}) + 0.0247(\text{DEC.2}) + 0.0083(\text{finj2km1}) + 0.0028(\text{finj2km2}) - 0.002(\text{cat2km3}) + 0.0026(\text{cat3km2}) + 0.0003(\text{cat3km3}) - 0.0085(\text{finj1km3flagnox}) - 0.0068(\text{finj3km3flagnox})] \quad (11).$$

The algorithms shown in equations (4) to (6) indicate that, apart from AVGSPD, which appears in both the CO and NO_x models, a different collection of variables is needed to model each pollutant. This finding is in agreement with theoretical expectations. The production and distribution of all three pollutants are functions of the physico-chemical processes occurring in the engine. While CO and NO_x are principally produced as a result of chemical and kinetic mechanisms within the engine, the production of HC is heavily dependent on the physical processes within the engine. The phrase "physical processes" is used in an inclusive sense to embody both the physical structure of the engine combustion chamber and the physics of the combustion process within the combustion chamber. It has long been recognized that the crevices within the combustion chamber are a significant source of exhaust hydrocarbons (Heywood 1988). Therefore, it is not surprising that different variables are needed to model each pollutant. For example, the variable cyl8, which is positively correlated with the number of crevices in the engine, is a significant predictor variable in the HC model but is insignificant in both the CO and NO_x models.

VALIDATION DATA SET DESCRIPTION

Model validation consists of two types, internal and external. Internal validation consists of model-checking for plausibility of signs and magnitudes of estimated coefficients, agreement with past models and theory, and model diagnostic checks such as distribution of error terms, normality of error terms, and so forth. Internal validation was performed as part of the model estimation procedure and is documented in Fomunung et al. (1999) and Fomunung (2000). External validation is the process whereby a model is compared to data collected "outside" the modeling framework (i.e., data from another location or time). External validation is the only way to check if a model has been "overfit" to data, thus capturing spurious rather than real relationships or underlying structure in the data. It is also the only way to determine whether the relationships captured in the estimated model reflect the same relationships elsewhere or over time. Finally, external validation is the only objective way to compare two models estimated using different data. These objectives have motivated the validation of the MEASURE emission prediction algorithms: to assess its transferability and to compare its performance to the current in-practice emission predictions model, MOBILE5a.

The data used for MEASURE and MOBILE5a validation consist of 50 vehicles tested across 16 different hot stabilized driving cycles. Of the 50 vehicles, 4 are Chrysler-manufactured cars, 13 are Ford cars, 21 are GM cars, and the rest are imports. One of the four Chrysler cars is a 1983-model year car with 94,399 miles. Another is a 1989-model year car with 118,586 miles, and two are 1995-model year cars with 20,855 miles and 28,525 miles, respectively. The Ford cars are from model years 1985 to 1992 with between 53,000 and 123,000 miles. The GM cars are from model years 1985 to 1996 with 16,000 to 180,000 miles.

TABLE 4 Number of Tests, Average Speeds, Maximum Speeds, and Maximum Instantaneous Acceleration of Each Test Cycle for the Validation Data Set

Test cycle description	Name	Number of tests	Average speed (mph)	Maximum speed (mph)	Maximum acceleration (mph/sec)
Arterial LOS A-B cycle	ARTA	50	24.8	58.9	5.0
Arterial LOS C-D cycle	ARTC	50	19.2	49.5	5.7
Arterial LOS E-F cycle	ARTE	50	11.6	39.9	5.8
Hot running 505	F505	50	25.6	56.3	3.4
New York City cycle	FNYC	50	7.1	27.7	6.0
Freeway LOS A-C cycle	FWAC	50	59.7	73.1	3.4
High-speed freeway cycle	FWHS	50	63.2	74.7	2.7
Freeway LOS D cycle	FWYD	50	52.9	70.6	2.3
Freeway LOS E cycle	FWYE	50	30.5	63.0	5.3
Freeway LOS F cycle	FWYF	50	18.6	49.9	6.9
Freeway LOS G cycle	FWYG	50	13.1	35.7	3.8
CARB "unified" LA92 cycle	LA92	49	24.7	67.2	6.9
Local roadways cycle	LOCL	50	12.8	38.3	3.7
Freeway ramp cycle	RAMP	50	34.6	60.2	5.7
Start cycle	ST01	49	20.1	41.0	5.1
Areawide non-freeway cycle	WIDE	49	19.4	52.3	6.4

The model years of the imports are from 1987 to 1993 with 30,000 to 197,000 miles.

Neither the MEASURE nor MOBILE5a models were originally estimated using data from these vehicles. Table 4 lists the different cycles used and shows their average speed s, maximum speeds, and acceleration characteristics. EPA tested each vehicle on every cycle (three cycles only included 49 of the 50 vehicles), and the near-balanced sampling design results in the ability to segregate vehicle-to-vehicle, within vehicle, and cycle-to-cycle variability. A similar list of cycles used in the MEASURE model development is shown in table 5. There are minor differences between the two data sets. First, only two driving cycles, NYCC and the CARB "unified" cycle, were used in both instances. Second, the data ranges for the parameters of interest are slightly different: average speeds range from 2.45 to 59.9 mph in the model data and from 7.1 to 63.2 mph in the validation data; maximum speeds range from 10 to 71.3 mph in the model data and from 27.7 to 74.7 mph in the validation data; and maximum acceleration ranges from 1.5 to 6.9 mph per second in the model data and from 2.3 to 6.9 mph per second in the validation data. These differences notwithstanding, the independence of the validation data set lends itself well to purposes of model evaluation.

The aggregation of existing in-use EPA data obtained from past testing efforts by both the EPA and CARB and used to develop the aggregate modal emission models in MEASURE is different from that of the validation data set in several respects. First, not all vehicles were tested on all cycles. Second, the vehicles recruited, in aggregate, are not representative of the national on-road vehicle fleet. Finally, there is very little replication testing, so within-driver variability is not known. However, the relatively large size of the aggregate database provides an opportunity to 1) obtain precise estimates of a multitude of vehicle-specific technology effects, 2) predict emission rates over a wide range of makes and model years, and 3) assess the effect of mileage accrual.

PREPARING THE MEASURE AND MOBILE5A MODELS FOR VALIDATION AND COMPARISON

Before being able to assess the predictive abilities of both MEASURE and MOBILE5a, it was necessary to set some ground rules for model validation and comparison. First, it was necessary to determine which "classes" of vehicles would be compared. In other words, it seemed that for at least some comparisons it would be useful to see how the two models predict emission rates for classes of

TABLE 5 **Number of Tests, Average Speeds, Maximum Speeds, and Maximum Instantaneous Acceleration of Each Test Cycle for the Model Development Data Set**

Cycle name*	Number of tests	Average speed (mph)	Maximum speed (mph)	Maximum acceleration (mph/sec)
Arterial 1	23	14.30	44.9	6.9
Arterial 2	21	24.06	46.3	5.8
Arterial 3	23	34.39	54.9	6.9
CCDH (bag 2)	58	13.40	29.8	3.0
Cycle 1	21	59.90	71.3	1.5
Cycle 2	23	53.31	68.0	2.0
Cycle 3	22	39.28	68.9	4.6
Cycle 4	20	31.54	61.9	3.3
Cycle 5	23	23.60	56.5	3.9
Cycle 6	21	15.94	40.9	5.0
Cycle 7	23	9.17	39.7	3.1
HFET	6586	48.20	59.9	3.2
LSP 1	813	2.45	10.0	2.4
LSP 2	814	3.62	14.0	2.5
LSP 3	815	4.04	16.0	3.4
NYCC	1218	7.07	27.7	6.0
SC12	1199	11.70	29.1	3.3
SC36	1201	36.50	57.0	6.0
Unified cycle (bag 2)	88	27.40	67.2	6.9
FTP (bag 2)	All	16.20	34.3	3.3

*Arterial 1, 2, and 3 denote cycles developed in California for roadway specific testing.
CCDH denotes a cycle developed for use by the Colorado Department of Health for high altitude testing.
Cycles 1, 2, 3, 4, 5, 6, and 7 represent high-speed cycles developed in California for roadway facility testing.
HFET stands for Highway Fuel Economy Test.
LSP 1, 2, and 3 refer to EPA's low-speed testing cycles.
SC12 and SC36 refer to EPA speed correction factor cycles.
Unified Cycle (LA92) refers to a new California laboratory testing cycle providing greater engine loads.

vehicles that are fundamentally different since emissions are characteristically different across such classes. Second, emission factors need to be converted to comparable and meaningful units, i.e., emissions in grams per second. Finally, appropriate criteria for comparison needed to be established.

Technology Class Definition

Four different emissions-control technology types were investigated during model development: fuel injection, catalytic conversion, transmission, and supplemental air injection. Each technology can be represented by several different types, as indicated below (with coding shown):

- Fuel Injection (FINJ)
  1. Port fuel injection (PFI), coded as finj1
  2. Carburetor and all pre-1980 domestic cars, coded as finj2
  3. Throttle body fuel injection (TBI), coded as finj3
  4. Unknown type pre-1980 import and both 1980 domestic and import, coded as finj4
- Catalytic Converter (CAT)
  1. None, coded as cat1
  2. Oxidation only, coded as cat2
  3. Three-way catalyst, coded as cat3
  4. Oxidation and three-way catalyst, coded as cat4
- Supplemental Air Recirculation (SAR)
  1. Pre-1980 of unknown type, coded as sar1
  2. None, coded as sar2
  3. Air pump, coded as sar3
  4. Pulse, coded as sar4
- Transmission Speed (TRAN)
  1. Automatic, coded as tran1
  2. Semi automatic, coded as tran2

3. Three-speed manual, coded as tran3
4. Four-speed manual, coded as tran4
5. Five-speed manual, coded as tran5

To capture the effects of deterioration, accrued test vehicle mileage was used as a surrogate for deterioration. Fomunung et al. (1999) have previously determined that deterioration appears to occur more like a step function rather than a constant deterioration over time, so four deterioration mileage groups (or bins) are employed in the models. These groups are "25,000 miles or less," "25,000 to 50,000 miles," "50,000 to 100,000 miles," and "100,000 miles or more." They are labeled km1, km2, km3, and km4, respectively.

It was a fairly complex task to implement the regression equations inside the MEASURE model. First, it was necessary to define mutually exclusive technology groups that would interact uniquely with vehicle operating modes. In essence, it was necessary to employ classification rules that resulted in mutually exclusive and collectively exhaustive technology groups. To define preliminary classification rules, a matrix of all possible combinations of the four technology variables plus the mileage bins and high-emitter status that appear in the regression model (a total of $4 \times 4 \times 4 \times 5 \times 4 \times 2$ or 2,560 technology rules) was created for each pollutant. Then, using equation (3) and each of equations (9), (10), and (11), which include technology and modal interactions, for CO, HC, and NO_x, respectively, the predicted emission rate ratio for each pollutant was computed for each of the 2,560 initial classification rules using the modal variables from the highway fuel economy test.

Classification rules that yielded the same predicted emission rate ratio for any given cycle were then clustered together; that is, they were collapsed into the mutually exclusive technology groups that are represented in the regression equation. A cross-check with modal variables from other driving cycles (LA4, Low Speed 1, and High Speed Cycle 1) produced the same technology groups. Each technology group cluster was then assigned an aggregate definition to represent a "technology group," as distinct from the former "classification rule." Consequently, 44 technology groups were defined for CO, 120 for HC, and 13 for NO_x, and all were assigned consecutive numerical labels beginning from 1. Thus, CO technology groups were labeled from 1 to 44; HC, from 1 to 120; and NO_x from 1 to 13. The definition of each technology group can be found in Fomunung (2000). The vehicle activity of each of these technology groups is then tracked separately within the MEASURE model because the technology and modal activity interaction variables appearing in the regression equations are different for each group.

Emissions Rates

The next step was to predict emissions for each pollutant for any given cycle and technology group. To predict emissions for each technology class one at a time, equation (7) is modified to

$$P_i = R_{ij} \times \psi_{ij\text{bag2}}, \tag{12}$$

where P_i is measured in g/sec and $\psi_{ij\text{bag2}}$ is now defined as the average of the base emissions rate (FTP bag 2), in g/sec, of pollutant i for technology class j. Note again that in this form, the term R_{ij} in equation (12) represents a cycle-specific correction factor for each technology class. The R_{ij} is the predicted rate ratio of pollutant i for technology class j. The values for $\psi_{ij\text{bag2}}$ are obtained from the FTP bag2 measurements in the original data set, while values for R_{ij} depend on the modal variables put into the model.

Criteria for Model Validation and Comparison

There are a number of model goodness-of-fit criteria that can be used to assess the difference between the emissions predicted by MEASURE and MOBILE5a and the emissions observed in the validation data. The focus in this paper is on point estimates of emissions. That is, an independent validation sample is used to compare the performance of MEASURE and MOBILE5a in predicting emissions of CO, HC, and NO_x. Overall model bias, the mean difference between predicted and observed emissions for a sufficiently large validation sample, reflects perhaps the most important criterion for comparing whether a model is working well in practice.

This study assesses the relative performance of the two models, MEASURE and MOBILE5a, using three statistical measures of effectiveness: the linear correlation coefficient, the root mean squared error (RMSE) (Neter et al. 1996), and the mean prediction error. The linear correlation coefficient reflects the degree to which a linear relationship exists between observed and predicted emissions. A high linear correlation coefficient would imply a close correspondence between paired data (predicted and observed emissions for vehicle i), whereas a low coefficient would imply the reverse. The RMSE is a measure of the prediction error. When comparing two models, the model with a smaller RMSE is a better predictor of the observed phenomenon. In addition, low values of RMSE accompanied by a high linear correlation coefficient is a good indicator that a model predicts well. The third measure of comparison is mean prediction error, ideally close to zero.

The MOBILE5a hot stabilized emission rates for each vehicle in the data set were predicted from the FTP bag 2 hot stabilized emission rate for each vehicle. The MOBILE5a input file provided by the EPA Region 4 office for Atlanta was modified to reflect 100% hot stabilized operations by setting the fractions of cold and hot start vehicle miles traveled (VMT) to zero. The model was set in a model year mode to predict emission rates for each model year. The model was then run in five-mph average speed increments to develop an emission rate matrix by model year and average speed for calendar year 1997. A matrix of emission rate ratios was developed from the emission rate matrix, with the 20-mph emission rate serving as the baseline emission rate (to conform with MOBILE5a internal assumptions related to the 19.6-mph average speed of the composite FTP). The emission rate ratio is equivalent to the speed correction factor implemented by MOBILE5a for each model year. The emission rate ratio for the average speed of the test cycle (found in the matrix using cubic spline interpolation) was then multiplied by the hot stabilized FTP bag 2 emission rate for that vehicle to estimate emissions on the alternative test cycle.

It is worth mentioning briefly that models were not compared based on the confidence in mean emission predictions, despite the fact that these comparisons may be useful. These comparisons are omitted for two important reasons. First, the data set used to estimate the emissions models within MEASURE is much larger than that used for estimating MOBILE5a, and thus statistical estimates are likely to be inherently more precise for MEASURE. Second, the regulatory arena in which models are employed has yet to embrace the use of confidence intervals on model outputs; therefore, comparisons of model efficiency would not likely lead to a strong argument for one model over another since precision is not applied in practice. It is not without hesitation that these comparisons have been omitted; the authors strongly believe that these types of comparisons are valid criteria for mounting evidence in favor of one model over another and could be useful in policy arenas.

RESULTS OF THE MODEL VALIDATION EXERCISE

This section describes the results of the validation of the MEASURE and MOBILE5a emission factor modules by comparing their prediction abilities across the set of validation data. Using validation vehicle characteristics and emissions results for each of the three pollutants CO, HC, and NO_x, the MEASURE and MOBILE5a emissions algorithms, shown in equations (9), (10), and (11), respectively, for the MEASURE model algorithms, were used to predict the observed data.

Because a vehicle fleet is usually tracked, in practice, by characterizing the number of vehicles in each technology class and by model year, model validation results were computed both for aggregate data (all vehicles) by driving cycle and by technology class. The results provided on a driving cycle basis yield information on how well the models explain variability in emissions due to differences in modal activity or driving profiles, while technology-class based results yield information on how well the models explain emission differences caused by disparate vehicle technologies.

The results of the performance evaluation are shown in tables 6 through 11. The linear correlation results on a cycle basis are shown in table 6, while table 7 (a, b, c) lists the results on a technology class basis for CO, HC, and NO_x, respectively. The number of vehicles tested on each cycle is shown in table 4, whereas table 7 (a, b, c), shows that the 797 vehicle tests in the validation data set are distributed into the following: 16 CO technol-

TABLE 6 Correlation Coefficients:* Observed vs. Predicted Using MEASURE and Observed vs. Predicted Using MOBILE5a, by Cycle

	Observed vs. predicted MEASURE (g/sec)			Observed vs. predicted MOBILE5a (g/sec)		
Cycle	CO	HC	NOx	CO	HC	NOx
ARTA	**0.559**	**0.702**	**0.391**	0.268	0.243	0.339
ARTC	**0.463**	**0.577**	**0.411**	0.368	0.199	0.269
ARTE	**0.432**	**0.606**	**0.398**	0.314	0.252	0.280
F505	**0.602**	**0.688**	**0.372**	0.266	0.302	0.313
FNYC	**0.399**	**0.601**	**0.446**	0.314	0.263	–0.001
FWAC	**0.642**	**0.647**	**0.496**	0.221	0.255	0.282
FWHS	**0.634**	**0.686**	0.428	0.229	0.232	**0.519**
FWYD	**0.581**	**0.522**	0.400	0.395	0.373	**0.513**
FWYE	**0.545**	**0.672**	0.428	0.370	0.323	**0.451**
FWYF	**0.464**	**0.598**	**0.465**	0.368	0.225	0.299
FWYG	**0.458**	**0.581**	**0.389**	0.336	0.264	0.265
LA92	**0.579**	**0.630**	**0.424**	0.355	0.230	0.321
LOCL	**0.465**	**0.616**	**0.434**	0.314	0.235	0.260
RAMP	**0.610**	**0.630**	0.361	0.306	0.357	**0.379**
ST01	**0.665**	**0.689**	**0.509**	0.192	0.151	0.205
WIDE	**0.561**	**0.682**	**0.424**	0.323	0.269	0.273

*Highest value for each comparison is **bolded**

TABLE 7a Linear Correlation Coefficients* for CO: Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

Tech class I.D.	Number of tests	Observed vs. predicted MEASURE (g/sec)	Observed vs. predicted MOBILE5a (g/sec)
3	16	0.512	**0.514**
6	16	**0.781**	0.548
11	32	0.164	**0.533**
14	190	**0.293**	0.120
19	16	**0.626**	0.467
20	32	0.599	**0.635**
21	64	0.578	**0.877**
22	112	**0.501**	0.456
23	176	0.433	**0.476**
27	16	**0.849**	0.765
33	15	**0.975**	0.908
36	32	0.500	**0.535**
39	16	**0.880**	0.809
40	16	**0.952**	0.908
41	32	**0.735**	0.534
42	16	**0.624**	0.439

*Highest value for each comparison is **bolded**

ogy classes out of a possible 44, 13 HC technology classes out of 120, and 5 NO_x technology classes out of 13. In addition, the number of vehicle tests in each technology class is shown. The results in table 6 show that for CO and HC, the MEASURE model outperforms MOBILE5a across all test cycles (the highest linear correlation coefficient in each comparison is bolded), while for NO_x both models perform equally well across almost all cycles, with MEASURE doing better in the rest of

TABLE 7b Linear Correlation Coefficients* for HC: Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

Tech class I.D.	Number of tests	Observed vs. predicted MEASURE (g/sec)	Observed vs. predicted MOBILE5a (g/sec)
32	16	**0.597**	0.555
34	16	**–0.099**	–0.065
38	16	0.095	**0.108**
51	16	**0.126**	0.115
54	16	–0.452	**–0.459**
77	304	**0.370**	0.092
80	191	0.145	**0.213**
84	79	**–0.110**	–0.042
95	64	**0.296**	0.111
96	16	**0.539**	0.460
97	15	**0.946**	0.915
108	16	0.075	**0.126**
112	32	0.085	**0.306**

*Highest value for each comparison is **bolded**

Table 7c Linear Correlation Coefficients* for NO_x: Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

Tech class I.D.	Number of tests	Observed vs. predicted MEASURE (g/sec)	Observed vs. predicted MOBILE5a (g/sec)
4	16	–0.288	**–0.296**
5	161	0.368	**0.449**
6	556	0.452	**0.497**
7	48	0.746	**0.939**
8	16	**0.926**	0.952

*Highest value for each comparison is **bolded**

the cycles. For the CO and HC results in table 7, no general trend is discernible, but it can be noted that for a majority of the results MEASURE performs equally well or better than MOBILE5a. For NO_x, however, MOBILE5a performs slightly better than MEASURE in four technology classes and significantly better in technology class seven.

Tables 8 and 9 contain the results from the root mean square error analysis (the smallest RMSE is bolded in each comparison). Table 8 shows the results on a cycle basis and table 9, on a technology class basis. As with the case of the linear correlation coefficient, the results on a cycle basis indicate that for CO and HC, MEASURE performs better than MOBILE5a, but for NO_x, MEASURE performs equally well or slightly better than MOBILE5a. On a technology class basis, MEASURE is only marginally better than MOBILE5a for CO and HC, and results are mixed for NO_x.

Table10 shows the result of the mean prediction error on a cycle basis, and table 11 shows the results on a technology class basis (smallest mean prediction error is bolded in each comparison). Also shown in both tables in underlined italics are the overall weighted average mean prediction errors per pollutant. To provide the reader with a quick assessment of the relative improvement of one model over the other, a column with the ratio of mean prediction error using MOBILE5a to that of MEASURE is highlighted in table 10. The same comparison on a technology class basis is shown in table 12. When comparing mean prediction error, it can be seen that MEASURE consistently overpredicts, while MOBILE5a consistently underpredicts, both on cycle and technology class bases. However, the same results indicate that across all cycles and

TABLE 8 Root Mean Square Prediction Error:* Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Cycle

	Observed vs. predicted MEASURE (g/sec)			Observed vs. predicted MOBILE5a (g/sec)		
Cycle	CO	HC	NO_x	CO	HC	NO_x
ARTA	**0.1038**	**0.0069**	**0.0071**	0.1362	0.0084	0.0072
ARTC	**0.0900**	**0.0078**	**0.0064**	0.1149	0.0086	0.0065
ARTE	**0.0763**	**0.0063**	**0.0048**	0.0980	0.0071	0.0050
F505	**0.0990**	**0.0056**	**0.0072**	0.1268	0.0068	0.0075
FNYC	**0.0679**	**0.0056**	**0.0024**	0.0830	0.0064	0.0029
FWAC	**0.2763**	**0.0095**	0.0166	0.3232	0.0110	**0.0162**
FWHS	**0.3271**	**0.0107**	0.0194	0.3615	0.0121	**0.0180**
FWYD	**0.1947**	**0.0084**	**0.0129**	0.2429	0.0099	0.0138
FWYE	**0.1189**	**0.0073**	**0.0074**	0.1529	0.0089	0.0080
FWYF	**0.0787**	**0.0069**	0.0059	0.1029	0.0078	0.0059
FWYG	**0.0658**	**0.0057**	0.0029	0.0784	0.0061	0.0029
LA92	**0.1099**	**0.0073**	**0.0080**	0.1505	0.0089	0.0085
LOCL	**0.0684**	**0.0058**	0.0042	0.0832	0.0064	**0.0041**
RAMP	**0.2153**	**0.0107**	**0.0121**	0.2666	0.0130	0.0128
ST01	**0.2031**	**0.0202**	**0.0092**	0.2564	0.0238	0.0093
WIDE	**0.0923**	**0.0064**	**0.0057**	0.1266	0.0080	0.0058

*Smallest value for each comparison is **bolded.**

TABLE 9 Root Mean Square Error (RMSE):* Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

MEASURE CO			MEASURE HC			MEASURE NO_x			MOBILE5a CO			MOBILE5a HC			MOBILE5a NO_x		
Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE
3	16	0.109	32	16	**1.09E-02**	4	16	**2.5E-03**	3	16	**0.020**	32	16	1.2E-02	4	16	3.4E-03
6	16	**0.071**	34	16	**7.41E-03**	5	161	8.4E-03	6	16	0.104	34	16	7.7E-03	5	161	**7.5E-03**
11	32	0.153	38	16	8.33E-03	6	556	7.9E-03	11	32	**0.145**	38	16	**1.1E-03**	6	556	**7.6E-03**
14	190	**0.076**	51	16	**1.10E-02**	7	48	**9.9E-03**	14	190	0.079	51	16	1.2E-02	7	48	1.3E-02
19	16	**0.121**	54	16	1.20E-02	8	16	**1.9E-02**	19	16	0.129	54	16	1.2E-02	8	16	3.5E-02
20	32	**0.225**	77	304	**6.28E-03**				20	32	0.230	77	304	6.4E-03			
21	64	0.083	80	191	3.80E-03				21	64	**0.072**	80	191	**3.7E-03**			
22	112	**0.055**	84	79	5.11E-03				22	112	0.064	84	79	**5.8E-03**			
23	176	**0.038**	95	64	**4.68E-03**				23	176	0.040	95	64	4.8E-03			
27	16	**0.109**	96	16	**2.14E-02**				27	16	0.123	96	16	3.8E-02			
33	15	**0.575**	97	15	**3.82E-02**				33	15	0.930	97	15	4.2E-02			
36	32	**0.420**	108	16	**5.49E-03**				36	32	0.557	108	16	2.5E-02			
39	16	**0.382**	112	32	**4.33E-02**				39	16	0.588	112	32	5.8E-02			
40	16	**1.255**							40	16	1.303						
41	32	**0.251**							41	32	0.480						
42	16	**0.088**							42	16	0.338						

*Smallest value for each comparison is **bolded.**
**n is the number of tests in each technology class of the validation data set.

TABLE 10 Mean Prediction Error:* Observed vs. Predicted Using MEASURE and Observed vs. Predicted Using MOBILE5a, by Cycle

Cycle	Observed vs. predicted MEASURE (g/sec)			Observed vs. predicted MOBILE5a (g/sec)			MOBILE5a error/MEASURE error (absolute ratio values)		
	CO	HC	NO_x	CO	HC	NO_x	CO	HC	NO_x
ARTA	**0.074**	**0.004**	0.006	–0.134	–0.008	–0.006	1.8	2.0	1.0
ARTC	**0.054**	**0.005**	0.006	–0.113	–0.008	–0.006	2.1	1.6	1.0
ARTE	**0.037**	**0.003**	**0.004**	–0.096	–0.007	–0.005	2.6	2.3	1.3
F505	**0.064**	**0.003**	0.006	–0.125	–0.006	–0.006	2.0	2.0	1.0
FNYC	**0.022**	**0.002**	**0.002**	–0.082	–0.006	–0.003	3.7	3.0	1.5
FWAC	**0.265**	**0.007**	**0.002**	–0.323	–0.010	–0.011	1.2	1.4	5.5
FWHS	**0.314**	**0.008**	**0.001**	–0.360	–0.011	–0.011	1.2	1.4	11.0
FWYD	**0.174**	**0.005**	**0.005**	–0.242	–0.009	–0.010	1.4	1.8	2.0
FWYE	**0.089**	**0.005**	**0.006**	–0.151	–0.008	–0.007	1.7	1.6	1.2
FWYF	**0.042**	**0.004**	0.005	–0.100	–0.007	–0.005	2.4	1.8	1.0
FWYG	**0.017**	**0.002**	0.002	–0.075	–0.006	–0.002	4.4	3.0	1.0
LA92	**0.087**	**0.005**	**0.007**	–0.150	–0.009	–0.008	1.7	1.8	1.1
LOCL	**0.021**	**0.002**	0.004	–0.079	–0.006	–0.004	3.8	3.0	1.0
RAMP	**0.201**	**0.009**	**0.011**	–0.266	–0.013	–0.012	1.3	1.4	1.1
ST01	**0.203**	**0.020**	0.009	–0.256	–0.024	–0.009	1.3	1.2	1.0
WIDE	**0.004**	**0.004**	0.005	–0.124	–0.008	–0.005	31.0	2.0	1.0
Weighted average	*0.104*	*0.005*	*0.005*	*–0.167*	*–0.0091*	*–0.007*	*1.6*	*1.8*	*1.4*

*Smallest value for each comparison is **bolded.**

TABLE 11 Mean Prediction Error:* Observed vs. Predicted Using MEASURE and Observed vs. Predicted Using MOBILE5a, by Technology Class

MEASURE									MOBILE5a								
CO			HC			NO_x			CO			HC			NO_x		
Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error
3	16	–0.109	32	16	**0.011**	4	16	**–0.001**	3	16	**–0.019**	32	16	–0.012	4	16	0.003
6	16	**0.045**	34	16	**0.007**	5	161	0.007	6	16	–0.102	34	16	–0.008	5	161	**–0.006**
11	32	**0.120**	38	16	–0.008	6	556	**0.004**	11	32	–0.141	38	16	**–0.001**	6	556	–0.006
14	190	**0.051**	51	16	**0.011**	7	48	**0.006**	14	190	–0.077	51	16	–0.012	7	48	–0.013
19	16	**0.118**	54	16	0.012	8	16	**0.009**	19	16	–0.129	54	16	–0.012	8	16	–0.035
20	32	**0.221**	77	304	**0.005**				20	32	–0.230	77	304	–0.006			
21	64	**–0.029**	80	191	**0.001**				21	64	–0.072	80	191	–0.003			
22	112	**0.050**	84	79	**–0.000**				22	112	–0.064	84	79	–0.005			
23	176	**0.028**	95	64	**0.003**				23	176	–0.036	95	64	–0.005			
27	16	**0.109**	96	16	**0.021**				27	16	–0.123	96	16	–0.038			
33	15	**0.575**	97	15	**0.038**				33	15	–0.930	97	15	–0.042			
36	32	**0.318**	108	16	**0.001**				36	32	–0.557	108	16	–0.025			
39	16	**0.382**	112	32	**0.036**				39	16	–0.588	112	32	–0.058			
40	16	**1.255**							40	16	–1.303						
41	32	**0.244**							41	32	–0.480						
42	16	**0.081**							42	16	–0.338						
		0.108			*0.005*			*–0.005*			*–0.167*			*–0.009*			*–0.007*

*Smallest value for each comparison is **bolded**.
** n is the number of cases in each technology class of the validation data set.

TABLE 12 Absolute Ratios of Prediction Error of MOBILE5a Results to MEASURE Results

CO MOBILE5a/MEASURE error		HC MOBILE5a/MEASURE error		NO_x MOBILE5a/MEASURE error	
Tech Class I.D.	Absolute ratio values	Tech Class I.D.	Absolute ratio values	Tech Class I.D.	Absolute ratio values
3	0.2	32	1.1	4	3.0
6	2.3	34	1.1	5	0.9
11	1.2	38	0.13	6	1.5
14	1.5	51	1.1	7	2.2
19	1.1	54	1.0	8	3.9
20	1.1	77	1.2		
21	2.5	80	3.0		
22	1.3	84	*		
23	1.3	95	1.7		
27	1.1	96	1.8		
33	1.6	97	1.1		
36	1.8	108	25.0		
39	1.5	112	1.6		
40	1.04				
41	2.0				
42	4.2				
Weighted average	*1.6*		*1.8*		*1.4*

* The ratio in this cell is a number divided by zero, which is undefined.

FIGURE 1 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Cycle Basis for Carbon Monoxide Emission Rates)

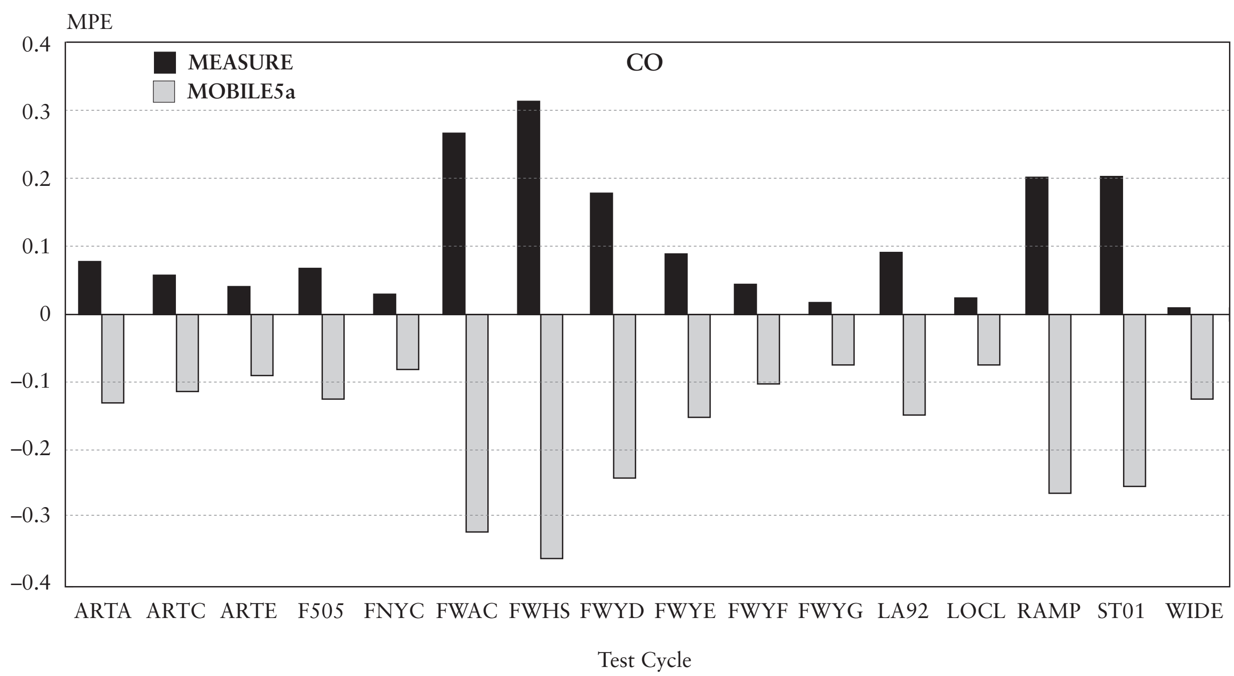


FIGURE 2 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Cycle Basis for Unburned Hydrocarbons Emission Rates)

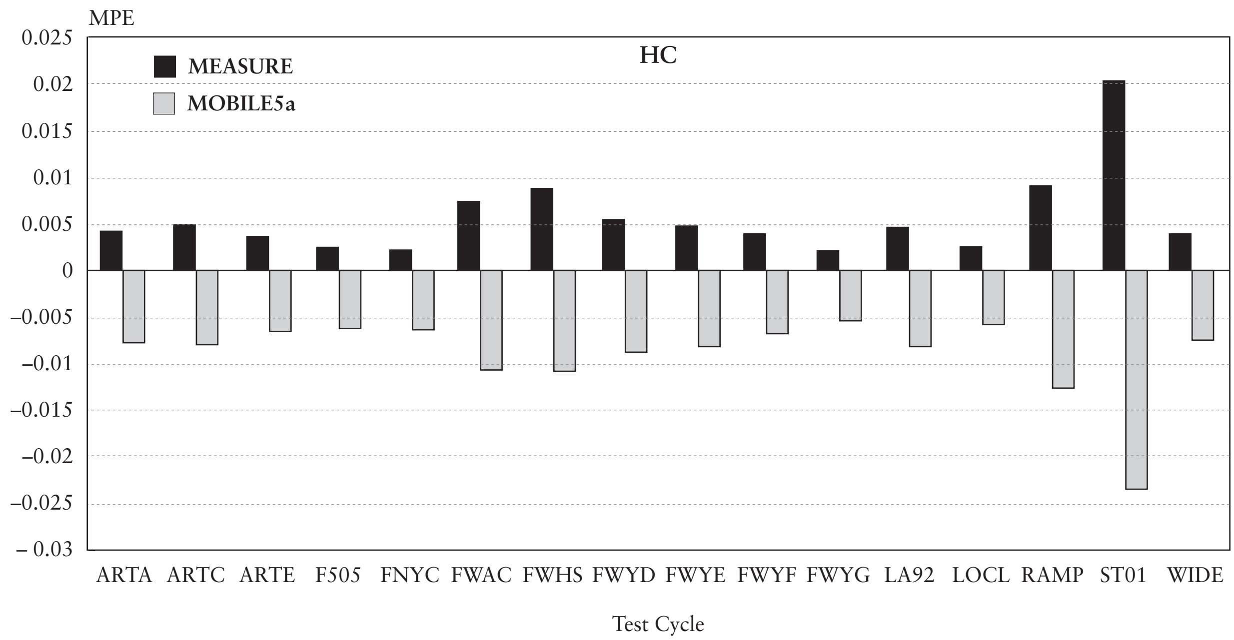


FIGURE 3 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Cycle Basis for Oxides of Nitrogen Emission Rates)

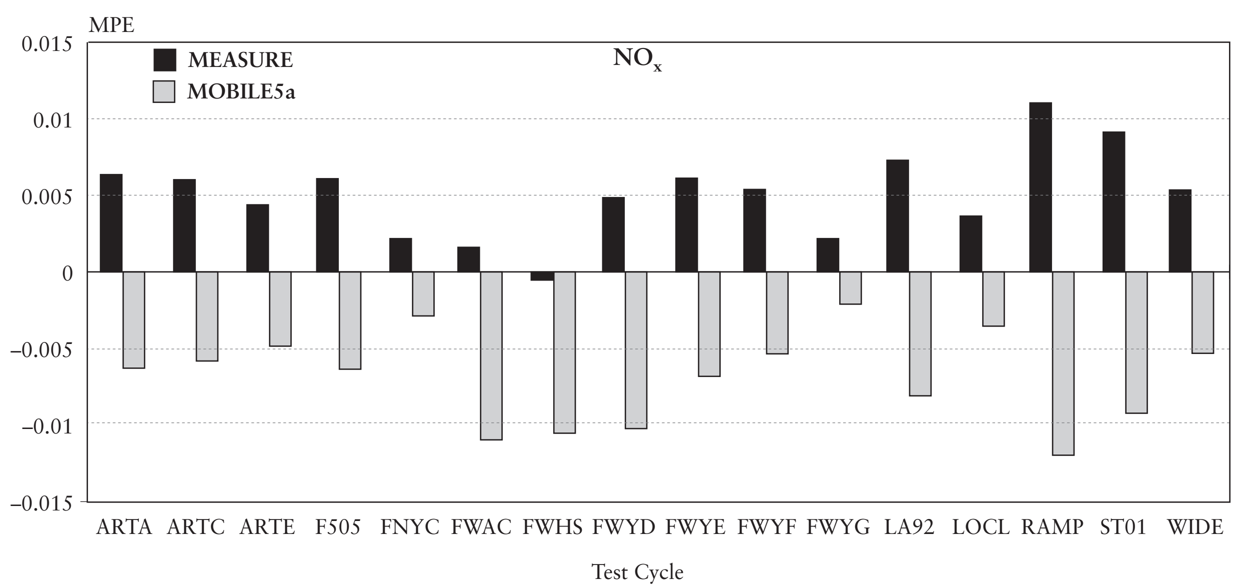
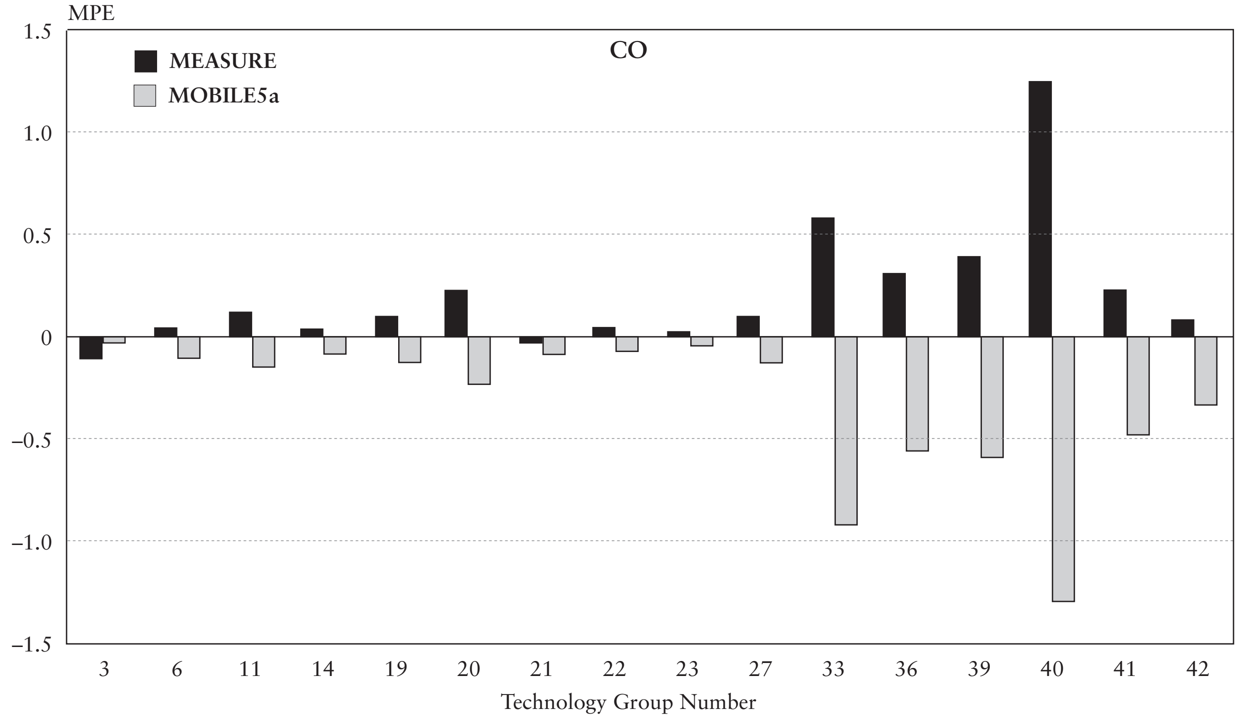


FIGURE 4 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Technology Group Basis for Carbon Monoxide Emission Rates)



technology classes the degree of overprediction (as measured by the magnitude of the errors) by MEASURE is lower than that of underprediction by MOBILE5a, demonstrating once again by this measure of assessment that MEASURE performs better than MOBILE5a. Pictorial representations of the mean prediction errors on a cycle basis are provided in figures 1 through 3 for CO, HC, and NO_x, respectively, and on a technology class basis in figures 4, 5, and 6, respectively.

CONCLUSIONS

The MEASURE model consistently showed larger correlation coefficients between observed and predicted emissions for the validation data set compared to MOBILE5a. The larger correlation coefficients suggest that the additional modal variables (beyond average speed) and their interactions employed in the MEASURE model provide additional explanatory power. The relatively smaller improvement in NO_x predictions stems from the fact that the average-speed approach to modeling NO_x emissions is not significantly inferior to using improved vehicle activity information; average speed seems to perform quite well for this pollutant.

Some of the driving cycles used in the validation study were designed by EPA contractors to represent on-road driving conditions under varying levels of congestion. Many of these cycles are significantly different from those that were used to develop the MOBILE5a and MEASURE models. The strong performance of the MEASURE model on these new cycles reveals the strength of applying the model to cycles outside those used to develop the model. These findings provide empirical support for the underlying principle that, although the models are cycle-based and aggregate, the discrete contributions of various modal contributions have been well modeled in MEASURE's modeling algorithms and can be used to model the emissions resulting from a variety of "off-cycle" vehicle activities.

In general, the results provided here are encouraging for MEASURE. The general superiority of MEASURE on mean prediction error suggests that if MEASURE and MOBILE5a were applied in practice for forecasting, MEASURE predictions would be more accurate, on average, by a factor of 1.6, based on the validation sample. On the basis of each pollutant, MEASURE would be more accurate

FIGURE 5 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Technology Group Basis for Hydrocarbon Emission Rates)

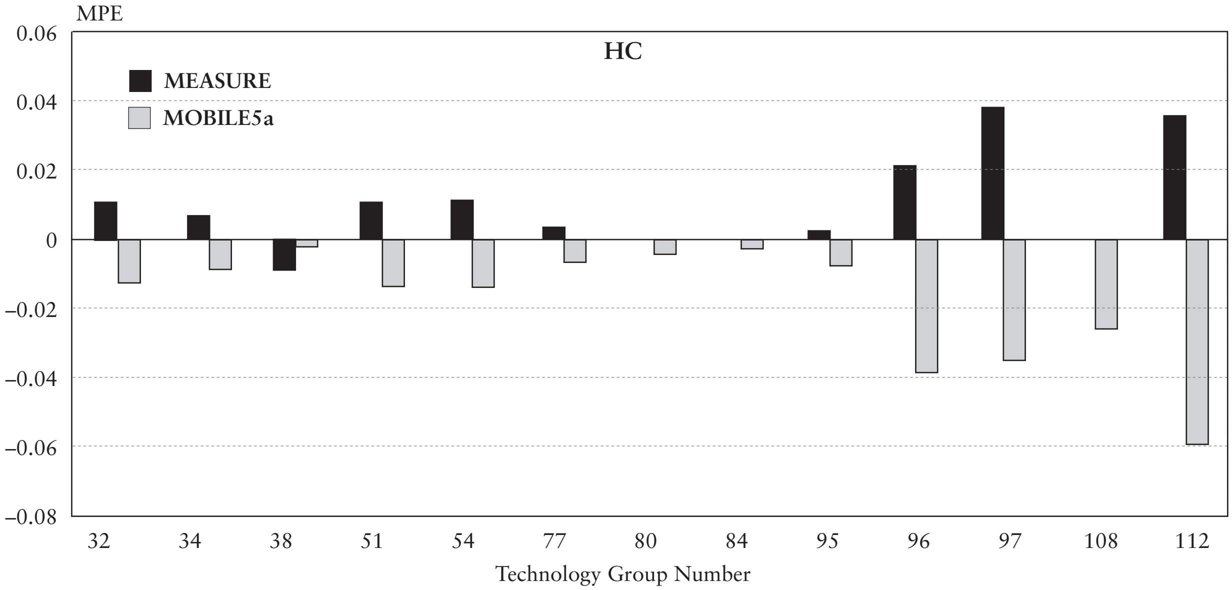
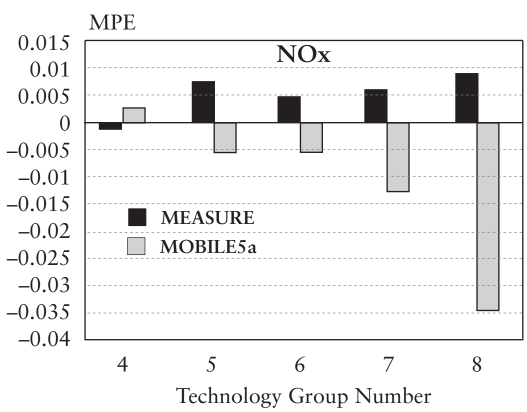


FIGURE 6 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Technology Group Basis for NOx Emission Rates)



by a factor of 1.6 for CO, 1.8 for HC, and 1.4 for NO_x. These factors are shown in underlined italics in the last column of table 10 and at the bottom of table 12. This is a compelling reason to favor MEASURE over MOBILE5a since systematic errors in emission rates will in practice be multiplied by the number of vehicles in an urban area and then again by the amount of mileage driven on a "typical day." MEASURE does slightly overpredict emissions for the validation sample, but this is not a significant concern since MEASURE would also slightly overpredict emissions reductions likely to be garnered from proposed control strategies. Thus, there is no expected major impact from using the model for control strategy modeling (i.e., as a comparative tool across control strategies and time).

Furthermore, the data used to develop MEASURE contained very few test results from 1994 and later model year vehicles. When new data from laboratory studies, such as the University of California, Riverside study by Barth et al. (1997), are included in the data set and the MEASURE algorithms are re-derived, the authors expect further improved performance in applications to the modern vehicle fleet.

REFERENCES

An, F., M. Barth, J. Norbeck, and M. Ross. 1997. The Development of a Comprehensive Modal Emissions Model: Operating Under Hot-Stabilized Conditions. *Transportation Research Record* 1587:52–62.

Barth, M., T. Younglove, T. Wenzel, G. Scora, F. An, M. Ross, and J. Norbeck. 1997. Analysis of Modal Emissions from a Diverse In-Use Vehicle Fleet. *Transportation Research Record* 1587:73–84.

Bishop, G.A., D.H. Stedman, and L. Ashbaugh. 1996. Motor Vehicle Emissions Variability. *Journal of the Air & Waste Management Association* 46:667–75.

Brieman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wardsworth International Group.

Cadle, S.H., R.A. Gorse, and T.C. Belian. Real-World Vehicle Emissions: A Summary. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Cicero-Fernandez, P., J. Long, and A. Winer. 1997a. Effects of Grades and Other Loads on On-Road Emissions of Hydrocarbons and Carbon Monoxide. *Journal of the Air & Waste Management Association* 47.

Cicero-Fernandez, P., W. Wong, and J. Long. 1997b. Fixed Point Mobile Source Emissions Due to Terrain Related Effects: A Preliminary Assessment, in *Proceedings of the 90th Annual Meeting of the AWMA*, Pittsburgh, PA, June.

Fomunung, I., S. Washington, and R. Guensler. 1999. A Statistical Model for Estimating Oxides of Nitrogen Emissions from Light-Duty Motor Vehicles. *Transportation Research D: Transport and Environment* 4/5:333–52.

Fomunung, I. 2000. Predicting emissions rates for the Atlanta on-road light-duty vehicular fleet as a function of operating modes, control technologies, and engine characteristics. Ph.D. diss., Georgia Institute of Technology.

Guensler, R. 1993. Data Needs for Evolving Motor Vehicle Emission Modeling Approaches. *Transportation Planning and Air Quality II*. New York, NY: American Society of Civil Engineers.

Guensler, R., S. Washington, and W. Bachman. 1998. Overview of the MEASURE Modeling Framework. *Transportation Planning and Air Quality III, Emerging Strategies and Working Solutions*. New York, NY: American Society of Civil Engineers.

Heywood, J.B. 1988. *Internal Combustion Engine Fundamentals*. New York, NY: McGraw-Hill.

Jimenez-Palacio, J.L. 1999. Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing. Ph.D. diss., Massachusetts Institute of Technology.

Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. 1996. *Applied Linear Statistical Models,* 4th edition. Boston, MA: Richard D. Irwin, Inc.

Rodgers, M.O. 1995. An Improved Power Demand Model for Evaluation of On-Road Vehicle Emissions. The Emission Inventory: Programs and Progress, EPA/AWMA. October.

Shih, R., S. Fable, and R.F. Sawyer. 1997. Effects of Driving Behavior on Automotive Emissions, in *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop, Coordinating Research Council,* Atlanta, GA.

Washington, S. 1994. Estimation of a vehicular carbon monoxide modal emissions model and assessment of an intelligent transportation technology. Ph.D. diss., University of California at Davis.

Webster, W.J. and C. Shih. 1996. A Statistically-Derived Metric to Monitor Time-Speed Variability in Practical Emissions Testing, in *Proceedings of the Sixth CRC On-Road Vehicle Emissions Workshop,* San Diego, CA, 18–28 March, Atlanta, GA: Coordinating Research Council.

Wolf, J., R. Guensler, S. Washington, and W. Bachman. 1998. High-Emitting Vehicle Characteristics Using Regression Tree Analysis. *Transportation Research Record* no. 1641:58–65.

Air Quality Assessment at a Congested Urban Intersection

MUTASEM EL-FADEL

MAJDI ABOU NAJM

HAYSSAM SBAYTI

Department of Civil and Environmental Engineering

American University of Beirut

ABSTRACT

Urban areas of Beirut suffer severe traffic congestion due to a deficient transportation system, resulting in significant economic losses. Grade separations are proposed at several congested intersections to alleviate this problem. Air quality, which greatly depends on the geometric configuration of an intersection, is a major environmental concern at these locations. This paper presents an air quality impact assessment at a typical urban intersection and addresses potential mitigation strategies for air quality management in urban areas. For this purpose, air quality measurements were conducted at representative locations to define existing pollutant exposure levels. Mathematical simulations were performed for several scenarios, both with and without grade separations, changes in vehicle mix, and level of service. Assessment of air quality impact significance was conducted by comparing simulated exposure levels with relevant air quality standards. Sensitivity analysis indicated that the introduction of a grade separation, changes in vehicle mix, and level of service lead to decreased exposure to air pollutants by up to 80%.

Mutasem El-Fadel, Department of Civil and Environmental Engineering. Faculty of Engineering and Architecture, American University of Beirut, 850 Third Avenue, New York, NY 10022. Email: mfadel@aub.edu.lb

INTRODUCTION

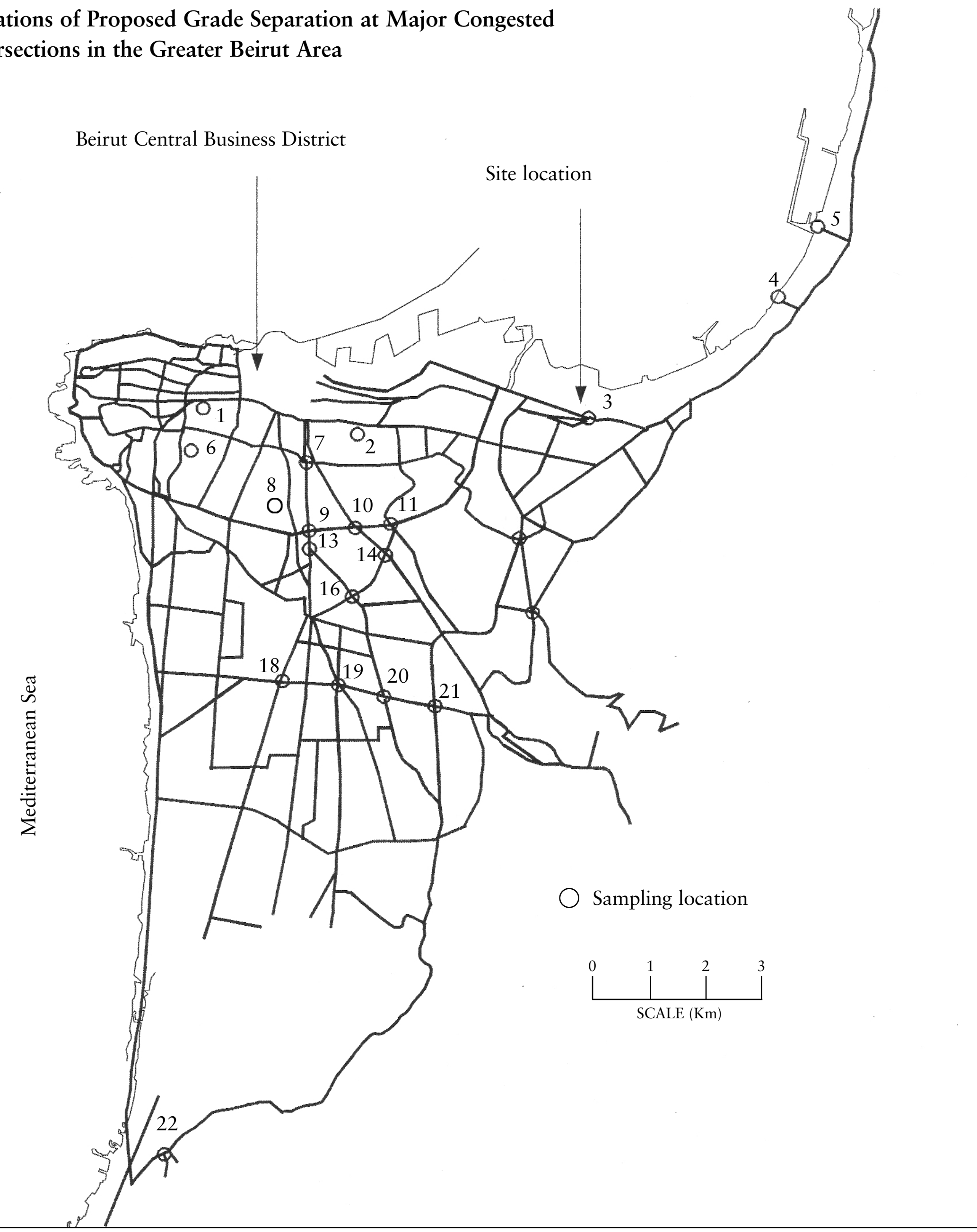
Traffic-induced emissions have been closely correlated with adverse impacts on air quality, especially in over-populated and highly congested urban areas. When industrial facilities are located away from urban centers, traffic circulation remains the most significant source of air pollutants. Urban centers are characterized by severe traffic congestion and an increased number of vehicles, leading to a longer peak-hour duration and higher air pollutant concentrations. This is particularly true in developing countries plagued with a general lack of traffic management and transport planning policies. The Greater Beirut Area (GBA) is a typical example, with 1.5 million passenger trips per day occurring on a relatively inferior road network, with a weak public transportation system, and without regulation enforcement (Staudte et al. 1997). Moreover, a relatively old and poorly maintained vehicle fleet increases the contribution of vehicle-induced emissions (TEAM International 1994). Consequently, residents are exposed to elevated concentrations of air pollutants, especially in hot summer periods when minimal air circulation and high humidity prevail.

Peak pollutant concentrations in urban areas are mostly encountered at heavily congested intersections, reflecting a high traffic volume and long delays, decreased average speeds, and a poor level of service (LOS). Vehicle-induced emissions at major intersections are dependent on many factors, including road geometry, traffic volume, vehicle fleet/fuel characteristics, driving patterns, and meteorological conditions (Hoglund 1994; Meng and Niemeier 1998; Faiz et al. 1996; Hallmark et al. 1998; Hoydysh and Dabberdt 1994). Various traffic management alternatives are used to improve traffic at congested intersections. Signalization, lane addition or lane widening, and the addition of roundabouts are among the most common traffic management alternatives. Historically, they have been implemented at intersections with various degrees of success. As congestion increases, however, these alternatives are often inadequate to accommodate the rise in traffic volume. Consequently, a grade separation or an interchange, where traffic flows without interruption in one or more directions, may be introduced.

Air quality monitoring provides the best means to characterize the state of emissions in the atmosphere, to evaluate the impact of various emission sources, and to assess the effect of new geometric configurations. However, monitoring can be inhibitively expensive to implement at every intersection. Economic considerations, coupled with the need for on-demand control and management of air pollution, have resulted in the development of a variety of guidelines and modeling techniques (mathematical algorithms) (Schattanek 1992; Schewe 1992; Zamurs et al. 1992). The most widely used models are Gaussian-based and require the definition of several factors: meteorological conditions, such as wind speed, wind direction, and atmospheric stability; emission rates, which depend on vehicle type and age, driving patterns, the type of pollution control equipment, and the level of inspection and maintenance; and the geometry of the specific intersection, including lane length and width, slope, receptor locations, and surface roughness. Some mathematical models can be used to determine fleet-average emission factors for each pollutant expressed as mass per distance traveled (grams/kilometer). Others, such as line source dispersion models, can be used to simulate atmospheric exposure levels.

This paper evaluates the impact on air quality of traffic-induced emissions at a typical intersection in a highly congested urban area. For this purpose, field measurements were first conducted to define existing pollutant exposure levels and to serve as a baseline for model calibration. The pollutants of interest were carbon monoxide (CO), nitrogen dioxide (NO_2), and total suspended particulate (TSP). The Mobile Vehicle Emissions Inventory (MVEI7G) and the California Line Dispersion Model (CALINE4), a roadside air dispersion model, were used to simulate vehicle fleet emission factors and atmospheric pollutant concentrations, respectively. Simulations were performed for worst case scenarios, including three specific factors: presence/absence of grade separations, changes in vehicle mix, and LOS. An assessment of the impact of vehicle-induced emissions was then conducted by comparing simulated pollutant concentrations to baseline air quality levels and relevant air quality standards. The overall objective was to optimize

FIGURE 1 Locations of Proposed Grade Separation at Major Congested Intersections in the Greater Beirut Area



intersection management in such a way as to minimize the impact of emissions on air quality.

Site Description

The site is one of several major intersections where a grade separation is proposed (see figure 1). The intersection under study provides the northern entrance to the future Beirut Central Business District. Traffic-counting meters were installed at these intersections to determine traffic volume at morning and afternoon peak-hour conditions. The Equilibre Multimodal-Multimodal Equilibrium (EMME/2), which models and forecasts the volume of traffic at any link based on the Multimodal Equilibrium Theory, was used to determine future traffic conditions at the intersection. EMME/2 offers the tools necessary to forecast future traffic conditions based on changes in road networks and socioeconomic conditions and is commonly used for traffic planning purposes (EMME 1998). The site layout, along with future (year 2010) peak-hour traffic volume, is depicted in figure 2.

FIGURE 2 Intersection Layout, with Projected Peak-Hour Traffic Volumes

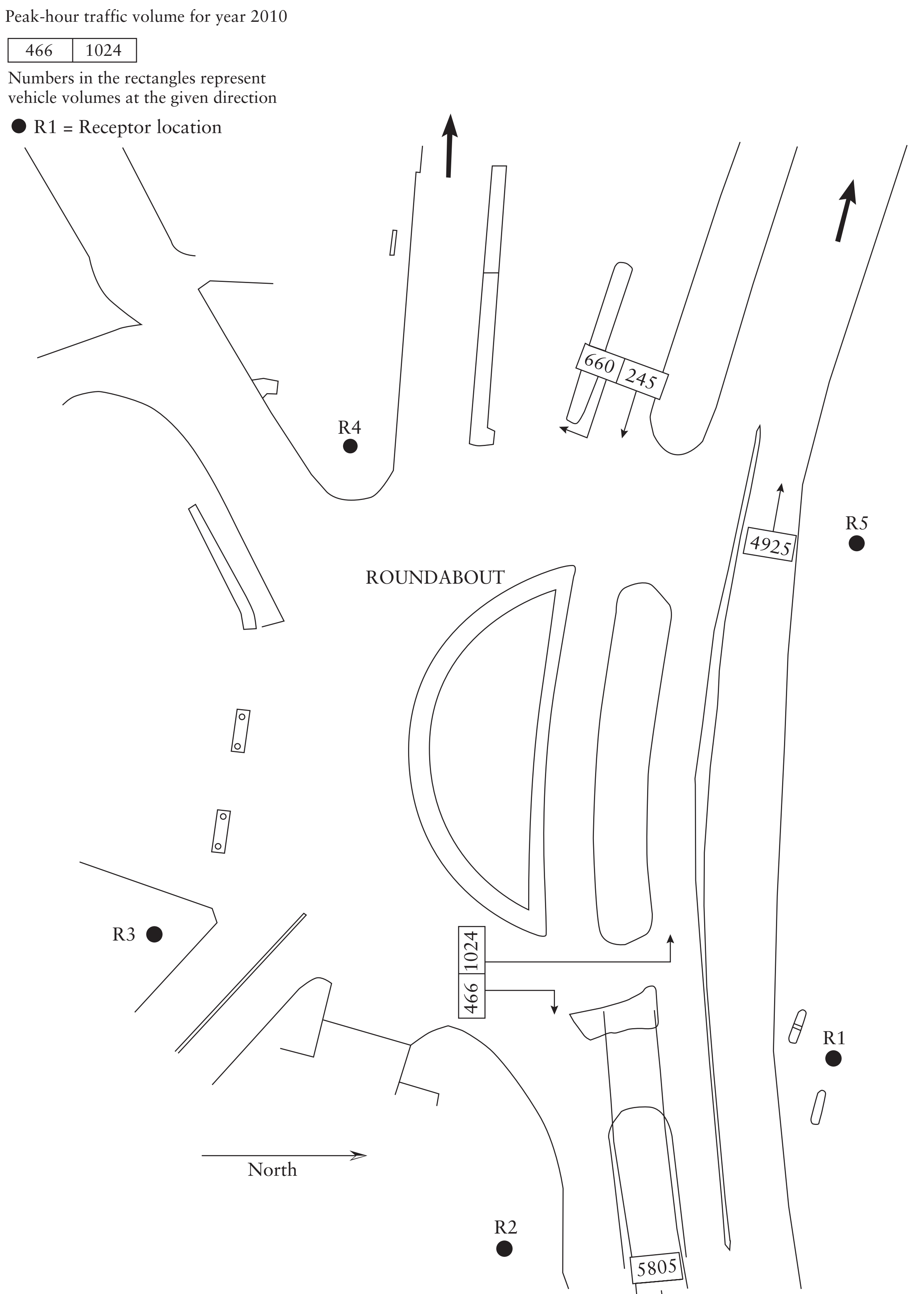


TABLE 1 Vehicle Fleet Composition, Occupancy, and Average Age

Vehicle type	Percent	Average occupancy (persons)	Average age (years)
Cars	90	1.5	14
Medium trucks	6	3	16
Heavy trucks	2	1	18
Buses	2	10	18

Sources: TEAM International 1994 and Dar Al-Handasah 1995

Traffic Fleet Characteristics

The fleet is mainly comprised of passenger cars and is characterized by relatively old and poorly maintained vehicles (see table 1). Although predictions indicate a decrease in passenger cars and an increase in bus trips resulting from the introduction of a more efficient mass transit system, implementation of such changes in infrastructure is unlikely to occur in the near future, due to minimal changes in the fuel taxation policy, weak urban planning practices, and a lack of enforcement of traffic regulations (TEAM International 1994; 1998).

IMPACT ASSESSMENT METHODOLOGY

The impact assessment methodology used consisted of the five consecutive steps identified in table 2. Applicable or relevant standards are defined first, followed by the determination of baseline conditions through field measurements, previous surveys, or mathematical modeling. In the third step, future traffic and air quality conditions are estimated using mathematical models. Potential impact is assessed in the fourth step by comparing future conditions with applicable standards and baseline conditions. Finally, mitigation measures to improve urban air quality and to achieve compliance with regulatory standards are addressed.

Using the approach described above, 13 different scenarios were developed and analyzed (see table 3). These scenarios include several variations of the three factors of interest. Scenario 1 represents the 2010 condition with no grade separation. Scenarios 2, 3, 4, and 5 represent future conditions with a grade separation coupled with or without speed improvement. The current vehicle mix was compared with two other alternatives, doubling or tripling bus ridership with or without speed improvements (scenarios 6, 7, 8, and 9). Finally, scenarios 10, 11, 12, and 13 represent the 2010 conditions under LOS E, D, C, and B, respectively (HCM 1994). Scenario 0 represents the present situation without any changes.

TABLE 2 Impact Assessment Methodology

Step	Description	Tools
1	Definition of applicable standards	Comparison with WHO and EPA standards
2	Determination of baseline conditions	■ Field measurements ■ Previous studies ■ Mathematical modeling
3	Simulation results	■ EMME: simulate vehicle volumes ■ MVEI7G: simulate emissions factors ■ CALINE4: simulate concentrations
4	Identification of potential impacts	■ Comparison with standards ■ Comparison with baseline
5	Mitigation of potential impacts	■ Regulatory ■ Technical

Definition of Applicable Air Quality Standards

Ambient air quality standards were proposed in Lebanon in the 1994 never-approved Proposed Law Number 1/52 (Ministry of Environment 1996). However, the proposed standards do not appear to have been developed on a scientific or country-specific economic basis (Staudte et al. 1997). The limit values are equal to or lower than thresholds employed in other parts of the world (see table 4) and seem to be unattainable, at least in the current regulatory and enforcement environment.

Definition of Baseline Conditions

Previous data on air quality in Beirut are practically nonexistent. Air samples were collected during the morning peak hours and analyzed for selected

TABLE 3 Description of Simulated Scenarios and Corresponding Road Traffic Conditions

Scenario	Year	Link number	Traffic volume (vehicles per hour)	Average lane speed (kph)	Purpose
0	1998	1	16	14	Definition of baseline conditions and model calibration
		2	100		
		3	3,120		
		4	126		
		5	283		
		6	3,522		
1	2010	1	660	14	Future conditions
		2	245		Implementation of proposed intersection
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
2	2010	1	660	14	Overpass
		2	245		Without speed improvement
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
3	2010	1	660	34	Overpass
		2	245		With speed improvement
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
4	2010	1	660	14	Underpass
		2	245		Without speed improvement
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
5	2010	1	660	34	Underpass
		2	245		With speed improvement
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
6	2010	1	660	14	Vehicle mix doubled bus ridership
		2	245		Without speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
7	2010	1	660	19	Vehicle mix doubled bus ridership
		2	245		With speed improvement
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
8	2010	1	660	14	Vehicle mix tripled bus ridership
		2	245		Without speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		

TABLE 3 Description of Simulated Scenarios and Corresponding Road Traffic Conditions *(continued)*

Scenario	Year	Link number	Traffic volume (vehicles per hour)	Average lane speed (kph)	Purpose
9	2010	1	660	24	Vehicle mix tripled bus ridership
		2	245		With speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
10	2010	1	660	48	LOS E
		2	245		With speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
11	2010	1	660	67	LOS D
		2	245		With speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
12	2010	1	660	75	LOS C
		2	245		With speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		
13	2010	1	660	80	LOS B
		2	245		With speed improvements
		3	4,925		
		4	466		
		5	1,024		
		6	5,805		

constituents. Table 5 contains measurements of NO_2 and TSP for the various intersections associated with the study. Measurement of CO concentrations was hindered by equipment malfunction. Generally, the results indicate the presence of greater NO_2 and TSP levels than allowed under ambient air quality standards. For example, the average measured concentration at the intersection under study for NO_2 and TSP were 28 parts per million (ppm), 581 μg/m^3 and 172.8 μg/m^3, respectively, both exceeding ambient air quality standards. Such levels are expected, due to several prevailing conditions conducive to air pollution, including:

1) the lack of periodic maintenance of the vehicle fleet,
2) a relatively old vehicle fleet,
3) the high frequency of acceleration and deceleration due to "stop-and-go" situations resulting from traffic congestion,
4) poor level of service of existing roadways,
5) the absence of regulations concerning vehicle emissions,
6) the minimal use of catalytic converters,
7) a weak and unreliable public transport system,
8) poor fuel quality, and
9) extensive construction activities (Staudte et al. 1997; TEAM International 1998).

Background levels were continuously monitored at a location away from traffic (on the campus of the American University of Beirut) for nearly one month. Average concentrations of 2 to 6 ppm (2,222 to 6,667 μg/m^3) for CO, 0.03 to 0.05 ppm (60 to 100 μg/m^3) for NO_2, and 50 to 80 μg/m^3 for PM_{10} (PM_{10} signifies particulate matter of 10 microns in diameter or smaller) were reported.

TABLE 4 Comparison of Selected Air Quality Standards

Parameter	Standard: Lebanese[a] µg/m³ (ppm)	US EPA[b] µg/m³ (ppm)	WHO[c] µg/m³ (ppm)	Averaging period
Nitrogen dioxide (NO_2)	200 (0.1)	NS	200 (0.1)	1 hour
	150 (0.075)	NS	150 (0.075)	24 hours
	100 (0.05)	100 (0.05)	NS	1 year
Carbon monoxide (CO)	30,000 (27)	40,000 (36)	30,000 (9)	1 hour
	10,000 (9)	10,000 (9)	10,000 (9)	8 hours
Total suspended particulate (TSP)	120	260	150-230	24 hours

NS = not specified
[a] Ministry 1996
[b] De Nevers 1995
[c] WHO 2000

TABLE 5 Summary of Average Air Quality Measurements (µg/m³)

Intersection	NO_2*	Particulate**
3	621	219.8
4	376	144.5
5	659	—[a]
7	659	176.6
10	847	151.7
11	884	136.0
12	470	194.5
13	376	101.7
14	715	192.6
16	753	130.8
17	282	165.9
18	658	179.4
19	753	291.0
20	564	207.4
21	339	139.1
22	339	160.5
Average	*581*	*172.8*

* ± 10 %
** ± 5 %
100 µg/m³ = 0.05 ppm for NO_2
[a] No data available

PM_{10} is equal to 0.55×TSP (Pearce and Crowards 1996; Vedal et al. 1987). Concentrations of 50 to 80 µg/m³ of PM_{10} correspond to 90 to 145 µg/m³ TSP), relatively high given that the measurements were collected as background readings at a non-congested location.

Emission Model: Mobile Vehicle Emissions Inventory (MVEI7G)

MVEI7G was used to determine emission factors for the Beirut fleet. The MVEI7G model, developed by the California Air Resources Board, estimates the total amount of pollutants released into the atmosphere by road transportation vehicles using statistical relationships based on emission tests for new and used vehicles. MVEI7G accounts for vehicle mix, the percentage of cold and hot starts, the existence and application of an inspection and maintenance program, the fraction of vehicles using catalytic converters, and the fraction of vehicles using gasoline or diesel. The model consists of four interrelated modules that operate together: CALIMFAC, WEIGHT, EMFAC, and BURDEN (see figure 3). The CALIMFAC and WEIGHT modules produce baseline vehicle emission rates and weighting factors for each model year, respectively. The EMFAC module uses this information, along with appropriate correction factors, to produce composite fleet emission factors. Finally, the BURDEN module combines emission factors with activity data to produce emission inventories (CARB 1996).

Emission Factors Assessment

An emission factor is the estimated average emission rate of a certain pollutant for a specific class of vehicles. Pollutants emitted from vehicles vary depending on vehicle characteristics; operating conditions; inspection and maintenance levels; fuel characteristics; and ambient conditions such as temperature, humidity, altitude, and wind speed and direction. Emission factors are strongly influenced by vehicle driving patterns, average speed, and the degree of acceleration and deceleration in the driving cycle (Garza and Graney 1996). These factors increase sharply at lower average

FIGURE 3 Flow Chart of MVEI7G Model

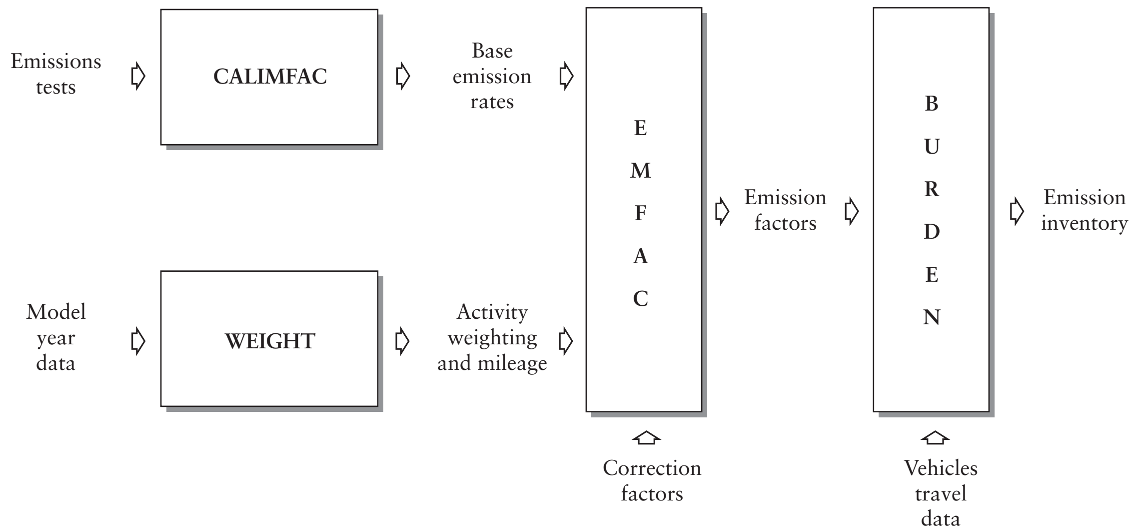


FIGURE 4 Variation of CO Emission Factor with Speed

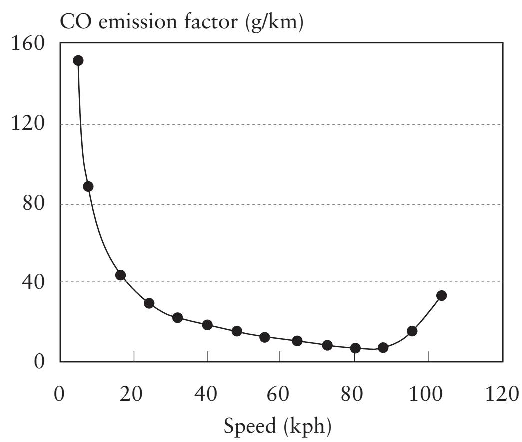
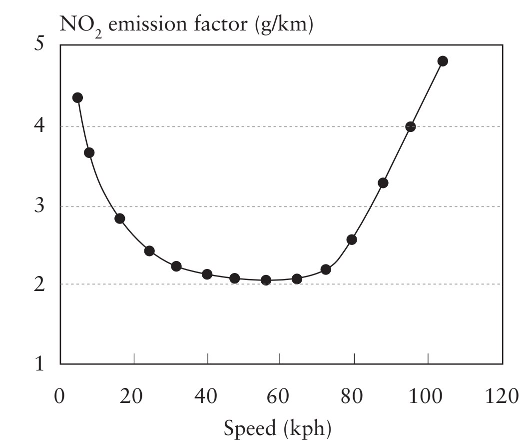


FIGURE 5 Variation of NO_2 Emission Factor with Speed



Note difference in vertical scale between figures 4 and 5.

speeds (see figures 4 and 5), typical of highly congested, stop-and-go urban driving. They decrease in free flow traffic at moderate speeds then increase again under relatively high-speed conditions (Krupnik 1991). Poorly maintained vehicles are responsible for a disproportionately higher share of total emissions (Faiz et al. 1996). Emission factors for U.S. gasoline-fueled passenger cars and medium-duty trucks equipped with different emission control technologies are presented in table 6. These factors are used as benchmark indicators in assessing emission factors of the Beirut fleet.

Model Application

MVEI7G was calibrated and applied to the Beirut area (El-Fadel and Bou-Zeid 1999). For this purpose, field surveys were conducted to measure tailpipe emissions and inspection and maintenance levels for the Lebanese fleet. Fuel composition, as well as the age distribution of the vehicle fleet, were determined and incorporated into MVEI7G. The calibration data and corresponding emission factors for selected pollutants are summarized in table 7. The estimated CO and NO_2 emission factors for the Beirut fleet fall between the noncatalyst control and the uncontrolled category of the estimated emission factors for the U.S. gasoline-fueled passenger cars. The estimated particulate emissions factor falls within the range of comparable vehicle fleets (Faiz et al. 1996).

TABLE 6 Estimated Emissions Factors for U.S. Gasoline-Fueled Vehicles

	Passenger cars		Medium-duty trucks	
Type of control	CO g/mile	NO_2 g/mile	CO g/mile	NO_2 g/mile
Advanced three-way catalyst control	9.92	0.83	16.32	0.83
Non-catalyst control	44.32	3.26	76.18	5.54
Uncontrolled	68.27	4.32	270.61	9.14

Estimated with the U.S. Environmental Protection Agency (USEPA) MOBILE5 model for a temperature of 24ºC, a speed of 31 kph, gasoline Reid vapor pressure of 62 kpa, and no inspection and maintenance program in place.

Source: Faiz et al. 1996

TABLE 7 Calibration Data for MVEI7G and Emissions Factors Obtained for Vehicle Fleet

Parameter	Value
Average fleet age in years	14
Sulfur content in fuel in ppm	40
Lead content in ppm	0.3
Unleaded fuel, percent	10
Fraction of hot starts, percent	10
Fraction of cold starts, percent	10
Hot stabilized conditions, percent	80
Inspection and maintenance, percent	10
Vehicles with catalytic converters, percent	1
Temperature in degrees Celsius	30
Number of starts per day	3.5
CO emission factor in grams per mile	60
NO_2 emission factor in grams per mile	3.15
PM emission factor in grams per mile	0.05

Source: El-Fadel and Bou-Zeid 1999

Dispersion Model: CALINE4

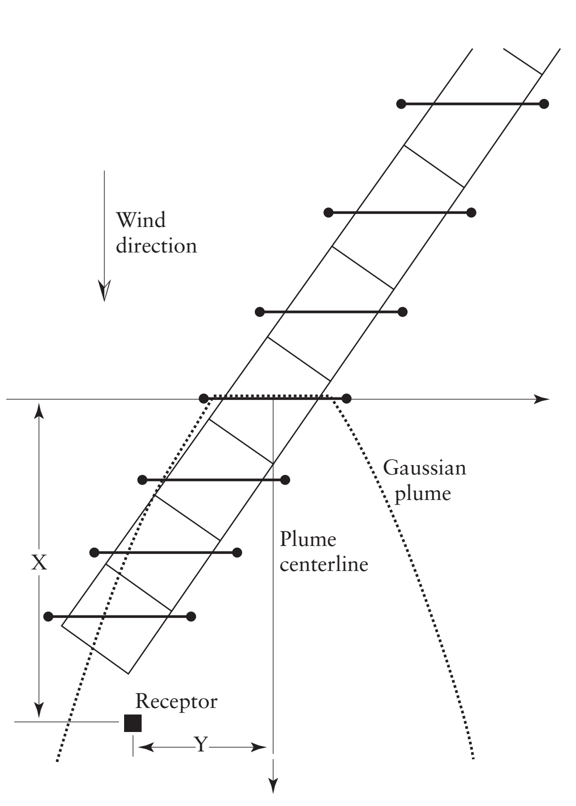
CALINE4, developed by the California Department of Transportation (Caltrans), is an atmospheric dispersion model used for predicting air pollutant concentrations. As noted, the model uses various factors to project air pollutant concentrations away from roadway line sources. For example, it estimates the concentration of CO, NO_2, and TSP from a roadway link at monitoring points within 500 meters from the road and can simulate air quality at intersections, street canyons, or parking facilities. The model subdivides a road into segments in a way similar to a finite element analysis, then sums up the contributions of these elements to the background air pollution levels at a specified receptor location (Benson 1989). A mixing zone concept is employed to evaluate pollutant mixing due to mechanical and thermal turbulence from car exhausts. The mixing zone consists of the road width, plus a three-meter border on either side, with a height equivalent to the mixing height attainable in the region. CALINE4 requires the input of emissions factors, obtainable from actual field measurements or models similar to MVEI7G.

The initial step is the specification of the first roadway element, whose position is a function of the roadway-wind angle (see figure 6). The first element remains constant and equal to its position at a roadway-wind angle of 45 degrees. Subsequent elements are longer since they are less critical for the total concentration value. This length adjustment saves computational time and is within the accuracy limits of other factors in the model. Equation (1) is used to compute the length of a roadway element. Each element is perpendicular to the wind direction, and its emissions are assumed to obey the Gaussian dispersion concept expressed in equation (2). As previously noted, worst case conditions are assumed in simulating pollutant concentrations for the present study due to several factors conducive to air pollution (see table 8).

$$EL = W\left[1.1 + \frac{PHI^3}{25\times10^5}\right]^{NE} \quad (1) \text{ and}$$

$$C(x,y,z) = \frac{q}{2\pi\sigma_y\sigma_z u}\exp\left(-\frac{1}{2}\frac{y^2}{2\sigma_y^2}\right) \times \left\{\exp\left[-\frac{1}{2}\left(\frac{z-H}{\sigma_z}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{z+H}{\sigma_z}\right)^2\right]\right\} \quad (2)$$

FIGURE 6 Element Series Represented by a Series of Equivalent Finite Line Sources



where C = pollutant concentration in grams/meter^3
EL = element length in meters
H = road height above receptor location in meters
h_m = source and receiver average height in meters
NE = element number in meters
PHI = wind-roadway angle in degrees
q = linear source strength in grams/meter^3
u = wind speed in meters per second
W = road width in meters
y_1, y_2 = distance from plume centerline to receptor in meters
z = height of receptor in meters
σ_y = horizontal dispersion parameter in meters
σ_z = vertical dispersion parameter in meters.

Mixing Height and Surface Roughness Justification

The effect of mixing height and surface roughness on CO was determined. For the present study, the critical mixing height, above which no change of pollutant concentration is observed, is 20 meters (see figure 7). The effect of surface roughness is less apparent (see figure 8), with the average value set at 100 centimeters, typically recommended as a default value.

TABLE 8 CALINE4 Calibration Data

Parameter	Value
Wind speed in meters per second	1
Wind direction	Worst case[a]
Wind standard deviation in degrees	5
Settling velocity of pollutants in meters per second	0
Surface roughness in centimeters	100
Mixing height in meters	20
Ambient concentration in ppm	0
Stability class[b]	G

[a] This is accounted for by the model itself.
[b] Atmospheric stability has been split into seven categories, labeled A through G, A being the most unstable and G being the most stable.

FIGURE 7 Projected CO Concentration (2010 Baseline) Based on Mixing Height

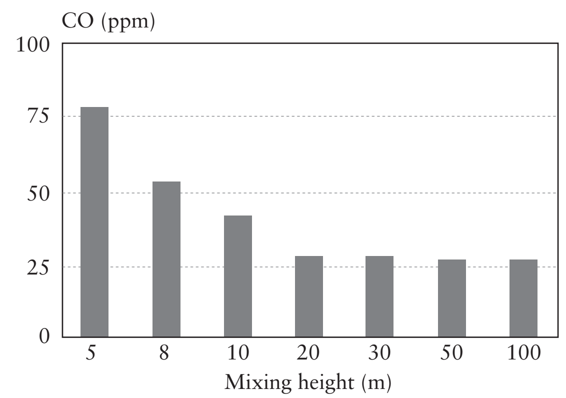
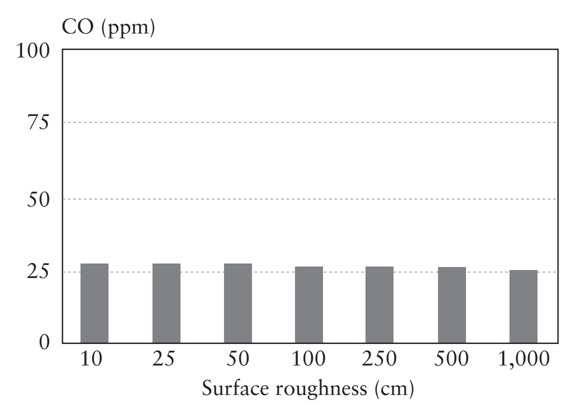


FIGURE 8 Projected CO Concentration (2010 Baseline) Based on Surface Roughness



Air Quality Simulations

During the operation phase, emissions are a function of the expected traffic conditions, volume and speed, at a particular location as well as the fleet characteristics. For this study, simulations were first conducted using the 1998 traffic conditions (see table 8). The simulated CO and NO_2 levels far exceed background levels where no traffic is present, indicating that traffic-induced emissions constitute the major contributor to CO and NO_2 levels. Since the simulated concentrations for NO_2 are of the same order of magnitude as those measured in the field (see table 9), this is another indication that traffic emissions are the major source of NO_2. However, the simulated TSP concentrations are far lower than field measurements, indicating that traffic is not the only contributor to particulate matter (see table 9).

Note that emissions during the road construction phase are a function of the excavation scheme and machinery used onsite. They consist primarily of particulate dust matter released as a result of earth removal activities and, to a lesser extent, of emissions from the onsite use of heavy construction equipment. While the extent of this impact cannot be reasonably quantified in a scientific manner due to the random nature of construction activities, it is typically temporary and confined to the immediate site vicinity, particularly if proper management measures are adopted to mitigate it.

Sensitivity Analysis

For the present study, model simulations were conducted to evaluate changes in the geometric configuration, vehicle mix, and LOS. Geometric changes consisted of the construction of a grade separation (overpass/underpass) to accommodate the heaviest traffic running in the north-south direction (see figure 2). Vehicle mix was modified, assuming the implementation of a mass transit system. Finally, the LOS was modified by adding an extra traffic lane along both directions of the overpass.

Type of Grade Separation

A grade separation is an effective transportation strategy aimed at increasing the average cruising speed and thereby reducing traffic delays at an intersection. Although its primary function is traffic management, a grade separation may help reduce pollutant concentrations. Pollutant emissions factors can be five to ten times higher in situations involving stop-and-go traffic due to the acceleration and deceleration processes (Faiz et al. 1996). On the other hand, increased average cruising speed can reduce emissions factors significantly (see figures 4 and 5). Note, however, that there is an upper limit of 50 and 90 kilometers per hour (kph), above which emissions factors start to increase again for NO_2 and CO, respectively. In addition, by virtue of its elevation, an overpass reduces exposure to air pollutants due to the increased time before a pollutant reaches a receptor at ground level. Similarly, an underpass confines air pollutants and, hence, reduces expo-

TABLE 9 Field and Simulated Concentrations of CO, NO_2, and TSP at Several Locations[a]

	CO		NO_2		TSP	
Intersection number[b]	Measured (ppm)	Simulated (ppm)	Measured (ppm)	Simulated (ppm)	Measured ($\mu g/m^3$)	Simulated ($\mu g/m^3$)
3	NM[c]	15.5	0.33	0.88	220	13.6
4	NM	28.9	0.20	0.51	144	27.1
19	NM	18.9	0.35	0.46	291	4.0
20	NM	8.5	0.30	0.32	207	7.6
21	NM	15.1	0.18	0.47	139	14.2

[a] The simulated values are the contribution of traffic emissions to the concentration of a particular pollutant in the air. They do not account for background levels or other potential sources in the area.
[b] See figure 1.
[c] NM = not measured because of equipment malfunction.

sure at ground level, particularly in the presence of an effective ventilation system.

In this study, concentrations of CO and NO_2 were simulated for the year 2010, given four of the different scenarios (2–5) previously described. Results are summarized in figures 9 and 10. The introduction of a grade separation reduced the concentration of both CO and NO_2. The reduction in CO concentration reached 56% and 86% for an overpass and an underpass, respectively, assuming a 20 kph increase in speed. The contribution of the geometry reached 7% and 71% for an overpass and underpass, respectively, assuming the same speed. The remaining decrease is attributed to the change in speed that directly affects emissions factors. The reduction in NO_2 concentration was less significant and reached 27% and 78% for an overpass and underpass, respectively, assuming 20 kph speed increase. Similarly, the contribution of geometry reached 7% and 70% for an overpass and underpass, respectively, assuming the same speed, with the remaining decrease attributed to the change in speed.

FIGURE 9 Effect of Grade Separation on CO Concentration

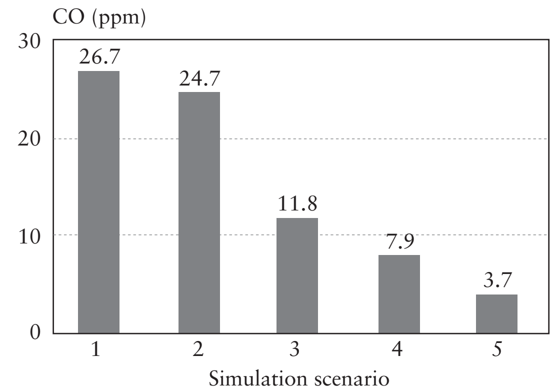
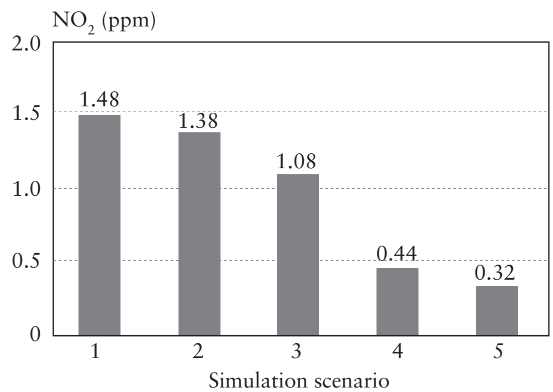


FIGURE 10 Effect of Grade Separation on NO_2 Concentration



Note difference in vertical scale between figures 9 and 10.

Vehicle Fleet Mix

Vehicle mix can have a significant effect on air quality. This is especially true when heavy vehicles, trucks and buses, are present at peak hours. In urban areas, traffic congestion relief and volume reduction can be accomplished through the expanded usage of mass transit systems. Such usage can change the vehicle mix and, hence, the extent of pollutant emissions. A traffic demand analysis indicates that the ridership share of a recently established public bus transport system in Beirut has reached 11%. The implementation of regulations and policies encouraging the use of this system can easily double this share to 22%. Assuming no additional demand for trips given the introduction of the new public transit system, the overall number of persons multiplied by trips will remain constant. The change will be in the mode of travel, with a shift from passenger cars, with average occupancy of 1.5, to buses, with average occupancy of 10 persons per vehicle trip (see table 1). It is assumed that trips gained by buses are lost from passenger cars. Doubling the percentage of buses in the fleet to account for doubling the demand for bus ridership results in reducing the traffic volume by 11.3% and increasing the average cruising speed by 5 kph. More stringent regulations can even triple the mass transit ridership share to reach 33%, reducing the traffic volume by 22.7% and increasing the average cruising speed by 10 kph. Certainly, the vehicle mix would vary if bus ridership is doubled or tripled (see table 10).

Simulation results for four alternatives, scenarios 6, 7, 8, and 9, are depicted in figures 11 and 12. The change in vehicle fleet mix has reduced CO concentrations because of traffic-volume reduction and increase in speed with a corresponding change in emissions factors. The interrelationship between these factors and emissions is illustrated in figure 13. The reduction in CO concentrations at the predefined receptor locations reached 29% and

TABLE 10 Vehicle Mix of Various Scenarios

Vehicle type	Scenario 1 bus ridership 11%	Scenario 6 bus ridership 22%	Scenario 8 bus ridership 33%
Cars	90	86.50	82.00
Medium trucks	6	4.50	7.75
Busses	2	6.75	7.75
Heavy trucks	2	2.25	2.50
Volume reduction	0	11.30	22.70

FIGURE 11 Effect of Fleet Mix on CO Concentration

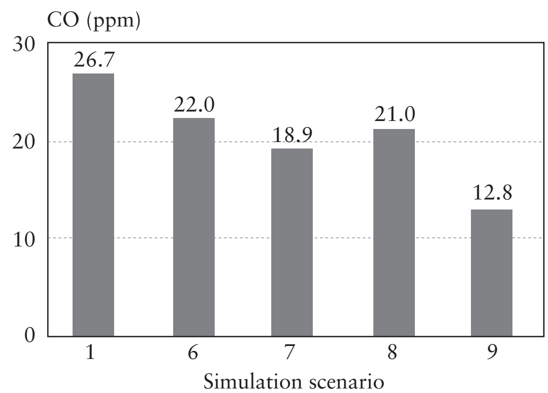
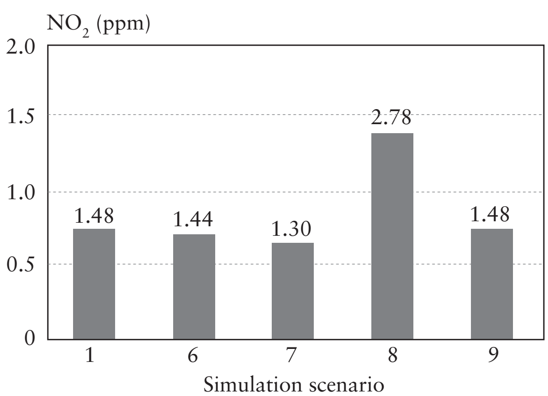


FIGURE 12 Effect of Fleet Mix on NO_2 Concentration



Note difference in vertical scale between figures 11 and 12.

52% for scenarios 7 and 9 with a 5 kph and 10 kph speed increase, respectively. As for NO_2 emissions, the situation is more complex. The contribution of fleet reduction, 11.3% and 22.7% for scenarios 6 and 8, did not reduce NO_2 concentrations. In fact, a 3% reduction, and an 89% increase in NO_2 concentrations were estimated for the 2 alternatives, scenarios 6 and 8. This can be attributed to the elevated emissions factors associated with heavy vehicles, leading to an increase in NO_2 concentrations. Moreover, the contribution of the 5 kph and 10 kph speed increase results in a net decrease of 12% and 0% in NO_2 concentrations for scenarios 7 and 9, respectively. This net reduction is not significant enough to bring the NO_2 to levels meeting air quality standards.

Level of Service (LOS)

LOS, which describes the performance of a roadway, can play a major role in establishing air quality levels. For the present study, the effect of LOS was analyzed using a slight modification to the intersection configuration. An overpass was assumed to serve the north/south traffic, the heaviest traffic flow, in both directions. The design speed for stopping sight distance (SSD) considerations is assumed to be 100 kph, with an actual average cruising speed of 60 kph (Papacostas and Prevedouros 1993). In addition, the maximum service flow (MSF) rate along the north/south direction is around 1,500 vehicle per hour per lane, which corresponds to LOS D (see table 11).

If the LOS is improved by adding another lane (both directions) in the overpass, then the facility would have an MSF of around 1,000 vehicles per hour per lane, making it eligible for both LOS C and B, corresponding to travel speeds of 75 kph and 80 kph, respectively (see table 11). Simulation results for the LOS scenarios are depicted in figures 14 and 15. Improvements in LOS reduced CO concentrations due to the increase in speed that results in a decrease in emissions factors (speed is less than 90 kph). On the other hand, changes in LOS increased NO_2 concentrations due to the increase in speed, resulting in an increase in emissions factors (speed is greater than 50 kph). This is true since for a LOS E and better, the average speed is greater than 50 kph, which falls in the range where the emissions factors increase with speed.

Assessment of Potential Air Quality Impacts

Assessment of the impact of emissions is conducted by comparing the simulated future exposure levels with existing air quality conditions and relevant local and World Health Organization (WHO)

FIGURE 13 Interrelationship between the Change in Vehicle Mix and Emissions

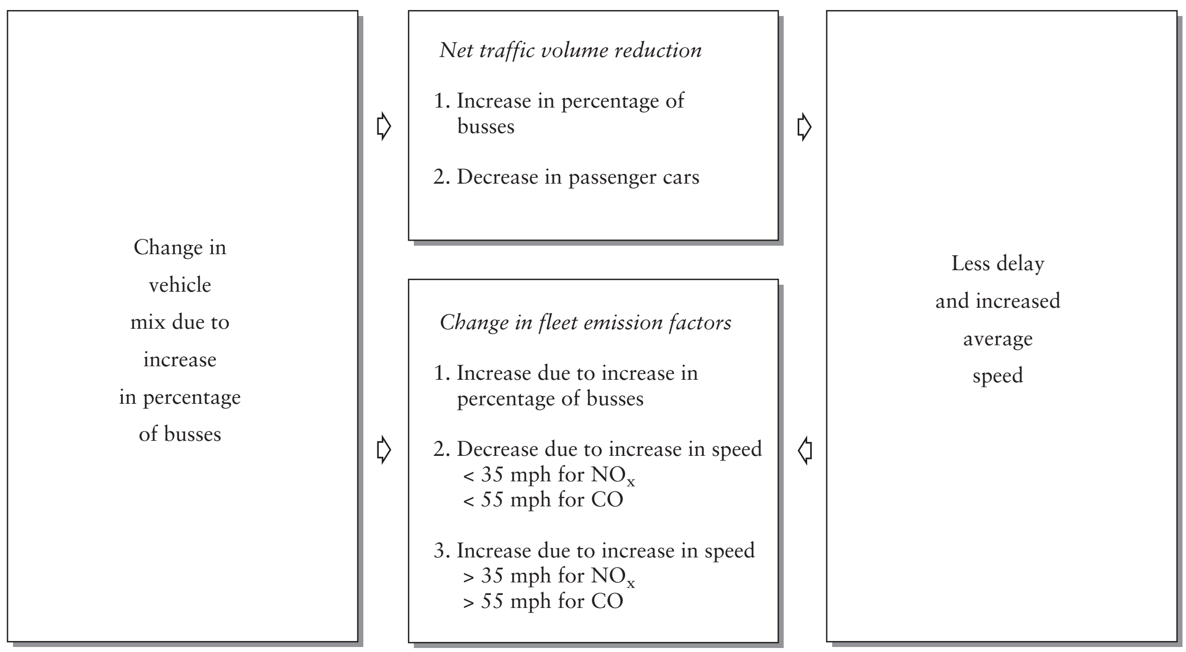


TABLE 11 Level of Service for a 100-kph Basic Freeway Section

LOS	Density vehicle/mile/lane	Speed kph	V/c[1]	MSF[2] vehicle/hour/lane
A	≤ 12	—[3]	—	—
B	≤ 20	≥ 80	0.49	1,000
C	≤ 30	≥ 75	0.69	1,400
D	≤ 42	≥ 67	0.84	1,700
E	≤ 67	≥ 48	1.00	2,000
F	> 67	< 48	—	—

[1] Volume to capacity
[2] Maximum service flow rate
[3] No data available
Source: HCM 1994

standards. During the operational phase, which can last indefinitely, the impact on air quality will be of a continual nature. In the present study, traffic emissions alone do not cause CO concentrations to exceed the WHO standards, but the additional background concentration does result in levels exceeding those standards. On the other hand, NO_2 levels from traffic emissions alone are higher than recommended standards, regardless of background levels.

Mitigation

In Lebanon, as in many developing countries, there is a lack of institutional capacity and technical expertise to deal with environmental issues. New legislation may fail to meet its objective unless a broad mix of measures is simultaneously implemented. In this context, project-specific measures and long-term policies are needed to mitigate potential impact on air quality. While technologies that ensure the removal of air pollutants are expected to grow in importance, mitigation measures must focus on separating pollution sources

FIGURE 14 Effect of LOS Change on CO Concentration

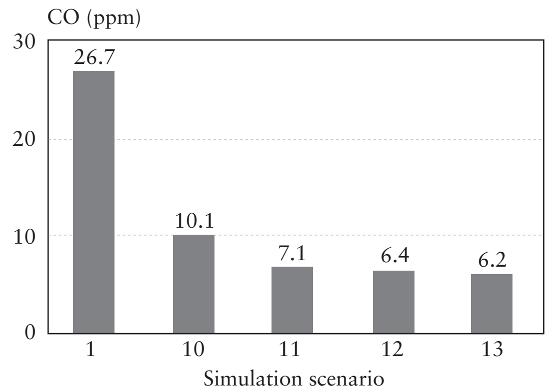
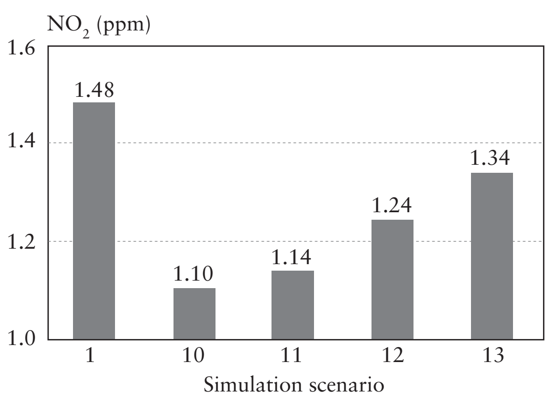


FIGURE 15 Effect of LOS Change on NO_2 Concentration



Note difference in vertical scale between figures 14 and 15.

TABLE 12 Summary of Possible Mitigation Strategies

Phase	Mitigation measure
Construction	■ Site and stockpile enclosure
	■ Spraying of stockpiles with chemical bonding agents
	■ On-site mixing in enclosed or shielded areas
	■ Proper unloading operations
	■ Water damping of stockpiles when necessary (dry conditions)
	■ Sealing of completed earthworks
	■ Re-vegetation as soon as possible
	■ Medium and heavily used haul routes permanently surfaced
	■ Damping unsurfaced haul routes
	■ Keep hauling routes free of dust and regularly cleaned
	■ Minimal traffic speed on-site with proper enforcement
	■ Maintenance and repair of construction machinery
Operation	■ Diverting traffic away from heavily populated areas
	■ Converting high-use vehicles to cleaner fuels
	■ Development of a comprehensive vehicle inspection and maintenance programs
	■ Imposing emission-related taxes
	■ Development of air quality standards and monitoring plans
	■ Increasing the share of less polluting traffic modes
	■ Using fuel-efficient vehicles
	■ Installing catalytic control devices

and receptors, reducing pollution activity and its characteristics, controlling emissions with filtering devices, and adopting and enforcing proper operational procedures.

Table 12 presents possible mitigation strategies. Emissions reduction will have multi-dimensional benefits since air pollution affects both public health and the environment. With the improved status of the health of the population, there will be less absence from work because of health problems and lowered costs of health insurance. Note that mitigation measures in the context of an intersection study are related mostly to construction activities if they are to be limited to site-specific measures. The interrelation between general mitigation measures that apply at the vehicle-fleet level and an intersection study is primarily through the estimation of emission factors. The reduction of the latter is not necessarily related only to an intersection study but rather to the entire vehicle fleet and should therefore be considered in this context.

SUMMARY

An air quality assessment was conducted for a typical congested urban intersection in Beirut. Air quality measurements were first obtained to define the existing levels of pollution. Mathematical simulations were then conducted to estimate vehicle emissions factors and to define concentrations of selected pollutants 1) with and without the grade separation, 2) with changing the fleet vehicle mix, and 3) with improvement of level of service. The simulations showed that these three factors can reduce exposure to CO concentrations in the air at ground level. Exposure to NO_2 was also reduced in most scenarios but not as significantly. Simulated TSP concentrations were far less than field measurements, indicating that traffic is not the only contributor to particulate matter. A summary of simulation results is shown in table 13, depicting the simulated concentrations and the net change in the pollutant concentration from baseline conditions. Note that in construction situations, which can arguably be a critical time for air quality, TSP concentrations are expected to exceed all standards. The extent of this impact, however, cannot be reasonably quantified in a scientific manner due to the random nature of construction activities.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to TEAM International for providing the project site design layout and data on current and future traffic volumes. Special thanks are extended the United States Agency for International Development for its continuous support for the Environmental Engineering and Science programs at the American University of Beirut.

REFERENCES

Benson, P.E. 1989. *Caline-4: A Dispersion Model for Predicting Air Pollutant Concentration Near Roadways,* FHWA/CA/TL-84/15. California Department of Transportation. Available at http://dot.ca.gov/hq/env/ext/soft.htm/CALINEManual.

California Air Resources Board (CARB). 1996. *Methodology for Estimating Emissions from On-Road Motor Vehicles,* Technical Support Division, Mobile Source Emission Inventory Branch, California EPA. California: California Air Resources Board. Available at http://www.arb.ca.gov/msei/mvei/mvdocs.htm.

Dar A.-H. 1995. *Maintenance & Rehabilitation of Lebanon Road Network*. Beirut, Lebanon: Council for Development and Reconstruction.

De Nevers, N. 1995. *Air Pollution Control Engineering*. New York, NY: McGraw-Hill, Inc.

El-Fadel, M. and E. Bou-Zeid. 1999. Greenhouse Gas Emissions from the Transportation Sector in Developing Countries: The Case of Lebanon. *Transportation Research, Part D*, no. 4:121–34.

TABLE 13 Summary of Simulation Results

		Simulated concentrations		Percent change in pollutant concentration[a]	
Scenario	Description	CO	NO_2	CO	NO_2
1	Do nothing (2010)	26.72	1.48	0.0	0.0
2	Effect of type of grade separation	24.74	1.38	7.4	6.7
3		11.76	1.08	56.0	27.0
4		7.88	0.44	70.5	70.3
5		3.72	0.32	86.1	78.4
6	Effect of vehicle mix	22.04	1.44	17.5	2.7
7		18.88	1.30	29.3	12.2
8		21.02	2.78	21.3	–87.8
9		12.76	1.48	52.2	0.0
10	Effect of level of service	10.14	1.10	62.1	25.7
11		7.08	1.14	73.5	23.0
12		6.42	1.24	76.0	16.2
13		6.16	1.34	76.9	9.5

[a] Percent change in pollutant concentration from baseline conditions (Scenario 1)

Equilibre Multimodal-Multimodal Equilibrium (EMME). 1998. EMME2 User's Manual Software Release 9. Montreal, Canada: INRO Co.

Faiz, A., C.S. Weaver, and M.P. Walsh. 1996. *Air Pollution from Motor Vehicles: Standards and Technologies for Controlling Emissions*. Washington, DC: World Bank.

Garza, V.J. and P. Graney. 1996. Transportation Project-Level Carbon Monoxide Protocol. Davis, CA: Institute of Transportation Studies, University of California, Davis.

Hallmark, S., R. Guensler, and J. Leonard II. Stopline Distributions of Speed and Acceleration for Signalized Intersections. Paper presented at the Air and Waste Management Association, 91st Annual Meeting and Exhibition, San Diego, CA, June 1998.

Highway Capacity Manual (HCM). 1994. Washington, DC: Transportation Research Board, National Research Council.

Hoglund, P. 1994. Alternative Intersection Design—A Possible Way of Reducing Air Pollutant Emissions from Road and Street Traffic? *The Science of the Total Environment* 146/147:35–44.

Hoydysh, W. and W. Dabberdt. 1994. A Fluid Modeling Study of Concentration Distributions at Urban Intersections. *The Science of the Total Environment* 146/147:425–32.

Krupnick, A.J. 1991. Transportation and Air Pollution in Urban Areas of Developed and Developing Countries. Discussion Paper QE91–08. Quality of the Environment Division, Resources for the Future. Washington, DC.

Meng, Y. and D. Niemeier. 1998. Project Level Carbon Monoxide Hot-Spot Analysis for Level of Service D Intersections. *Transportation Research Record* 1641: 73–80.

Ministry of Environment. 1996. *Standards and Regulations for Limiting Air, Water, and Soil Pollution*. Proposed Law Number 1/52. Beirut, Lebanon.

Papacostas, C.S. and P.D. Prevedouros. 1993. *Transportation Engineering and Planning*. Englewood Cliffs, NJ: Prentice Hall.

Pearce, D. and T. Crowards. 1996. Particulate Matter and Human Health in the United Kingdom. *Energy Policy* 24, no. 7:609–19.

Schattanek, G. 1992. Integrating Traffic and Air Quality Modeling Techniques to Predict Pollutant Concentrations Near Intersections, in *Conference Proceedings, Transportation Planning and Air Quality,* Santa Barbara, CA, 28–31 July, 1991. Urban Transport Division of the American Society of Civil Engineers.

Schewe, G.J. 1992. Modeling Guideline for Air Quality Analysis of Intersections, in *ASCE, Conference Proceedings, Transportation Planning and Air Quality,* Santa Barbara, CA, 28–31 July, 1991. Urban Transport Division of the American Society of Civil Engineers.

Staudte, M., M. Rau, and M. El-Fadel. 1997. Urban Air Quality Monitoring Program for the Greater Beirut Area. Transtec-Fitchner Consortium. Beirut, Lebanon: Ministry of Environment.

TEAM International. 1994. Greater Beirut Transportation Plan, Data Collection. Report no. 4. Beirut, Lebanon: Council for Development and Reconstruction.

_____. 1998. Environmental Impact Assessment of the Beirut Urban Transport Project. Beirut, Lebanon: Council for Development and Reconstruction.

Vedal, S., M.B. Schenker, A. Munoz, J.M. Samet, S. Batterman, and F.E. Speizer. 1987. Daily Air Pollution Effects on Children's Respiratory Symptoms and Peak Expiratory Flow. *American Journal of Public Health* 77:694–9

World Health Organization. (WHO). 2000. Guidelines for Air Quality, Air Management, and Information Systems (AMIS). Available at http://www.who.int/peh.

Zamurs, J., R. Conway, and S.S. Rosen. 1992. Intersection Air Quality Analysis, in *ASCE, Conference Proceedings, Transportation Planning and Air Quality,* Santa Barbara, CA, 28–31 July, 1991. Urban Transport Division of the American Society of Civil Engineers.

JOURNAL OF TRANSPORTATION AND STATISTICS

Guidelines for Manuscript Submission ▾ Instructions to Authors

After retaining a copy for your files, submit two double-spaced copies of your manuscript to:

Deborah Moore, Managing Editor
Journal of Transportation and Statistics
Bureau of Transportation Statistics
US Department of Transportation
Room 3430
400 7th Street SW
Washington DC 20590

As the manuscripts are subject to peer review, submission of a paper indicates the author's intention to publish in the *Journal of Transportation and Statistics*. Simultaneous submission of a manuscript to other journals is unacceptable. Previously published manuscripts, whether in an exact or approximate form, cannot be accepted. Check with the editor if in doubt.

All manuscripts must be prepared in an electronic format and double-spaced with a 1″ margin on all sides. Double-space all material, including quotations, the abstract, and any notes. All figures and tables should appear at the end of the paper, and each should be on a separate page. Include the figure name and filename on the bottom of the page.

Articles prepared in LaTeX software will be accepted.

Cover page. This page should include the title of the manuscript, the author's name, mailing and email addresses, and telephone number. The following page should contain a one-paragraph abstract of the manuscript of about 100 words. The abstract should include the title of the manuscript, briefly describe the contents of the paper, and explain the research method and results. The abstract should not contain any abbreviations or acronyms.

Graphic elements. Make sure that figures and tables are called out in the text. The title of your article, along with the figure or table number, should be written on the bottom of the page. Figures and tables should appear at the end of the paper, and each should be on a separate page. Figures need to be on high-quality paper, no more than 8½″ x 11″ in size, in black ink. Complicated graphics are to be avoided, and print should be large enough to allow for considerable reduction.

References. The reference style follows the *Chicago Manual of Style*. In the text, include the author, year, and page number. For example: Road dust also accounted for over 30 percent of emissions of very fine particles—those measuring 2.5 microns or smaller. (Barnard 1996, 17). Complete references must appear at the end of your manuscript. Include enough information in your reference so that the reader can find the document or contact its author. References should be alphabetized by authors' last names. When an author has more than one publication, place the oldest one first. Include the place of publication and publisher for books, and volume, month, or issue, and pages for journals. Examples of various types of references follow:

Gordon, R. 1990. *The Measurement of Durable Goods Prices*. Chicago, IL: University of Chicago Press.

in the text: (Gordon 1990)

Kargon, R.H., ed. 1974. *Maturing of American Science*. Washington, DC: American Association for the Advancement of Science.

in the text: (Kargon 1974)

Michaelis, L., D. Bleviss, J.-P. Orfeuil, and R. Pischinger. 1996. *Mitigation Options in the Transportation Sector. Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analysis, Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by R.T. Watson, M.C. Zinyowera, and R.H. Moss. Cambridge, England: Cambridge University Press.

in the text: (Michaelis et al. 1996)

National Academy of Sciences (NAS), Committee on National Statistics. 1985. *Sharing Research Data*. Washington, DC: National Academy Press.

in the text: (NAS 1985)

Yates, C. 1992. Eurocontrol Comes of Age. *Jane's Airport Review* 4:3.

in the text: (Yates 1992)

Papers selected for publication in the journal. An electronic copy of the manuscript must be submitted on a 3½″ diskette, along with two paper copies of the article. Make sure the hard copy corresponds to the electronic copy. Label the diskette with your name, the date, and the title of your paper. Please specify what software was used, including the release number (e.g., Word for Windows). Send a list of filenames. Please keep a copy of the electronic file. Send separate files for each table, figure, and the text. Do not imbed tables and figures in the text files. Indicate the approximate placement of each figure and table in the text.

Once the page proofs of your manuscript are returned to you, you will be required to return them to the Managing Editor within 48 hours. If the proofs are not returned, we will assume the article is correct as is.

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 3 Number 2
September 2000
ISSN 1094-8848

CONTENTS