♻ EPA A GIS-Based Modal Model Of Automobile Exhaust Emissions

NARSTO

Prepared for National Risk Management Research Laboratory

Prepared by Air Pollution Prevention and Control Division

FOREWORD

The U.S. Environmental Protection Agency is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the Agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The National Risk Management Research Laboratory is the Agency's center for investigation of technological and management approaches for reducing risks from threats to human health and the environment. The focus of the Laboratory's research program is on methods for the prevention and control of pollution to air, land, water, and subsurface resources; protection of water quality in public water systems; remediation of contaminated sites and groundwater; and prevention and control of indoor air pollution. The goal of this research effort is to catalyze development and implementation of innovative, cost-effective environmental technologies; develop scientific and engineering information needed by EPA to support regulatory and policy decisions; and provide technical support and information transfer to ensure effective implementation of environmental regulations and strategies.

This publication has been produced as part of the Laboratory's strategic long-term research plan. It is published and made available by EPA's Office of Research and Development to assist the user community and to link researchers with their clients.

E. Timothy Oppelt, Director
National Risk Management Research Laboratory

EPA REVIEW NOTICE

A GIS-Based Modal Model of Automobile

Exhaust Emissions

FINAL REPORT

Prepared by:

William H. Bachman

School of Civil and Environmental Engineering
Center for Geographic Information Systems
Georgia Institute of Technology
Atlanta, Georgia 30332

EPA Cooperative Agreement CR823020

Project Officer: Carl T. Ripberger
U.S. Environmental Protection Agency
Air Pollution Prevention and Control Division
Research Triangle Park, North Carolina 27711

Prepared for:

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

ABSTRACT

Suburban sprawl, population growth, and auto-dependency have, along with other factors, been linked to air pollution problems in U.S metropolitan areas. Addressing these problems becomes difficult when trying to accommodate the needs of a *growing* population and economy while simultaneously *lowering* or maintaining levels of ambient pollutants. · Growing urban areas must, therefore, continually develop creative strategies to curb increased pollutant production.

This report presents progress towards the development of a computer tool called MEASURE, the Mobile Emission Assessment System for Urban and Regional Evaluation. The tool works towards a goal of providing researchers and planners with a means for assessing new mobile emission mitigation strategies. The model is based in a geographic information system (GIS) and uses modal emission rates, varying emissions according to vehicle technologies and modal operation (acceleration, deceleration, cruise, and idle). Estimates of spatially resolved fleet composition and activity are combined with situation-specific emission rates to predict engine start and running exhaust emissions. The estimates are provided at user-defined spatial scales. A demonstration of model operation is provided using a 100 square kilometer study area located in Atlanta, Georgia. Future mobile emissions modeling research needs are developed from an analysis of the sources of model error.

TABLE OF CONTENTS

TABLE OF FIGURES

GLOSSARY AND ACRONYMS

AQL – Georgia Tech's Air Quality Laboratory.

ARC - The Atlanta Regional Commission, the MPO for Atlanta, Georgia.

ARC/INFO   UNIX Based GIS Software by Environmental Systems Research Institute.

CBD – Atlanta's Central Business District.

Conflation - The process of transferring textual information from one linear data representation to another.

Engine start - Term referring to the emission rate phenomenon occurring during the first few minutes of a vehicle's operation.

Enrichment   Term referring to the emission rate phenomenon occurring during high power demand driving.

FTP   Federal Test Procedure, the emission test cycle from which the MOBILE5a emission rates were derived.

Geocoding   The process of establishing locational parameters (coordinates) from textual data.

GIS   Geographic Information System, computer hardware and software used for storing, displaying, analyzing, and modeling spatial information.

GPS   Global Positioning System, a device used to determine one's position on the earth's surface by triangulating distances from satellites.

HC – Hydrocarbons.

High emitters   Term applied to a small portion of the fleet that produces higher emission rates, usually the result of malfunctioning equipment.

Hot-stabilized - Term referring to the 'stable' emission rates characteristic of vehicles operating with active emission control equipment, usually occurring after a vehicle has warmed sufficiently.

LOS – Level of service is used to characterize operational conditions within a traffic stream and their perception by motorists and passengers.

Makefile - A text script used to manage multiple programs and files.

MEASURE   Mobile Emission Assessment System for Urban and Regional Evaluation, the model developed by the research reported here.

MOBILE5a - The active mandated emission rate model developed by the USEPA.

Modal emissions - Emissions that have been separated by specific operating conditions that result in distinct changes in emission rate behavior.

MPO – Metropolitan Planning Organization.

NAAQS - National Ambient Air Quality Standards, health-based air quality standards that cities must not exceed.

Normal emitters - Term applied to vehicles with low to moderate emission rates due to normal operation of emission control equipment.

NO_x – Nitrogen oxides.

Photochemical models - Computer models used to predict ambient air quality.

Pollutants of concern   Carbon monoxide, hydrocarbons, and oxides of nitrogen.

Ozone   Pollutant caused by the complex mixing process of NO_x and HC in the presence of sunlight.

Raster - Cell-based spatial data structure.

Running exhaust - Term applied to non-start exhaust pipe emissions that occur while a vehicle is in operation.

SCF - Speed Correction Factor, the tech   e found in MOBILE5a for adjusting emission rates based on the average spe   i a vehicle, or sets of vehicles.

SOV – Single occupancy vehicle.

Sub-fleet - Term applied to any group of vehicles smaller than a regional operating fleet.

TAZ – The TAZ represents a spatial unit for aggregating socioeconomic data and resulting trip generation estimates.

Technology group Term applied to categories of vehicles with similar characteristics resulting in similar emission rates.

TIN – Triangulated Irregular Network.

TMIP - Travel Model Improvement Program, USDOT plan to improve the standard travel demand forecasting modeling capabilities used by cities.

TRANPLAN Travel demand forecasting software produced by the Urban Analysis Group.

TRANSIMS – Transportation Analysis and Simulation System.

Travel demand-forecasting models Models that follow the standard four-step modeling strategy to predict travel behavior based on socioeconomic and infrastructure data.

Unix – Unix is an operating system originally developed in the 1960's and 1970's by scientists at the University of California at Berkeley and at AT&T Bell Laboratories and was designed to be used for running scientific and engineering applications on large processors.

USDOT – United States Department of Transportation.

USEPA – United States Environmental Protection Agency.

UTPS – Urban Transportation Planning System, a travel demand forecasting model, developed in the 1960s.

Vector Topologic spatial data structure (points, lines, polygons).

VIN Vehicle Identification Number, a code number revealing many vehicle characteristics and found on most vehicles.

1. INTRODUCTION

Suburban sprawl, population growth, and auto-dependency have, along with other factors, been linked to air pollution problems in U.S metropolitan areas. Accordingly, the Clean Air Act and other federal legislation and regulations require metropolitan areas to develop strategies for reducing air pollution in those cases where air quality standards are exceeded. An emissions 'budget' is established in these metropolitan areas that provides a benchmark for comparing new emission-generating activity, and presumably not exceeded. Such a goal becomes difficult when trying to accommodate the needs of a *growing* population and economy while simultaneously *lowering* or maintaining levels of ambient pollutants. Growing urban areas must, therefore, continually develop creative strategies to curb increased pollutant production. Because the largest contributor of pollutant emissions in urban areas has most often come from transportation (or mobile) sources, transportation is targeted for new control strategies.

Developing measures of effectiveness and subsequent predictions of overall impact for control strategies require an understanding of the relationship between *observable* transportation system characteristics and emission production. Quantifying this effectiveness requires modeling these relationships. According to published research, motor vehicle emission rates are correlated to a variety of vehicle characteristics (weight, engine size, emission control equipment, etc.), operating modes (idle, cruise, acceleration, and deceleration), and transportation system conditions (road grade, pavement condition, etc.) [Guensler, 1994, Barth, 1996]. Exhaust emissions are produced when a vehicle is started and when it is in operation. Pollutants produced from starting a vehicle can be predicted using vehicle characteristics. Running exhaust emissions additionally require estimates of dynamic engine conditions that result from how the vehicle is driven. Estimating motor vehicle emissions requires the ability to predict or measure these parameters for an entire region at a level of spatial and temporal aggregation fitting the scope of control strategies. Current modeling approaches, however, do not have the capability to provide these estimates.

Today's motor vehicle emission modeling process is based on four separate models: a travel demand forecasting model, a mobile emission model, a photochemical model (for emission inventory), and a microscale model (for analyzing transportation improvements). The travel demand-forecasting model uses characteristics of the transportation system and socioeconomic data to develop

1

estimates of road-specific traffic volumes and average speeds. Mobile emission models use these travel demand estimates, operating fleet model year distributions, and environmental conditions to develop estimates of mobile source pollutant production. These estimates are fed into photochemical models (along with stationary source estimates and data regarding atmospheric conditions) and are used to predict ambient pollutant levels in space and time. These mobile source estimates can also be used by microscale models to predict pollutant levels near specific transportation facilities.

There are several problems with the four-model system that limit effective evaluation of motor vehicle emission control strategies. First, the estimates of vehicle activity (vehicle miles traveled and average speed) lack the accuracy and spatial resolution needed to evaluate control measures [Stopher, 1993]. Second, the mobile source emission rate modeling process uses highly aggregate fleet estimates and average emission rates which are not specific for the fleet in operation, mode of vehicle operation, or grade of the highway facility. As a consequence, the current modeling system has limited capabilities for meeting the modeling requirements of transportation planners. Transportation planners and environmental assessment and control officials have need for improved models that help identify the impacts of standard transportation system improvements (e.g., lane additions, signal timing, peak-hour smoothing).

While many researchers agree that new models and processes need to be developed to overcome these problems, there is disagreement over the best approach [Washington, 1996]. The U.S. Environmental Protection Agency and the Federal Highway Administration held a workshop in Ann Arbor, Michigan in May, 1997 for the purpose of identifying and discussing current emission modeling research efforts [Siwek, 1997]. After the workshop, it was clear that defining appropriate model aggregation levels is important, in defining how and what research should be conducted. A point of departure between the largest vehicle emissions research efforts (University of California at Riverside, and the Georgia Institute of Technology) and the currently mandated approach (MOBILE5a) is the level of aggregation required. Figure 1.1 demonstrates the spectrum of possible approaches. The figure shows that highly aggregate approaches limit explanatory power, but have reduced data intensity. Disaggregate models have the most explanatory power, but the highest data needs. An added dimension to the issue is the fact that estimates must be spatially and temporally resolved, suggesting that an undefined level of spatial and temporal aggregation must also be defined. In fact, the level of spatial and temporal aggregation of mobile source emissions needed by photochemical models may help define the minimum level of model aggregation currently being debated.

This report presents a *research* model that can guide future mobile emissions model development efforts. A major objective of the model is to incorporate the latest transportation / air quality findings at a low level of spatial aggregation (restricted only by data availability). By creating a model under these guidelines, information is developed that leads to the maximum level of disaggregation given user needs and data availability. The research model will be comprehensive, flexible, and user-oriented. It includes enhanced vehicle activity measures; starts, idle, cruise, acceleration, and deceleration. Vehicle technology characteristics (model year, engine size, etc.) and operating conditions (road grade, traffic flow, etc.) are developed at a large scale (small zones and road segments). Flexibility is achieved through a modular design that separates emission production based on thresholds determined in background research. Due to large gaps in the state of knowledge, technology, and practice regarding travel behavior, emission rates, and the urban system inventory, the accuracy of the model results remains unvalidated and therefore unknown. However, the model contributes to transportation and air quality research in that it aids research and software development endeavors.

Emission Rates					
Average Fleet (g/trip)	Vehicle Class Average Speed (g/mi)	Tech Groups Vehicle Mode (g/sec)	Tech Groups Vehicle Mode (g/sec)	Individual Vehicles Vehicle Mode (g/sec)	Individual Vehicles Engine Mode (g/sec)

◄───►

Aggregate				Disaggregate	
Total Trips/Day	Vehicle Class VMT Mean Link Speeds	Tech Group Speed/Accel. Profiles	Tech Group Traffic Flow Simulation	Vehicle Activity Simulation	Vehicle/Engine Activity Simulation

Vehicle Activity

Figure 1.1 - Emission Modeling Spectrum (tech groups refer to sets of vehicles with similar emission characteristics)

The intended model users include emission science experts, model developers, transportation planners, policy makers, and governmental researchers. Each user group has specific modeling interests that define how the model should be designed and presented. Central to the model design is a geographic information system (GIS). Geographic information systems are widely used computer tools that allow geographically referenced data to be organized and manipulated. Both transportation and air quality vary in spatial dimensions. Thus, GISs have the conceptual capability to manage the relationships between transportation activity and resulting air quality

3

changes based on their spatial characteristics. Further, GISs are already used by most planning organizations and government institutions. Thus, a GIS-based emissions modeling framework fits the character of emission science as well as fitting the technical environment of the expected users.

The variables included in the proposed research model are those whose relationship to vehicle activity and emission rates has been defined in research and available to public agencies (see below). They can be categorized as follows:

Spatial Character:

- *US Census block boundaries*
- *Land use boundaries*
- *Traffic analysis zone boundaries (from travel demand forecasting model)*
- *Grid cell boundaries (defined by user)*
- *Road segments (by classification)*
- *Travel demand forecasting network links*
- *Grade school and university locations*

Temporal Character:

- *Hour of the day*

Vehicle Technology:

- *Model year*
- *Engine size*
- *Vehicle weight*
- *Emission control equipment*
- *Fuel injection type*

Modal Activity:

- *Idle*
- *Cruise*
- *Acceleration*
- *Deceleration*

Trip Generation:

- *Home-based work trips*
- *Home-based shopping trips*
- *Home-based university trips*
- *Home-based grade school trips*
- *Home-based other trips*
- *Non-home-based trips*

Road Geometrics:

- *Number of lanes*
- *Grade*

Socioeconomic characteristics (for spatial allocation only):

- *Housing units*
- *Land use (residential, non-residential, and commercial)*

1.1. Summary of Contributions to Research

- *An automobile exhaust emissions model is developed maximizing comprehensiveness, flexibility, and user friendliness.*

Comprehensiveness is accomplished through the inclusion of variables and procedures identified in the literature as significant to emission rate modeling. Flexibility is achieved by organizing the model components by geographic location, and by maintaining a modular program design. User friendliness is achieved by including only current data available to planning agencies, and by using a GIS framework.

- *A research tool is provided that allows for the testing of variable levels of motor vehicle emission model spatial aggregation.*

By having the flexibility to use a variety of spatial entities, the model can become a 'testbed' for determining the spatial resolution needed for future models. This information is valuable in identifying future research needs, costs of emission estimation, model development, maintenance, and operation. A question this model could be used to help answer would be, "Given the current state of research, does a 1 sq. km aggregation of ozone precursors provide enough resolution to predict ozone formation, or would a 4 sq. km aggregation be better?"

- *The benefits of using GIS for emissions modeling are demonstrated.*

GISs provide the ability to organize data by location, in turn providing the capability to develop relationships with new or existing spatial datasets. This allows for the development of creative alternatives to model construction and provides the ability of prioritizing emission control strategies based on location.

5

- *Research and data needs for improved spatial and temporal emissions modeling are identified.*

A study of background research into emissions modeling coupled with an analysis of data available in Atlanta will determine gaps in important emission-specific variables. Further, a prioritization of the data needs based on balancing explanatory power and cost will guide future model development.

1.2. Report Organization

Chapter 1 presents introductory discussion of the research, providing a list of: significant contributions, modeling components, and modeling approach.

Chapter 2 discusses background research significant to automobile exhaust emission modeling, vehicle activity modeling, and geographic information systems. This chapter identifies a research foundation of knowledge that is used to develop model parameters.

Chapter 3 presents a conceptual model design that serves as the foundation of the research approach. Accuracy, comprehensiveness, user needs, and enterprise awareness are important considerations in developing this conceptual model.

Chapter 4 provides a physical model structure that can be used as a research tool. The model will reside in a UNIX operating system and use *Make*, the *C* programming language, and *ARC/INFO*. A step-by-step guide to model use is also provided in this chapter.

Chapters 5 and 6 analyze a model implementation for a 100 sq. km area in Atlanta. Each module of the system is studied using sensitivity analysis, or through comparison of observed data

Chapter 7 will discuss data needs and present final conclusions. An expanded model diagram will demonstrate how future vehicle types and operating modes can be added to the system.

Chapter 8 lists references cited in the report, and Appendix A is a data dictionary.

2.   BACKGROUND

This background chapter will review the key literature related to emissions modeling. Four general areas are reviewed: automobile exhaust emissions, emission rate modeling, motor vehicle activity modeling, and geographic information systems (GIS). The automobile exhaust emission section will focus on the cause and effect relationships of vehicle operation and emission production. The emission rate modeling section will focus on techniques used by different modeling approaches to determine vehicle emission rates. The vehicle activity section will review and identify techniques for developing estimates of emission-specific vehicle activity. The GIS section will discuss issues surrounding spatial and temporal modeling, and review past uses of GIS in the transportation and air quality arena.

2.1. Automobile Exhaust Emissions

This section discusses three topics that are important in motor vehicle exhaust emissions: the major pollutants, the cause and characteristics of their production, and the concept of modal emissions. Understanding these three is crucial to designing a system that is focused on cause and effect relationships.

2.1.1. Exhaust Emission Pollutants

The Clean Air Act of 1970 identified six air pollutants of concern in the United States: carbon monoxide (CO), hydrocarbons (HC), oxides of nitrogen (NOx), sulfur dioxide (SO_2), particulate matter (PM-10), and lead (Pb). Recently, PM-2.5 was added to this list. Nationally, in 1994, on-road vehicles were reported to contribute 62 percent of CO emissions, 42 percent of HC emissions, 32 percent of NOx emissions, ~5 percent of SO_2, 19 percent of PM-10 (PM-2.5 was unreported), and 28 percent of Pb [USEPA, 1995]. Carbon monoxide, hydrocarbons, and oxides of nitrogen are pollutants prevalent in automobile exhaust (PM-10 is produced by diesel engines and tire wear and Pb is being successfully reduced by its elimination from gasoline). For the purposes of this research, the term 'emissions' will hereafter refer to CO, HC, and NOx. All of the pollutants present health dangers to people, animals, and vegetation. Ozone (O_3) is produced through a complex series of chemical reactions that result from pollutants (HC and NOx) mixing in the atmosphere in the presence of sunlight. Generally, ozone concentrations are highest in urban centers and downwind of urban

centers. Ozone has been observed to vary spatially in an urban area, and that the production of ozone is the result of pollutants mixing in space and time. It is also interesting to note that biogenic sources of HC contribute significantly to ozone production. For example, in the southeast United States, eliminating all the anthropogenic (man-made) sources of HC would still not result in passing federally mandated ozone standards due to the levels of HC produced by biogenic (vegetation) sources. [SOS, 1994]. This indicates that a NOx reduction policy would better serve ozone reduction in the southeast [NRC, 1991].

On-road vehicles have been significant contributors to air pollution since the 1940s. The trends in new car emission rates of CO have shown significant improvement over the last thirty years. The improvements have been attributed to legislatively induced emission controls for new vehicles (see section 2.1.2). The actual transportation contribution to overall CO emissions, however, has not declined at the same rate, due in part to the fact that the mobile emission controls are designed to affect only a portion of the engine operating mode and because per capita vehicle miles of travel have increased. In fact, vehicle miles of travel (VMT), auto ownership, person trips, and fraction of single occupant vehicles (SOV) have increased disproportionally to population growth [Johnson, 1993, Meyer, 1997].

2.1.2. The Mechanics of Exhaust Emissions

In ideal combustion, oxygen and fuel (HC) are combusted and produce byproduct emissions of carbon dioxide (CO_2) and water (H_2O). Air, however, contains nitrogen (N_2) among other chemicals, and combustion is always incomplete, producing byproducts of HC, CO, oxygen (O_2), carbon dioxide (CO_2), water (H_2O), and NOx [Heywood, 1988, Jacobs, 1990]. The air to fuel (a/f) ratio is an important factor in determining the quantity of pollutants produced by combustion. Generally, rich fuel mixtures (low a/f ratios) produce high amounts of CO and HC because combustion is incomplete. Lean fuel mixtures (high a/f ratios) will typically produce higher amounts of NOx (especially during very hot, lean conditions) and lower amounts of CO and HC because combustion is more complete. When considering vehicle activity, high power demand (sharp accelerations, heavy loads, etc.) creates a rich fuel mixture resulting in elevated CO and HC emission rates while NOx generally decreases. At high speeds with low acceleration rates, a lean fuel mixture develops which increases NOx emission rates [Heywood, 1988].

Car manufacturers design automobile engines to maximize fuel efficiency and to comply with federal certification tests (Federal Test Procedure (FTP)), which means balancing the a/f ratio (through computerized engine controls) to its most efficient point (stoichiometry). However, car manufacturers also design automobile engines to provide power to meet consumer demand. The certification tests do not cover the high

speeds (maximum speed is 56.7 mph) and high accelerations (maximum acceleration is 3.3 mph^2/sec.) where rich and lean mixtures occur [Barth, 1996]. Therefore, all automobiles are allowed to have inefficient combustion at the high ends of the speed / acceleration spectrum in order to provide drivers with greater power on demand. New test cycles would provide incentives for car manufacturers to reduce the designed enrichment events resulting from power demand. This reduction could significantly lower new car emission rates.

Vehicle technology has changed dramatically over the last thirty years and great strides have been made in reducing emissions. In the 1960s, many vehicles were fitted with devices that controlled the amount of fuel used for combustion, thereby improving the efficiency of combustion and reducing exhaust emissions. In the late 1970s and early 1980s, catalytic converters were installed on new vehicles. Initially, these catalytic converters focused on controlling CO and HC emissions. The catalytic converters treated exhaust gas by removing much of the CO, HC, and NOx emissions [CARB, 1990]. Because there is variability over time (model year) in the types of emission control devices installed on new vehicles, it is probable that vehicle characteristics will play an important role in predicting emission rates, and thus be an important feature in model design for many years to come.

Because emission control technology significantly impacts emissions generation, there are large differences between vehicles with functional control systems, and those with malfunctioning, deteriorated, or nonexistent control systems. The latter group can have significantly higher emissions [Pollack, 1992]. The differences can be pronounced enough that researchers have termed the high emitting vehicles 'high emitters.' Correct representation of high emitters in the vehicle fleet will be crucial to accurate emission modeling efforts given the magnitude of these "above normal" emissions.

2.1.3. Modal Emissions

Modal emissions refers to the types of emissions related to specific modes of operation. Figure 2.1 conceptually represents the relative magnitudes of exhaust emissions for a vehicle trip in space and time. As seen in the diagram, the initial rate of emissions is high, indicating engine start mode. After the engine warms over a period of time, emissions drop and stabilize (hot-stabilized mode). The stabilized rate is interrupted by periods of high emissions (enrichment mode). Each of these three automobile exhaust-operating modes is discussed in the following sections.

9

Engine Rate

Engine Start

Acceleration Enrichment

Grade Enrichment

Hot Stabilized

Engine On

Space and Time

Engine Off
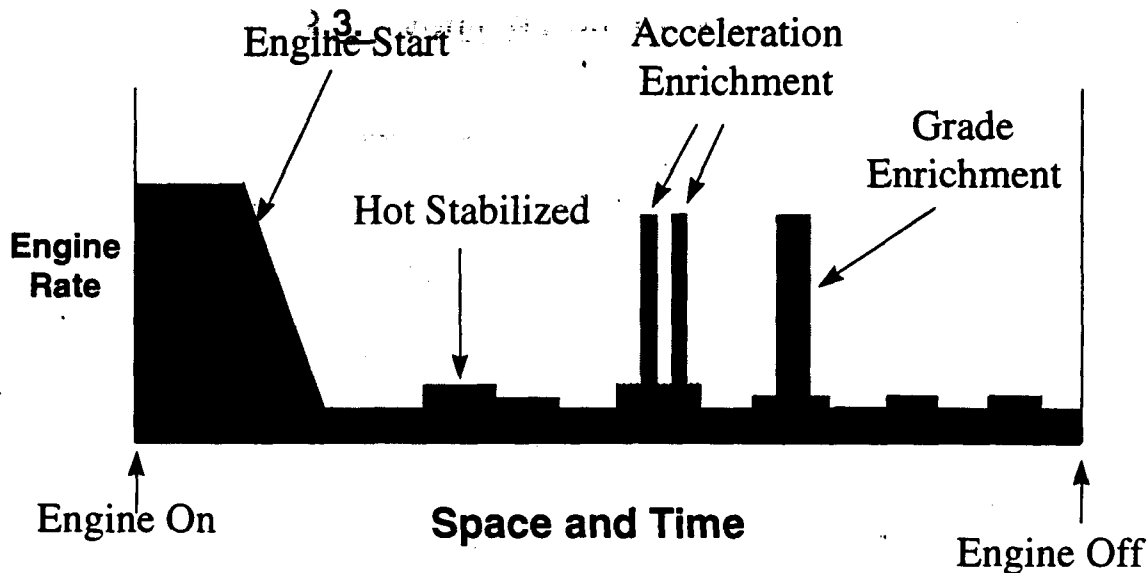
Figure 2.1 - CO Emissions for a Hypothetical Vehicle Trip

2.1.3.1.   Emissions in Start Mode

Motor vehicle emission rates are elevated during the first few minutes of vehicle operation. This is primarily caused by emission control equipment that functions well only at high temperatures. The magnitude of the emissions is a function of: commanded air/fuel ratios, catalyst temperature, and engine temperature (Jacobs, et al., 1990; Heywood, 1988; Joy, 1992; Pozniak, 1980). Most onboard computer control systems initially demand an enriched fuel mixture so the engine will not stall or hesitate during the warm-up period. Thus, the high emissions concentration in the exhaust plume is initially a direct function of the computer control system which varies from vehicle to vehicle. Commanded enrichment may cease when a specific time has passed or when a specific coolant temperature is reached. As engine temperatures rise, combustion efficiency improves and emissions concentrations are gradually reduced. Finally, to be effective, catalytic converters must reach "light-off" temperatures of roughly 300 °C. Until the catalyst reaches this temperature, emission concentrations in the exhaust plume remain high. Catalyst temperature rise is a function of initial catalyst temperature, exhaust gas temperatures, exhaust gas volumes passed through the converter, and emission concentrations. Thus, the magnitude of elevated emi ons associated with engine starts is also a function of the amount of time the vehicle has remained inactive (t affects the catalyst and exhaust gas temperatures), and a function of the manner in which the vehicle is operated after the

10

engine is started (which affects exhaust gas volumes and hydrocarbon loading). Cold starts, engine starts that occur when the engine temperature is below the catalyst light-off threshold, have higher CO and HC emissions.

Two approaches have typically been employed to model engine start emissions: 1) starts are modeled as discrete emission-producing activity, or a "puff," and 2) starts are modeled as a function of a base emission rate (hot-stabilized exhaust) adjusted for conditions that elevate emission rates (Guensler, 1994). The California Air Resources Board's (CARB's) emission rate model (EMFAC7F), for example, treats the elevated engine start emissions as a single "puff" (i.e. separate from running exhaust) and multiplies the number of engine starts by a cold start emission rate. The US Environmental Protection Agency's emission rate model (MOBILE5a), on the other hand, increases the calculated running exhaust emission rate for vehicles, based upon an assumed fraction of vehicles operating in cold start, hot start, and hot stabilized modes. MOBILE5a documentation recommends using 20.6% as the percentage of operating vehicles in cold start mode and 27.3% in hot start mode (based on the FTP analysis). These percentages do not consider location or functional class and were highly correlated to time of day and trip purpose [Venigalla, 1995a].

Historically, the number and location of cold starts have been based on trip generation models (see section 2.3.1) using socioeconomic predictors. Considering emission output, the major factor is not the actual number of starts, but the duration and location of a vehicle operating in start mode. Therefore, a vehicle trip lasting through the start mode will have significantly greater total pollutant production than the few seconds of a false start (an engine start that does not result in a vehicle trip). Research has shown that 180-240 seconds is the approximate average cold start mode duration. In 200 seconds, a vehicle traveling at 35 mph can travel over two miles. A spatially resolved model of start emissions must be able to identify the trip origin and the point on a traveled route where a vehicle moves from elevated emissions in start mode to reduced emissions in hot-stabilized mode. Given that the actual duration of the start mode is not necessarily 200 seconds but a function of a number of engine parameters and conditions, the ability to model on a large scale where the switch in operating modes occurs for a fleet of operating vehicles becomes quite complex. Because trip generation is estimated on a zonal basis, a zonal distribution of engine parameters and conditions may provide enough regional disaggregation and zonal aggregation to identify quantities of pollutants produced. Crucial to success, however, is the size of the zone.

The determination of whether a start is "cold" or "warm" (a warm start occurs when the engine is still warm and therefore closer to catalyst light-off temperature) is also a difficult problem. The duration of the engine soak time (length of time the vehicle is not running) has been used to determine whether a vehicle has a cold or

11

warm engine, thus affecting the duration of elevated emission rates [Sabate, 1994]. Cold starts occur after 4 hours of engine-off activity for non-catalytic converter vehicles, and after 1 hour for catalyst equipped vehicles. Therefore, the parking duration of vehicles indicates how long it will take before the engines warms sufficiently after a start.

The engine "cold" and "warm" start conditions pose a difficult modeling problem. The temporal characteristics of vehicle start activity play an important role in predicting appropriate emission rates. The travel patterns of vehicles also become important. A model including cold and warm start vehicle activity must be spatially and temporally resolved and include predictions of travel behavior and vehicle technology descriptions.

2.1.3.2. Emissions in Hot-stabilized Mode

Hot-stabilized emissions occur after a vehicle's engine has reached sufficient catalyst light-off temperature. When the emission control equipment runs efficiently, emission rates reach a low, fairly stable level. The stabilizing effect also occurs on non-catalyst vehicles due to decreased commanded enrichment, cylinder quenching, and engine oil viscosity. The stabilized emission rates actually fluctuate slightly according to vehicle characteristics, environmental conditions, and vehicle operating modes [Guensler, 1993a]. Vehicle characteristics that have been identified as possibly having explanatory power for a vehicle's emission rate include model year, engine size, accrued mileage, emission control equipment type (such as catalytic converter type) and condition, fuel delivery technology, engine monitoring and control strategies (integrated into the electronic control module), gear shift ratios, and vehicle weight and shape (for aerodynamic drag) [Guensler, 1994, Barth, 1996]. Environmental conditions include ambient temperature, altitude, and humidity [Guensler, 1994, Barth, 1996]. Vehicle operating modes include cruise, acceleration, deceleration, idle, and induced vehicle loads (e.g., number of passengers, trailer towing, grade, and air conditioning) [Guensler, 1994, Barth, 1996]. A vehicle can move in and out of hot-stabilized emission mode when sufficient power is demanded causing a rich air to fuel ratio. When power is demanded causing an enriched fuel condition, emission rates change dramatically (see section 2.1.3.3).

Current models account for some but not all of the factors listed above. Instead, surrogate factors, which are correlated to the factors of interest, are used because they are much easier to obtain for a regional fleet of vehicles. For example, in the EPA MOBILE5a model and the California Air Resources Board EMFAC7F models, the effects of acceleration, deceleration, cruise and idle are currently represented by a single surrogate factor, average operating speed. Average operating speed is correlated with different proportions of vehicle operating modes. Surrogate

12

vehicle attributes include model year, fuel delivery technology, catalytic converter type, accrued mileage, and vehicle condition, and are relatively easy to obtain or estimate for a regional fleet of vehicles from registration and inspection / maintenance databases .

2.1.3.3.   Off-Cycle Exhaust Emissions

Off-cycle emissions are those emission events which occur outside the envelope of the Federal Test Procedure (FTP). The FTP dynamometer test cycle was used as the basis for current model emission rates. Because the FTP cycle did not include vehicle activity with speeds above 57 mph and accelerations greater than 3.3 mph/sec, a certain portion of actual vehicle activity is unrepresented in the test dataset. Activity outside the tested ranges would represent high engine loads and throttle positions that push engines into enrichment conditions. These events are of crucial importance, not just because they aren't included in the analysis of emission rates used for current models, but because these events are known to produce the highest emission rates [Benson, 1989; Groblicki, 1990; Calspan Corp., 1973; Kunselman, et al., 1974]. In fact, one sharp acceleration may cause as much pollution as does the entire remaining trip [Carlock, 1993]. Emissions models may be underpredicting emissions by fairly high margin.

Spatial modeling of off-cycle exhaust emissions requires the ability to predict vehicle speeds and accelerations at a resolution deemed significant by emission rate research. Speeds and accelerations could identify the fraction of the fleet that may be unrepresented in current emission rates. Further, research into the reanalysis of second by second emission test data is discovering substantial amounts of test data outside the FTP envelope [Siwek, 1997]. The reanalysis could predict emission rates based on speed and acceleration characteristics. Further, there is a need to develop emission estimates at a facility level [Venigalla et al.,1995a]. That is, it must be able to predict the locations of enrichment events. If facility-level speed and acceleration profiles can be predicted, emission rates can be applied.

2.2.   Automobile Exhaust Emission Rate Prediction

Three emission rate modeling approaches are discussed in this section; an emission-factor approach, a physical approach, and a statistical approach. Each model type has particular advantages and disadvantages. All of the approaches suffer from two limiting factors:

13

Inadequate Vehicle Test Data. There is a significant amount of emission test data compiled over the years (over 700 vehicles and over 8000 vehicle tests). Most of the testing was done by agencies attempting to determine new car conformity to emission standards. New cars are run through the Federal Test Procedure (FTP) which is a set of three test cycles run on a dynamometer. There is a cold start cycle (bag 1), a running exhaust cycle (bag 2) and a hot start cycle (bag 3). The cycles are called 'bag' data because emissions are collected in a bag during the test. All of the test datasets suffer from at least one of two major limitations, sample size and (or) unrepresentative cycles. The FTP cycle, for example, does not test accelerations above 3.3 mph/sec or speeds above 57.5 mph. Other test cycles that have high speed and acceleration data do not have a representative sample of the on-road fleet.

Inadequate prediction of emission-specific vehicle activity. Emission-specific vehicle activity refers to the division of vehicle operation into groups that differ significantly in their resulting emission rates. All of the approaches require vehicle activity as an input. The best predictor of vehicle activity for metropolitan areas is currently the four-step Urban Transportation Planning System (UTPS) (see section 2.3.1). Although advances have improved the ability of these models to predict emission-specific vehicle activity, most MPOs still use models that have significant errors in facility-level estimates of volume and average speed [Stopher, 1993, Harvey, 1991, Outwater, 1994].

All of the modeling approaches focus on developing emission production estimates, but few present systems are designed to address facility-specific impact issues. This issue is crucial in defining which emission rate model approach best fits the technical capabilities and economic constraints of agencies required to make estimates. In other words, the most accurate model for predicting the emissions of an individual vehicle may not be the most useful for certain types of modeling situations. It may also become evident that the understanding of the causes of an individual vehicle's emission rate has greatly surpassed the ability to collect the input variables for a real-world operating fleet. Important to this issue is the level of aggregation manifested in deterministic or stochastic approaches.

2.2.1. A Speed Correction Factor Approach

Both the USEPA's and the CARB's modeling systems use a 'speed-correction factor' approach to predict aggregate emission rates. The systems are mandated for use in conformity determination despite their widespread statistical and theoretical criticism. The models select a base emission rate depending on a variety of vehicle technology and environmental parameters. The base emission rate is then factored or adjusted based on the ratio of the observed speed to the average FTP cycle bag 2 speed

(1-6 mph). As the models are currently used, the documentation suggests using default values for national fleet averages and other variables. On the positive side, the modeling system is not data intensive, it requires only inputs of total vehicle miles traveled (VMT), average speed, and a cursory knowledge of fuel type and climate data to get estimates of pollutant production. The system is easy to use and widely implemented by agencies without significant capital or operating expense. On the negative side, it is not responsive to changes in important variables (acceleration, fleet makeup, engine load, etc.).

The emission rate models are based on data collected from the FTP cycles developed for new car emission testing. Added to the problems noted earlier with the FTP cycle data, the modeling methodology is highly aggregate and therefore insensitive to microscale variability [Guensler et al., 1993b]. The approach, therefore, may not be able to accurately identify the best choice between small scale development alternatives (changes in lane widths, signal coordination, etc.).

The EPA's Office of Mobile Sources continues to support research which will help to identify incremental improvements to their modeling process. Currently, MOBILE 5a is the mandated emission rate model, and MOBILE 6 is under development. Modal issues, non-FTP cycle estimates, and other emission rate specific factors are planned for implementation.

2.2.2. A Physical Approach

The physical or deterministic approach to emissions modeling is designed to develop accurate emission estimates using many variables. The University of California at Riverside is currently developing such a modal modeling approach under a three year National Cooperative Highway Research Program project. The approach will track the vehicle components and conditions that affect emission rates. The model is designed to track an individual vehicle's power demand and engine equipment status. Power demand is predicted using environmental parameters (wind resistance, road grade, air density, temperature, and altitude), and vehicle parameters (velocity, acceleration, vehicle mass, cross-sectional area, aerodynamics, vehicle accessory load, transmission efficiency, and drive-train efficiency). Power demand is combined with other engine parameters (gear selection, air/fuel ratio, emission control equipment, and temperature) to develop dynamic vehicle or technology group emission rates. When combined with a vehicle's operating parameters, deterioration (the change in emission rate over time due to catalyst decay or equipment malfunction), and fuel type, the model produces highly time resolved emission estimates which promise to be more accurate at the microscale level than any model produced thus far. Vehicle test data for their model are being collected on

15

dynamometers (~300 vehicles) as part of the project. Final test data should be available in two to three years [Barth 1996].

Barth, et al. recognize that their approach is data intensive, but accurate emissions modeling forces it to be so. The vehicle data requirements are many and go beyond the availability of information found in vehicle identification numbers (VINs) that are maintained by state registration datasets. A lookup table could be developed for missing parameters based on vehicle make, model, and model year. Other data (environmental, and operating conditions) would have to be developed from other models. The physical approach fits well with a simulation model of vehicle activity (see section 2.3.2) because the simulations track individual vehicles.

The use of the physical model approach for regional impact modeling requires data aggregation. As with other models, the approach is plagued by poor estimates of emission-specific vehicle activity. Because the physical approach appears to be the most accurate model for predicting an individual vehicle's second-by-second emission rate, vehicle-specific second-by-second activities are needed to get accurate results. Because accurate prediction of these parameters relies on predicting human behavior among other highly variable data, it is likely the activity estimates will have high variability. Aggregating to statistical distributions of the data will lessen this problem, but departs from the original intention of highly accurate second-by-second estimates. The large number of input variables introduce error associated with the ability in predicting their values. The algorithms may be solid, but data input error could significantly degrade the accuracy of the final estimate.

2.2.3. A Statistical Approach

Researchers at Georgia Tech have developed a modeling approach that is based on statistical distributions of a variety of vehicle technologies and vehicle operating modes. The core of the emission rate model is based on hierarchical tree-based regression analysis (HTBR). The tree analysis is a statistical procedure that iteratively splits a dataset into two parts by; (1) selecting a variable that controls the most variability, and (2) determining a cutpoint of that variable that explains the most variability. The result is a 'tree' where each ending node is a set of predictor variable conditions, and an emission rate (for each pollutant and operating mode). Once a 'tree' is developed, adjustments are made to the values based on load (from wind resistance, grade, and accessories).

Georgia Tech researchers combined a variety of emission test datasets from a number of sources in order to maximize the comprehensiveness of the vehicle fleet and potential operating conditions. The data have been re-analyzed to allow modal parameters to be included. Although there are still limitations with the dataset

(representative fleet and cycle operating conditions), an extensive emission rate 'tree' has been developed. The HTBR approach is also plagued with the lack of availability of adequate data input. Extensive vehicle data (model year, engine size, fuel system, emission control, vehicle class, vehicle test weight) and vehicle operating data (speed, acceleration) are needed for predicting emissions. One benefit to the approach is that it can be adjusted for missing data. If one particular variable is missing from the dataset (vehicle test weight, for example, is not stored in the Vehicle Identification Number), the HTBR can be re-run and produce new emission rates that exclude that variable. The new rate may, however, be less accurate, depending on how significant the missing variable is to emission estimation. Another benefit to the statistical approach over current models is the ability to put confidence bounds around each estimate. This becomes important when estimates for a variety of conditions on a certain facility segment are added together to produce a single facility estimate, whose accuracy must be quantified.

Critics of this modeling approach have suggested that the inability to track causal variables results in a model that is unable to predict the effects of new technology. There are three counter-arguments to this criticism, (1) because control standards continue to tighten, it is more important to model the old technology instead of the new, (2) no model can expect to accurately predict future technology changes, they can only develop relationships based on known conditions, and (3) if surrogate variables are correlated to casual ones, the model will still continue to work.

2.2.4. Emission Rate Modeling Summary

The microscopic physical approach taken by Barth et al. has the potential to provide the most explanatory power, disregarding input data error issues that can't be quantified at this time. It is also clear from the research that the speed correction factor approach is highly aggregate and inappropriate for the modeling needs of research and planning agencies. The statistical approach provides near-term improvements and allows for facility level aggregations of data. An important factor in selecting a particular emission rate modeling approach is its ability to fit within the framework of the larger 'data model' issues regarding the user needs of measuring and predicting transportation impacts on air quality. The 'data model' in this context refers to the design of an entire modeling system from user needs to data structure and connectivity. The statistical approach seems most appropriate given the scope of this research because it appears to fit the balance between accuracy and implementability identified as a modeling objective in Chapter 1.

17

2.3. Vehicle Activity Modeling

2.3.1. Urban Transportation Planning System (UTPS)

The Urban Transportation Planning System, (or travel demand forecasting model), first developed in the 1960s, was designed to predict travel flows within an urban area. The primary purpose of the system was to guide new infrastructure investment [Outwater, 1994]. Because of its predictive nature and widespread use, the use of this modeling approach has expanded beyond the original intent to predicting emission-specific vehicle activity. Until recently, vehicle activity has meant vehicle miles traveled (VMT) and average speed, the inputs to mandated emission models. However, as understanding of emission behavior expands, so does the definition of vehicle activity. Emission-specific vehicle activity now encompasses detailed modal parameters which UTPS models are incapable of predicting. Researchers have identified numerous deficiencies in the approach (outside implementation problems); the facility level (link) estimates are highly variable, the models do not predict off-peak travel well, seasonal variations in travel are not considered, model size is limited, and the models are not sensitive enough to measure mandated TCM effectiveness [Stopher, 1993, Harvey, 1991, Outwater, 1994].

Along with theoretical problems, there have been a significant number of implementation problems including: lack of feedback components, insufficient current socioeconomic data and inadequate validation procedures [Harvey, 1991, Outwater, 1994]. Model results have indicated an accuracy range of 5-30% error in overall VMT estimates and 5-20 mph error in average speeds. [Miller, 1995]. Average error by models implemented by MPOs is 10% for VMT and 15 mph for average speed [Stopher, 1993]. Errors also increase as one moves from higher to lower road classifications. To add to the problems, the same models that are criticized as too simplistic are too complicated and costly for proper implementation by many agencies.

Despite these errors and theoretical deficiencies, the models represent the state-of-the-practice. In fact, they represent the only short and medium range alternative available for widespread implementation. There is a significant amount of research on techniques for improving the UTPS and hopefully improvements will result in better predictions of vehicle activity in time and space.

The Travel Model Improvement Program administered by the US Department of Transportation is attempting to improve the travel forecasting capabilities. Some of the potential improvements to predicting emission-specific vehicle activity are as follows. (1) There is a shift away from trip-based mode towards activity-based models. Activity-based models better represent temporal changes and mode alternatives. (2) Development of stochastic microsimulation techniques aggregated to

area traffic patterns will allow improved sensitivity to temporal changes. (3) The use of longitudinal panel surveys will more accurately identify cross-sectional survey (current technique) biases.

2.3.2. Simulation Models (TRANSIMS)

Simulation models are being viewed by many as the solution to the problems facing the UTPS. Simulation models generally come in three forms, microscopic, mesoscopic, and macroscopic. Microscopic models track individual vehicles and their relationships with other vehicles. Macroscopic models approximate traffic flow as a fluid and use a facility (road segment) as a the base unit. Mesoscopic models combine elements of both depending on the needed function. Simulation models have successfully been used for optimization (signal timing, traffic flow) and for forecasting (predicting results of a change). Models can be deterministic or stochastic (by allowing some randomness into the process). By their nature, simulation models have the theoretical and computational capability to predict facility-level activity at a resolution needed to predict emission-specific activity. The structure and data requirements of existing models have prevented their implementation for an entire urban structure, and force use at the facility level. Most models have been developed to answer specific problems instead of complete system simulation. However, a new generation of simulation models is taking a broader scope and the models are being designed around regional systems instead of specific traffic-flow issues. Recent advances in modeling theory, microscopic modeling, and computing power may have expanded the role of traffic micro-simulation modeling from the facility scale to the urban/regional scale.

Advances made by "TRANSIMS" (Transportation Analysis and SIMulation System) have led many to believe that they have found a replacement for the UTPS type models. TRANSIMS is being developed under the US Department of Transportation's Travel Model Improvement Program and funded by the Federal Highway Administration and other federal agencies. The intent of the project is to develop a system that will be able to answer questions regarding policy and infrastructure change for an entire urban area. One of their major selling points is their focus on predicting air quality and other environmental impacts.

TRANSIMS will be a set of modules that can be run separately or together. The first module is a household and commercial activity module that uses US Census data to develop a synthetic population of individuals for Census Tracts and Block Groups and predicts synthetic economic activity and resulting travel demand. The second module is the intermodal route planner that takes the activity-based travel demand and develops trip plans for every individual that can be adjusted depending on the activities of other individuals over time. The third module is the travel

19

microsimulation module that tracks individuals and their vehicles, and their relationships to other vehicle activities, on the road network using a 'cellular automata' technique. The final module is the environmental module that predicts a variety of environmental conditions including mobile source emission prediction, atmospheric mixing, and concentrations. The outputs of TRANSIMS will be summaries of second by second data at cells of 7.5 meters.

TRANSIMS promises to provide unique solutions to the integration of macroscopic and microscopic transportation modeling and provide advances in a number of simulation issues. Issues that the developers must address are validation and implementation. All of the new algorithms and techniques must be individually validated against observed data. The time frame and cost of implementation at a new urban area may be extensive due to the input data requirements.

Despite a number of issues that must be addressed by TRANSIMS developers, it is apparent that the spatial and temporal resolution of emission-specific vehicle activity could be substantially improved by TRANSIMS in the future. This aspect identifies the emission modeling need for incremental research that builds towards a future system that can move toward the objectives defined by the Los Alamos researchers.

2.4. Geographic Information Systems

A geographic information system (GIS) is "a computer-based information system that enables capture, modeling, manipulation, retrieval, analysis and presentation of geographically referenced data" [Worboys, 1995]. The rise of GIS technology and its use in a wide range of disciplines provides transportation and air quality modelers with a powerful tool for developing new analy⋅⋅ capability. The organization of data by location allows data from a variety of sources to be easily combined in a uniform framework. For example, vehicle registration information can be combined with census data to develop driver-vehicle profiles. Or, high traffic volume areas can be combined with satellite analysis of vegetation decay to study environmental impacts. Another important feature of GIS is its ability to bridge the technical gap between analysts' and decision-makers' need for easy-to-understand information . The communication power of GIS (thematic maps, GUIs, 3-D surface plots, etc.) is a feature th⋅ as made GIS one of the most used platforms for planning in the U.S. GIS provide⋅ the ability to get quick answers to technical questions. Literature on GIS data structures, applications, and vendor products is substantial. The following section will briefly cover, 1) the extent of GIS implementation by

transportation and air quality agencies, and the past use of GIS in transportation and air quality analysis, and 2) the issue of spatial data quality.

2.4.1. GIS in the Transportation / Air Quality Agencies

The National Cooperative Highway Research Program (NCHRP) *Report 359* studied GIS in an effort to define its potential for transportation agencies. The document, which presented a comprehensive overview of GIS technology, its potential role for serving the needs of a variety of agencies, and strategies for successful implementation, stated,

" The potential impact of GIS-T is profound. If this technology is exploited to its fullest, it will become ubiquitous throughout all transportation agencies and will become an integral part of their everyday information processing environments. ... The potential impact of GIS is more than just agency wide. The problems of today require the interaction of agencies at all levels of government. ... the broad problems that are driving the interaction typically involve environmental and economic development issues; and their solutions will require the integration and analysis of geographically referenced data of many kinds from many sources."

2.4.2. Applications of GIS in Mobile Emission Modeling

2.4.2.1. Emission Inventories

Bruckman et al. presented a paper in 1991 at an Air and Waste Management Association conference describing the use of GIS in developing gridded, hourly estimates of emissions. They also developed a model called CAL-MoVEM that utilized GIS in developing mobile source estimates for input into photochemical models. The main function of the GIS in their model was the spatial aggregation of travel demand forecasting model features into a grid. They used spatially defined vehicle mixes by trip purpose, temporal factors, hourly temperatures, trip volumes, trip speeds, and modal percentages as inputs. The spatially defined inputs were combined with EMFAC7E emission rates to produce gridded hourly estimates of pollutants [Bruckman, 1991]. The work was accomplished as part of a study on ozone levels in the San Joaquin Valley in California. Zonal estimates were allocated to TAZ (traffic analysis zone) centroids that were re-allocated to grid cells. Link estimates were allocated to nodes and re-allocated to cells. The use of points to represent these features did not take full advantage of the spatial structure provided by the original input data. TAZs falling along grid cell boundaries should have their portions divided. This strategy would limit grid cell sizes to those significantly larger than TAZs, which

21

can be quite large (30-40 square km) for some metropolitan areas. Also, no mention is was made of strategies for identifying the confidence ranges of the estimates.

The model supports the use of GIS, but did not take full advantage of the research value of GIS. Further, the model did not have the flexibility to answer the diverse impact or mitigation questions that arise from estimating emissions.

2.4.2.2. GIS for Transportation Planning and Air Quality Analysis

Researchers used GIS as a preprocessor and postprocessor to mobile emission modeling. Although they relied on existing models to estimate emissions, they showed how GIS could be valuable in the management of emission related data. They made the connection between the needs of transportation planners and decision-makers and the spatial tools and features of GIS [Souleyrette 1991].

2.4.2.3. Microscale Analysis

Researchers at Utah State University used GIS in developing microscale analyses of a small group of intersections. They linked a GIS with CALINE3 and CAL3QHC to predict pollutant concentration levels [Hallmark, 1996]. The value of GIS (outside of spatial data storage and data visualization) was its ability to compare concentration results to other non-related data. The contribution is significant to this research because it provides a foundation for the argument that a GIS approach is not restricted to developing emission inventories, but can be easily expanded to a number of other related issues.

2.4.2.4. Influencing Decision-makers

Othofer et al. developed an interesting approach to predicting location specific emission production estimates for changing control strategies. Instead of developing estimates using detailed location-specific emission producing activities and emission rates, they disaggregated large zonal estimates using emission-producing activities [Orthofer, 1995]. The advantage of this approach is its simplicity and its straightforward recognition that the data needed to predict emissions at smaller levels does not exist or the relationships are undefined. The disadvantage is that the ability to predict changes among the disaggregated levels is a function only of the change of the overall larger units. Thus, the true effects of activity changes on emissions cannot be measured. The proj produced high-quality graphics that indicated locational variation in emission-proc..cing activities. The project was successful because elected officials could 'see' areas that have potentially high emissions and therefore had evidence for developing actions for those specific areas. Although, the modeling

capability of the project is limited, its ability to influence action through spatial communication is a noteworthy contribution to the use of GIS in this arena.

2.4.3. Spatial Data Issues

Spatial data refers to points, lines, or polygons that maintain a digital connectivity with other entities in regards to their relative position. Spatial data comes in two forms, raster or vector. Raster data is information in a regular unit, usually a grid cell. The grid cell maintains an attribute value and locational information pertaining to its place in a matrix. Raster data is preferred when representing continuous data (natural features, environmental features, air quality, etc.) or when developing complex spatial data models. Vector data is information in the form of points, lines, or polygons. Vector data better represents features with discrete edges (anthropogenic features, rivers, transportation, etc.). An issue of prime importance to both data structures, and for this research, is spatial data quality. The quality refers to a number of issues regarding the accuracy and resolution which are discussed in the next two sections.

2.4.3.1. Positional Accuracy

Positional accuracy refers to the variability of the represented position from the actual position. Relative positional accuracy refers to the relational position between represented features and absolute position accuracy refers to the relationship between represented features and the Earth's surface. A good relative accuracy and poor absolute accuracy indicate a positional problem that is important when bringing different databases together. Because of the development of US National Map Accuracy Standards and the advent of improved surveying techniques, relative positional accuracy within a single dataset is not significant given the scope of inventory modeling. Absolute positional accuracy becomes an issue when joining multiple spatial databases. Variations in position can result from using different projections, datums, or transformations. Any attempt to join spatial databases must address the issue and provide solutions (stretching, fuzzy tolerance, etc.) to reduce the impacts.

2.4.3.2. Data Resolution

Data resolution concerns the level of spatial aggregation, or density of observed values. Data resolution usually refers to the scale at which the original data observations were made, and the level of interpolation used in developing the final dataset. The level of resolution is important in determining the confidence of a represented value at a particular coordinate. As in positional accuracy, data resolution problems usually occur when trying to combine databases of varying resolution. The

23

combination of two datasets will result in a dataset that has a resolution equivalent to the one with less detail. This is frequently overlooked in analysis resulting in the presentation of data with significant variance. For example, soils data at a scale of 1:24,000 can be overlaid with 1:100 parcel data in an attempt to identify the parcel's soil type. The result represents the 'best guess' as to soil type variations within a parcel, but the variability is high. It is good practice to question whether the scale of database fits the spatial character of what is being represented. For example, does a 1 km or 4 km aggregation of ozone precursor pollutants provide enough resolution given the scale of ozone formation?

2.4.3.3. Data Content Accuracy

Data content accuracy refers to the accuracy of the attribute data represented by the spatial feature. Data content accuracy can be limited by a number of procedural problems (coding error, measurement error, etc.) or by the change of the data over time. Data content can be estimated by validation techniques, but they are usually cost-prohibitive for the large spatial datasets available. Usually spatial databases can be tracked to an original collection technique that may have been validated. It is also possible to compare two or more datasets for agreement to develop a qualitative appraisal. Most publicly available datasets have quantitative information on the accuracy of their data content.

3. MODEL CONCEPTUAL DESIGN

This chapter presents a conceptual design of the GIS-based automobile exhaust emissions model. The background research from the previous chapter is summarized into a series of 'research foundation points' that define modeling parameters. User requirements are also identified, guiding model form and presentation. By the end of the chapter, a modeling approach is recommended.

While the overall purpose of the model is defined in Chapter 1, more specific model objectives that guide the development of such a model include:

- *The model must produce automobile exhaust emission estimates that are capable of being statistically verified.*

It is vital for the model to be able to determine errors in estimates that result from input data error and algorithm error. One of the biggest criticisms of the currently mandated modeling approach is that there is no information available for users to estimate errors resulting from the algorithms. A design open to outside review and analysis prevents avoidable extrapolation because the confidence intervals are known.

- *All estimates and input parameters (emissions, vehicle activity, etc.) must be capable of being validated.*

All model components must be capable of being validated either through previously published research or through designed experiments. Given the complicated process of predicting emissions, it is important that all intermediate modeling steps be designed to be tested. This objective will influence the data model because many elements regarding vehicle technology and vehicle activity will have to have identifiable characteristics that can be observed in the field.

- *The model must be designed to easily incorporate new findings.*

Because research into emissions modeling is occurring in a number of institutions, significant findings are expected in the near future. Keeping the research model up to date to research from other institutions is crucial if it is going to be used to influence research direction and software development decisions.

- *The model must use available data.*

Although the model is not designed to be implemented on a widescale basis for official reporting, it must still be constrained by real-world conditions of data

availability and cost. Without considering these factors, one can spend a significant amount of time and resources developing models from variables (dynamic engine parameters, etc.) that cannot be collected by a regional modeling agency.

- *The model must use as large a spatial scale as data will allow.*

It is important to use available data, but it is also important to use the largest scale possible. One of the uses of the model will be to identify the level of spatial aggregation required for useful emission estimation. In order to do this, it is important to start with the most detail and aggregate up, thereby identifying locations with high emission production.

- *The model must be portable.*

The model should be transferable to other urban areas without substantial model alteration. This means that all the input parameters should be available to major metropolitan areas, and that model assumptions should not be limited to the study area.

3.1. Model Design Parameters

The following model design parameters are based on material discussed in the background chapter, user requirements, and good modeling practice. These parameters will establish minimally acceptable guidelines for model development. The ability of the model design to abide by the parameters will depend on the data and technology available to a clearly-defined user group. Some parameters may have to be scaled back due to limitations in data availability.

This summary of the background knowledge is presented in this section to clearly identify model development parameters. The initial goal of model development is to include all listed parameters. At some point limited data availability or excessive data development expense will likely remove or scale back some parameters from consideration. The research backed parameters are:

- *Develop estimates of the production of automobile exhaust pollutants CO, HC, and NOx in space and time (from section 2.1.1)*

Research has shown that the major exhaust pollutants of concern are CO, HC, and NOx. Considerations should be given to including particulate matter greater than 2.5 microns in diameter (PM2.5) due to its recent identification as a health risk. However, there is very little data on the cause and effect relationships of PM2.5 production by automobiles. The emission estimates represent only the production of pollutants, not the resulting air quality. The spatial and temporal scale should be

developed according to anticipated user needs. Existing photochemical models (used to predict ambient air quality) currently use hourly, 4-5 sq. km aggregations. Future photochemical model improvements are expected to use 1 sq. km estimates of mobile sources.

- *Anthropogenic NOx estimate accuracy important in predicting ground-level ozone (from section 2.1.1)*

Major cities in warmer climates have air quality problems resulting from ground level ozone concentrations. NOx and HC are precursors to ozone formation. HC, however, can be produced in significant amounts by biogenic sources. Therefore, a more accurate, verifiable, estimate of NOx may prove more useful in predicting the impact of motor vehicles.

- *Comprehensive representation of vehicle technologies (from section 2.1.2)*

Differences in vehicle technologies / characteristics have been shown to significantly affect vehicle emission rates. As seen in the physical model approach by Barth, et al., it is actually the dynamic status of a number of vehicle parameters that causes emission rate variability (see section 2.2.2). At the same time, a number of vehicle characteristics have been tied to emission rate variability because they are surrogate variables for causal parameters (see section 2.2.3). From a research model perspective, it is important to be able to include both sets of conditions. However, modeling individual vehicle engine dynamics for an urban area is not practical due to extensive data requirements. Instead, only those specific static inventory variables involved with the dynamic conditions, and those variables identified as surrogate variables are included. The list of desired vehicle characteristics are: model year, engine size, weight (or mass), emission control type(s), fuel delivery type, transmission type, cross-sectional area, and number of cylinders.

- *Separate and quantify high-emitting vehicle emissions (from section 2.1.2)*

A small percentage of the fleet disproportionally contributes to total mobile source emissions. By separating this small high-emitting portion of the operating fleet, it will be easier to predict the impacts of control strategies that may target high emitters. Further, model attention should be focused on factors that result in higher emissions, wisely using resources in the most important areas.

- *Separate start, hot-stabilized, and enrichment emission quantities and locations (from section 2.1.3)*

By separating estimates into specific emission modes, mode-specific impact strategies can be more efficiently evaluated. Further, emission rates for each mode are predicted using different variables. Engine starts are primarily influenced by vehicle characteristics and engine temperature. Hot-stabilized and enrichment emissions are

primarily influenced by vehicle characteristics and operating condition (speed, acceleration, etc.).

- *Include Speed Correction Factor (SCF) emission rates (from section 2.2.1)*

The inclusion of SCF emission rates provides an alternative modeling approach. One of the objectives of this research is to be comprehensive and flexible. Inclusion of the SCF estimate provides a flexible framework and a way to compare between emission rate modeling approaches. The model may indicate that the highly aggregate SCF approach is suitable for regional inventory modeling at a certain spatial level. The SCF approach to modeling start emissions will not be included because the approach lacks the ability to show spatial variability between start and running emissions.

- *Include emission rates from the statistical approach (from section 2.2.3)*

Emission rates from the statistical approach need to be included because the research indicates that modal parameters better characterize accurate emission rate estimation. Because the modal emission rates models are available, they can be immediately integrated into the model framework. The approach also produces separate start and running exhaust emission estimates, addressing one of the previously defined model design parameters.

- *Include activity measures from travel demand forecasting models (from section 2.3.1)*

Travel demand forecasting models are the primary predictive tool for regional level vehicle activity. They are also used by almost every transportation planning agency (MPO) in the country. Further, their use in developing emission estimates is currently mandated. Despite their well-documented problems, they have characteristics that make them very attractive for a spatially-resolved model. First of all, they have a defined structure and connectivity that translates into a spatial form (zones, links, and nodes). Second, they develop estimates using socioeconomic information, allowing the model to be indirectly affected by social and economic changes.

- *Prepare for inputs from future simulation models (from section 2.3.2)*

Simulation models provide vehicle activity measures at a larger spatial scale and resolution than macroscopic travel demand models. The value of producing estimates at the microscopic or mesoscopic scale becomes evident when studying the types of vehicle activity that produce high emissions. Just as high emitters disproportionally contribute to total emissions, so do high power demand situations. These driving situations can be characterized by comprehensive representation of traffic flow dynamics.

- *Use geographic information systems (from section 2.4)*

Using GIS is important because it is designed to handle the spatial data management and modeling functions key to the research goals. Without GIS, complex spatial analysis and manipulation algorithms would have to be re-created. Its widespread use and popularity among planning agencies is significant enough to warrant its use.

3.2. User Requirements

In designing any analysis model, it is crucial to clearly understand the analysis needs of the proposed users. There are several user groups that could be expected to interact with the research model.

- *Emission Science Experts*

These experts are those individuals who help define the emission science domain. They provide the knowledge regarding the cause and effect relationships in automobile emissions modeling. Although their interaction with model design is conceptual, it is tremendously beneficial if they can interact with specific model components to ensure that the science is being accurately represented. Therefore, one data model requirement is that the system be composed of well-documented and appropriately termed modules that can be easily reviewed by the specific component's experts. The model vocabulary should be defined by the experts' terminology (i.e., transportation components use standard traffic engineering terminology).

- *Model Developers*

Model developers can also have significant knowledge of the cause and effect relationships among key variables. If a model requires significant software development, it would be prudent to organize it using standard programming techniques, terminology, and comments. This may require that comments in the code explain the underlying scientific concepts to the point that clear understanding of the importance of the various pieces is evident to developers. If a developer could improve program efficiency by slightly altering a process, it would be beneficial that the explanatory cost of the change be evident. Therefore, well-documented code is a specific data model requirement.

- *Emission Researchers*

Emission researchers are individuals who use the model to get a better understanding of the impacts of new findings, or develop criteria for future research efforts. This adds a dimension to the model by requiring that measures of confidence be included with the estimates. The estimates produced by the model must be capable of being accompanied with certain measures of accuracy and descriptions that clearly

identify what is known and unknown in the process. The model inputs and outputs should be in a format that allows easy import/export to various software packages that may be used for more detailed analysis. Outputs must include detailed summaries of assumptions and discussions of accuracy to prevent false conclusions from being drawn.

- *Government Experts*

    Government experts would be individuals who would look at all levels of model development to ensure quality and accuracy in order to approve or disapprove results that could hold legal bearing. If model results are to be used for conformity or inventory reporting, the model elements must be validated and peer-reviewed. This is important in a developing model because government experts and researchers must be included in the design process to prevent efforts from moving in directions that contradict policies and mandates that govern air quality modeling. This user requirement strengthens the need for modular, clearly communicated model code.

- *Transportation and Environmental Planners*

    Transportation and environmental planners are the eventual 'users' of the system. They will be the ones who develop the emission estimates for their particular project. Although the level of development discussed in this report is for a 'research-grade' model not to be used for legal reporting, the intention is that the model or some of its components eventually be targeted for widespread public use. By including the eventual user needs in the early design, complications in future development can be avoided. By including transportation and environmental professionals, who may or may not have model development experience, the system design becomes intuitive and flexible. Planners should not be burdened with extensive command and syntax requirements. Results should be designed towards the reporting needs of the planners.

- *Non-technical Decision-Makers*

    Decision-makers (policy-makers, managers, planning boards, etc.) need model results to make informed decisions. Decisions range from guiding the direction of research to broad-based policy analysis and to local transportation alternative analyses. Many of these users are removed from the modeling process, but must be familiar with the process of modeling so they can have confidence in the model results and be aware of assumptions made by modelers. By maintaining the model framework within an off-the-shelf GIS, questions about model inputs and outputs can be asked and answered by non-technical users. Further, thematic maps and user-def·..d spatial queries, graphical results can be produced along with standard spreadsheet and textual reports.

    In summary, the user-defined needs include:

- *Appropriate documentation*

30

- *Appropriate terminology*
- *Modular system design*
- *Open input and output data formats*
- *Intuitive modeling process*
- *Easy to understand and use*
- *Model should reside in a GIS*

3.3. The Spatial Data Model

A good data model has five dimensions [Reingruber, et al., 1994]: conceptual correctness, conceptual completeness, syntactic correctness, syntactic completeness, and enterprise awareness. *Conceptual correctness* refers to the degree to which the model represents the real world, or the accuracy of the model estimates. For this model, it refers to the accurate representation of the cause and effect relationship between motor vehicle behavior and emissions. *Conceptual completeness* refers to the wholeness of the represented science. In this model, it refers to the ability to represent the cause and effect relationships in a comprehensive manner for the entire urban area. *Syntactic correctness and completeness* refer to the quality of the use of language and proper communication. This would involve the use of organized and structured programming techniques as well as the use of accepted transportation, air quality, and GIS terminology. *Enterprise awareness* refers to the idea that the model does not work in a vacuum, but that it represents only a portion of a much larger system. An automobile exhaust model is a portion of the much larger scope of environmental and transportation modeling. Keeping the model open to connectivity with other systems ensures an adaptable and open system.

As the design of model components develops, it will be analyzed in respect to these five concepts, paying particular attention to completeness and correctness. These measures will guarantee that the design will have an organized framework.

Spatial data entities are the spatial forms used to characterize an object. For example, a road 'object' can be characterized by the digital representation of a line, the spatial data entity. Generally, the best entity type to use for spatial modeling or environmental data representation is a raster cell. A raster cell structure handles continuous variables better than a vector (points, lines, and polygons) structure. This occurs because regular grid cells that fall between observed values can have statistically interpolated values. A vector representation forces observed values to be discrete within its structure, possibly misrepresenting boundaries. Another possible spatial data entity is a triangulated irregular network (TIN). TINs interpolate points found on a line between two observation points based on the values of the points. It is

31

most often used in representing topography, but the concept could be translated to other areas. The selection of an appropriate spatial structure is controlled by the model objectives, model parameters, and user needs.

A vector approach has the following advantages:

- *Intermediate estimates must be validated*

Because the model being developed is research-oriented, all of the algorithms and data must be represented in a format that can be field validated. A clearly defined beginning and ending point for a segment of road (intersection to intersection), or a field-evident boundary (zone bounded by roads) makes validation simpler.

- *Users require facility-level estimates*

Transportation modelers (eventual users of the model) work with vector-based facility entities (see section 3.4.1). By providing and receiving data in a similar format, the integration and transfer of data is more efficient.

- *Pollutant production is discrete*

Vehicles are discrete objects. Because the model predicts pollutant production, not resulting air quality, the emission estimate should also be discrete. Given that the model will not actually model individual vehicles, there is an argument that an aggregate characterization of factors is more efficiently handled in a raster approach. However, the appropriate level of aggregation is undefined; in fact, one of the model objectives is to provide a way to determine the appropriate level of aggregation. Once it is determined, a raster approach for software development may be warranted.

A raster approach to the model also has benefits associated with its use in the research model:

- *Inventory estimates are gridded*

The final outputs of mobile emissions models are inputs into photochemical models. The photochemical models are raster due to the nature of the phenomenon being modeled (ambient air quality). Gridded, hourly estimates are currently required.

- *Data will have to be aggregated eventually*

It is unlikely that widespread data availability allows modeling on a vehicle per vehicle basis, nor is it likely that modeling at that level is practical or useful for regional scale modeling. Some level of aggregation will have to be used. In that regard, it may become more appropriate to predict continuous distributions rather than discrete polygon values.

Overall, if the user is concerned with the location of high auto emissions, a raster approach would be better, although technically questionable for linear features

such as roads. If the user is concerned with the emissions on a specific entity (i.e., road segment, TAZ), a vector approach would be better. Given that the specified users are concerned with both issues, both raster and vector entities should be used. The issues of validation and emission science suggest that initially, vector data models are warranted. At some point, the vector structure needs to be converted to raster for further photochemical modeling, and regional data visualization.

3.4. Model Approach

The conceptual design of the proposed research model meets all of the stated goals and objectives that have been stated earlier while avoiding the constraints of extensive data development and cost. The model will be deterministic and spatially and temporally identify: the types of vehicles that are being operated, the types of activities the vehicles are involved in, the resulting emission rates, and the resulting emissions. The level of aggregation and spatial scale is flexible, depending upon the user's needs, data availability, and accuracy requirements of post-processing. Regardless of the spatial scale, the conceptual design remains the same. Figure 3.1 shows a schematic drawing of the design. The top row represents the spatial environment. The second represents vehicle characteristic assessment. The third represents the vehicle activity. The fourth and fifth rows represent facility-level and gridded emission estimation. Detailed information about each component is provided in chapter 4.

Central to the model design is the identification of the source of vehicle activity data. While travel demand forecasting models have significant limitations providing inputs to emissions modeling, they represent the only widely used prognostic planning tool available. Until regional micro-simulation models become widely accepted, validated, and implemented, emission models must rely on the forecasting capabilities of the tools in use. With all of its disadvantages, there are components of the travel demand forecasting model that defend its use for emissions modeling. First of all, trip generation results can be easily translated to engine starts, an important emission activity. Second, poor estimates of average speed can be supplemented with observed speed and acceleration data given certain traffic flow parameters. Third, the travel models have spatial characteristics that can form a foundation for spatial modeling. The research model is tied to traditional travel demand models. At some point in the future, regional simulation models may become the primary source of travel behavior prediction. This should be reflected in the model design by avoiding strategies that lock into specific travel model types.
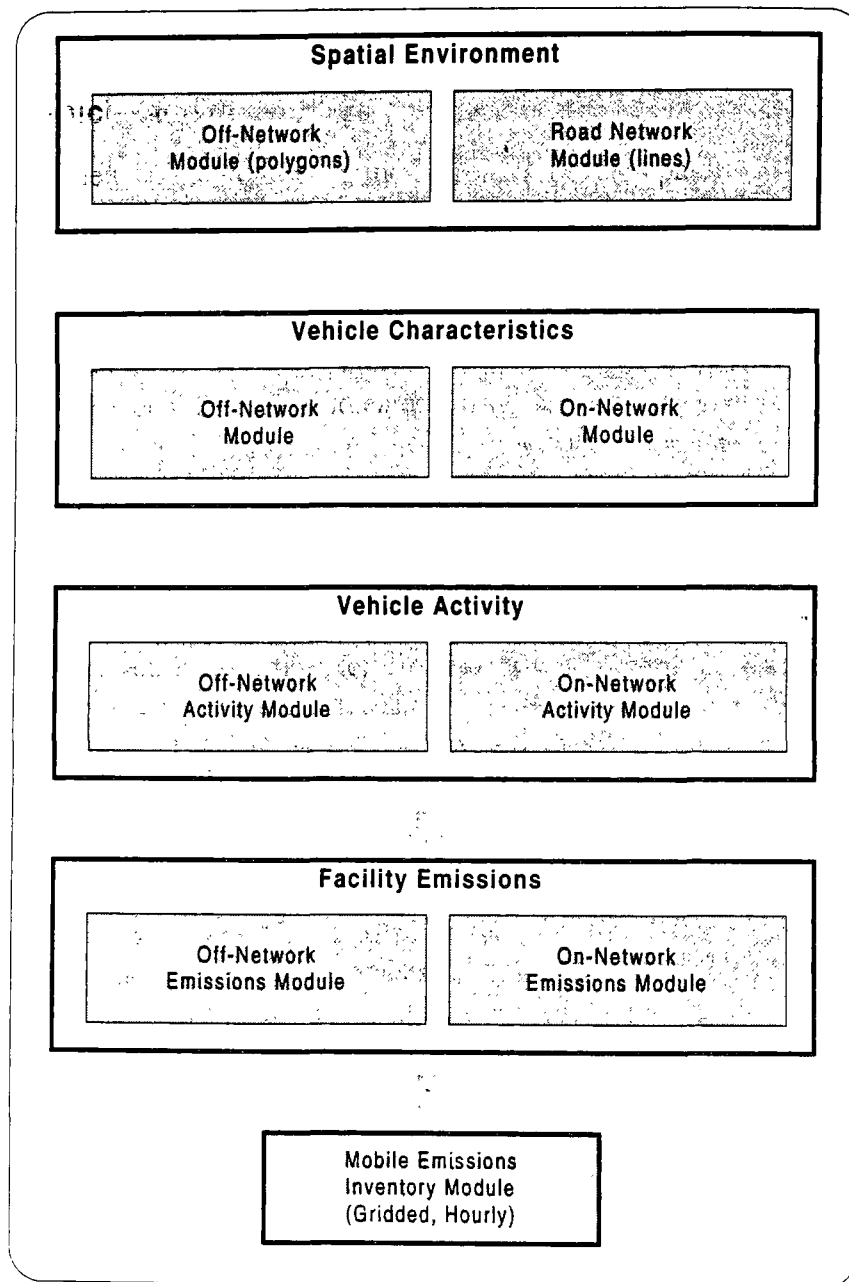
Figure 3.1 - Conceptual Model Design

The following sections describe the five major tiers of the model design. In each section, descriptions of the roles, data needs, and processes are provided. Three of the five parts of a good data model are discussed for each major component: conceptual correctness, conceptual completeness, and enterprise awareness (the other two parts are related to syntax and deemed less significant). Each section is supplemented with specific model descriptions in chapter 4.

3.4.1. Spatial Environment

The objective of the spatial environment tier is to unify input data under a common zonal and lineal structure. The size and scope of the zones and lines depend on the users and their specific needs. Historically, exhaust emissions are divided into start and non-start (running exhaust) emissions. Most prognostic travel models provide a zonal (TAZ) estimate of the number of trip origins, and a lineal (link) estimate of road volume and average speed. By defining an engine start as being synonymous with a trip origin, TAZs become the base spatial entity used for engine start emissions. Running exhaust emissions occur on the road network, suggesting that a 'link' should be the base spatial entity. Improvements in the spatial resolution of the zonal estimates can be made outside the travel model by disaggregating trips to smaller zones. The lineal estimate can similarly be improved by conflating (see section 3.4.1.3) the links to comprehensive and accurate road datasets. These issues are discussed further in the next two sections.

3.4.1.1. Zonal Data

The zonal module defines the polygon structure used to represent data and emission estimates for engine starts. It is the role of the zonal module to combine the polygons of various input data (i.e. socioeconomic, land use, TAZ) into a single polygon dataset. As mentioned before, TAZs represent the base spatial entity for engine starts. However, disaggregating trip origin estimates from large TAZs to smaller zones can be accomplished if good socioeconomic and land use data are available. For example, home to work trip origins can be assumed to start from the residential areas within the TAZ. Likewise, return trip origins can be assumed to start from land uses representing workplaces. While the process of disaggregation is discussed later, it is the role of the zonal module to establish the data linkages that make it possible.

Due to the fact that polygon data usually come from a variety of original sources and therefore a variety of spatial representation differences, significant errors may occur when trying to bring the datasets into a unified structure. It is unlikely that boundaries that represent identical features from different datasets will match perfectly. The result of this problem is the creation of a series of 'sliver' polygons whose attributes may be misaligned. However, there is no loss of information, only a zone structure that is as spatially accurate as the original data. The model does not make any assessment about the spatial accuracy, but uses whatever data are available. This allows users to define their spatial accuracy needs through the accuracy of the input data. Thus, if one wants estimates of engine start pollutant production within 100 meters, one must provide input data with equal or better spatial accuracy than 100 meters. Each of the new polygons maintains key fields tying them to their original

35

datasets, allowing all engine start emission estimates to be aggregated to any of the input polygon structures.

3.4.1.2. Lineal Data

The road module defines the lineal data used for predicting running exhaust emissions. While the travel demand forecasting models continue to be criticized for inaccurate roadway volume and speed estimates, they represent the only available prognostic regional vehicle activity tool. Most of the models predict travel (volume and average speed) only for major roads, aggregating minor roads to TAZ 'centroid connectors'. Further, the lineal representations of the road networks are usually spatially abstract structures. 'Links' represent actual road segments, but given modeling tasks, detailed shape points are unnecessary. In Atlanta, the absolute spatial errors resulting from the abstract representation exceeded 2 km in some instances [Bachman, 1996]. Improving this error is important to generating emissions for grid cells of 4 sq. km or smaller.

3.4.1.3. Conflation

Conflation is the blending of two line databases. Conflating the abstract travel demand forecasting network and a spatially accurate comprehensive road database is needed to improve the spatial accuracy of the travel model results. Because travel models' abstract 'links' represent actual road segments, it is possible to assess the connections to other road dataset lines based on link configuration and attributes. The process requires a link by link assessment and conflation by the user, resulting in a time-consuming and tedious task. Many planning organizations that develop travel demand forecasting models have already conflated the networks for purposes outside emission modeling. Conflation is required for the research model and not considered to be a task beyond the users needs. The model design can function without it, but at a significant loss of spatial error, thereby eliminating one advantage of the approach.

The purpose of the model's road module is to separate the conflated road dataset into modeled and unmodeled roads. Modeled roads, usually the roads with the most volume, become the major lineal structure used to represent running exhaust emissions. Unmodeled roads are aggregated into zones (bounded by modeled roads) used to represent minor running exhaust emissions. An argument that supports the zonal representation is the assumption that half the vehicles traveling on minor roads have a higher chance of operating under start conditions because they are closer to their origin (start conditions can last 2-3 minutes).

- *Spatial Environment Conceptual Correctness*

    The spatial environment modules have the task of defining the locational parameters for the rest of the model. The structure of the spatial environment needs to reflect the spatial characteristics of automobile exhaust emission production. Automobile exhaust emissions are produced by operating vehicles traversing a road network. The road network becomes the crucial component of the spatial environment. For major roads, there is little loss in the conceptual correctness of the spatial representation. Minor roads, however, suffer from insufficient prognostic data forcing zonal aggregation. The zonal aggregation uses discrete polygons in representing urban information. While the conceptual correctness suffers, important linkages become straightforward and the needs for minor road modal activity are lessened.

- *Spatial Environment Conceptual Completeness*

    The conceptual completeness of the spatial environment refers to the comprehensiveness of the spatial representation. The zonal aggregation of all roads not modeled by the travel demand forecasting model ensures comprehensive spatial representation by being a 'catch-all' for minor roads. The age of the input data will impact the completeness of the data. Recent land use changes or new road construction will be left out unless the input data are continuously maintained.

- *Spatial Environment Enterprise Awareness*

    The spatial environment structure is based on zonal and lineal representations of spatial structures used in a variety of agencies. By maintaining connections to the original input dataset identifiers, solid linkages to these agencies are provided. Further, by using a GIS and organizing data based on location, an indirect linkage to many enterprises is possible.

3.4.2. Fleet Characteristics

    Although different emission modeling approaches are being developed, all research efforts indicate that an improved capability to identify the emission significant components of the operating fleet is important to emission rate accuracy [Siwek, 1997]. Currently, emission models use model year distributions to describe the fleet. However, many other vehicle characteristics hold significant explanatory capability for predicting emission rates [Guensler 1994, Barth 1996]. Further, spatially variant emission estimates are needed, requiring spatially resolved sub-fleet characterization [Bachman, 1996]. Therefore, there is a need for identifying procedures that can accurately predict spatially resolved vehicle characteristics for urban areas. The fleet characteristics modules described in this section develop emission-specific and location-specific estimates of the distribution of automobiles.

37

~~Regional vehicle registration data provide information that allow emissions important vehicle characteristics to be determined for individual vehicles. The data also provide clues to identifying the vehicle's location, the owner's registered address, and the owner's ZIP code.~~ The fleet characteristics tier will develop estimates of technology distributions for each of the zonal and lineal representations. There are four general tasks: (1) attaching location parameters to the individual vehicle registration data; (2) determining important characteristics of the vehicles; (3) determining technology groups for the vehicles, and (4); aggregating to spatial entity-specific technology distributions.

The first vehicle characteristics module has two major tasks: determine individual vehicle location parameters and emission-specific characteristics. Each task is time-consuming due to the size of the registration datasets found in a metropolitan area. In Atlanta in 1995, 2.2 million vehicles were registered in the nonattainment area. The initial intense processing tasks need to be completed only one time per year, following new registration database development. ~~Therefore, the first vehicle characteristics module becomes a 'pre-processing' step, residing outside the formal model.~~

3.4.2.1. Vehicle Geocoding

Address geocoding is a process whereby standard address fields of road name, road type, and ZIP code are used to identify corresponding lines in a road database. The address number is used to identify the position of the address on a matched line based on the left and right address ranges. Address-matching usually results in success ranges of 60-80% dependent on the quality and comprehensiveness of the road dataset, and the number of errors associated with miscoding, duplicate or multiple road names, apartment numbers, and rural route identification. Growing urban areas have difficulty keeping road datasets current with new housing developments, adding significant bias to the geocoding errors.

The geocoding process results in two types of records: matched and unmatched. The matched vehicles are associated with a point entity. The unmatched vehicles maintain a default location identifier of ZIP code, a polygon entity. While ZIP codes can be rather large, they provide a degree of spatial information that can help determine regional fleet distribution variability.

3.4.2.2. VIN Decoding

Raw registration data can usually provide a few important vehicle characteristics (VIN, make, model, model year, and number of cylinders), but more information can be developed from the vehicle identification number (VIN). All

vehicles after 1980 are given a 17 digit VIN that consists of a code containing information about the types of emission control systems, the fuel delivery systems, the engine size, etc. Prior to 1980, VINs existed but lacked universal standards. Decoding the VIN for each vehicle requires the use of software (VIN decoder) developed by Radian International Corporation. Missing vehicle characteristics and the lack of updates prevent sole reliance on Radian's VIN decoder [Bachman, 1998]. Missing characteristics, pre-1972 autos, and post-1994 autos need to be developed from lookup files using the make, model, and model year. Research efforts at Georgia Tech have resulted in a datafile that can be used to determine the test weight of vehicles. While significant errors are expected, enough information should be available to develop a clear view of the operating fleet distributions.

The vehicle characteristics module results in two groups of vehicles and their emission specific characteristics; point-based (successfully matched) and zone-based (unmatched). These files should represent a comprehensive description of the region's fleet characteristics. These files are further processed to develop the emission-rate specific fleet distributions.

The zonal technology group module takes the spatially-resolved vehicle characteristics' files and determines zone-based engine start and running exhaust technology group distributions. The technology group (TG) definitions are defined by the emission rate modeling approaches included in the system. Currently, they are the aggregate modal approach (see section 2.2.3) and the speed correction factor approach (see section 2.2.1) because they are the only currently available models.

3.4.2.3. High and Normal Emitters

The aggregate modal approach developed emission-specific technology groups using a regression-tree analysis of emission test vehicles [Wolf, 1998]. In the analysis, all vehicles are divided into technology classes, indicating high or normal emitter fraction likelihoods. High and normal technology groups are then defined for each pollutant and emission mode (engine start and running exhaust). A sample engine start, normal emitter, CO, 'tree' is provided in figure 3.2. By starting at the top of the tree, conditions are identified based on the vehicles characteristics. True statements move to the left side of the tree, false statements move to the right. Each ending node is a set of conditions that are assigned a grams per start emission rate.

39
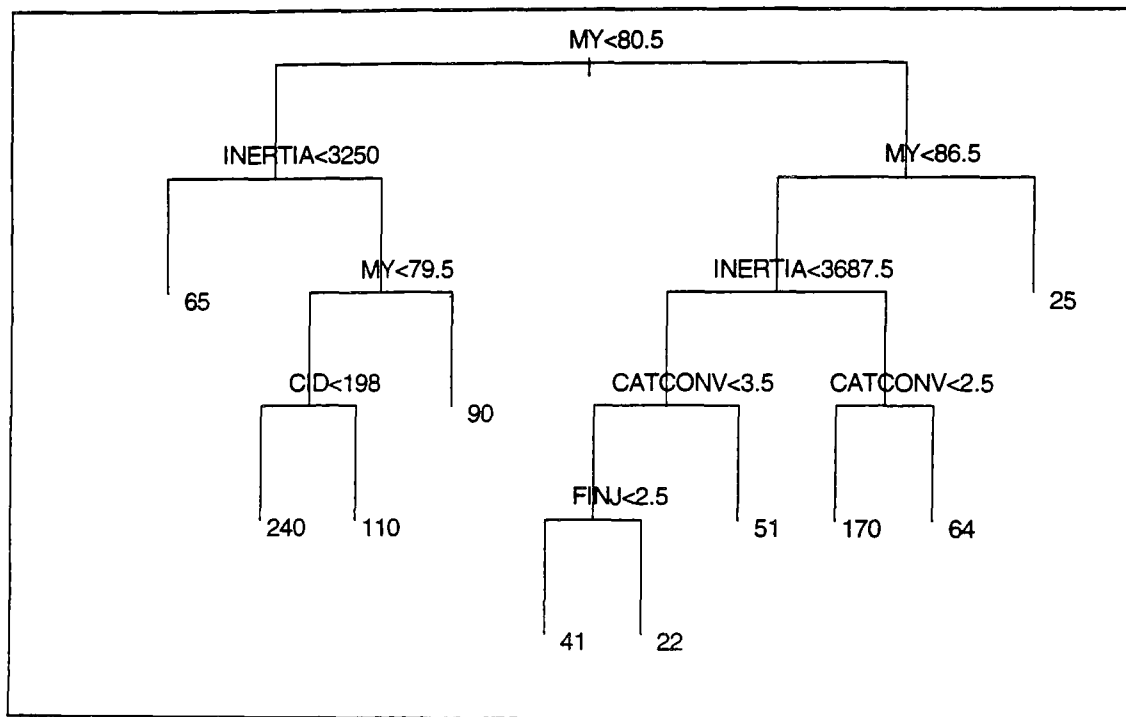
Figure 3.2 - Sample Regression Tree for Normal CO Engine Starts (grams/start)

A high emitting vehicle is one that has malfunctioning or tampered with emission control systems causing higher than normal emissions. It is expected that a small percentage of high emitting vehicles account for a large percentage of total emissions. High emitter determination is an important model design parameter and therefore it is appropriate to characterize these vehicles differently. The fraction of high emitters in the fleet, and the rate of malfunction among different vehicle types are unknown, but currently being researched. The regression-tree results by Wolf et al. divide the fleet into four groups that have different likelihoods of being high emitters. Lacking better information, a random sample for each group will be separated and labeled as high emitters, with sample sizes based on the group's likelihood. All other vehicles will be modeled as normal emitters.

3.4.2.4. Technology Groups

Once vehicles are identified as high or normal emitters, they are characterized into technology groups. Technology groups are combinations of vehicle characteristics and operating conditions that have been identified in the regression tree analysis as having significant emission rate differences. There will be separate technology groups for high and normal emitters, each pollutant, and each operating mode. Each vehicle will fall into six technology groups (engine start CO, HC, and

40

NOx, and running exhaust CO, HC, and NOx). Specific technology group descriptions are provided in chapter 4.

Engine start technology groups only include vehicle characteristics. For each emission-significant combination of vehicle characteristics, an associated gram per start emission rate is identified. Running exhaust technology groups include vehicle characteristics and(or) modal operating parameters (idle, cruise, acceleration, etc.). Unlike engine start groups, running exhaust technology groups can have different emission rates based on modal operating conditions. By the end of the fleet characteristics' modules, distributions of technology groups will exist for every zone in the model.

- *Fleet Characteristics Conceptual Correctness*

The representation of vehicles into emission-specific high and normal emitter technology groups is based on observed relationships discovered through test datasets. The ability of the technology groups to correctly represent the emission-specific characteristics of the operating fleet directly relates to the representativeness of the emission test dataset. Clearly this is not the case [see section 2.2]. However, alternative conceptual approaches suffer from the same limiting factors. As new vehicle tests are performed, and as re-analysis of past vehicle tests continues, progress towards a representative fleet will be accomplished. In fact, the technology group approach provides greatest potential for correct representation when representative samples are not provided.

- *Fleet Characteristics Module Conceptual Completeness*

The 'conceptual completeness' of the vehicle characteristics approach is fairly good, all operating automobiles are considered. However, data limitations severely hamper comprehensive development. By using a region's entire registration dataset, a comprehensive view of the region's vehicle characteristics is possible. Geocoding errors and decoding errors result in a significant loss in data [Bachman, 1998]. Problems with completeness are resolved by developing distributions of technology groups instead of frequencies. While lost data have bias and cannot be fully represented, the use of distributions provides a 'best guess' given the data limitations.

- *Fleet Characteristics Enterprise Awareness*

The vehicle characteristics can be tied to other users of spatially-resolved fleet descriptions because the locational parameters are defined before the fleet is segmented into emission-specific technology groups. This allows the individual vehicle characteristics to be available for other analyses. Inclusion of technology groups from two separate modeling approaches directly results in added flexibility and openness for the users.

41

3.4.3. Vehicle Activity

As mentioned previously, the core prognostic capability of the model rests on the ability of travel demand forecasting models to accurately predict regional travel. The emission-important vehicle activity estimates provided by the regional travel models are: the number and location of peak hour (or daily) trip origins, road segment volumes, and road segment average speeds. Important activity not provided by current models are; temporal travel behavior and modal (idle, cruise, acceleration, and deceleration) operations. As indicated in the background chapter, the average speed estimates can be very poor. Therefore, it is the role of the vehicle activity modules to transfer usable travel model information into the modeling environment, and develop estimates of the missing important parameters.

3.4.3.1. Engine Start Activity

Engine starts are equivalent to trip origins determined by the trip generation component of travel demand forecasting models. Travel demand forecasting models divide an urban region into traffic analysis zones (TAZs). The TAZs represent a spatial unit for aggregating socioeconomic data and resulting trip generation estimates. The designation of a TAZ should be based on homogeneous socioeconomic characteristics, reducing the variability of the trip estimates. However, many urban areas use zonal definitions based on cadastral (property) boundaries or US Census boundaries. TAZs are usually large (2-5 sq. km) due to the original objectives of the travel demand models (major infrastructure investments). Unless the TAZs can be disaggregated to smaller zones, the TAZ structure will determine the spatial resolution.

Estimates of trip generation are made for each TAZ for a variety of trip purposes. Trip purposes usually include trip production and attraction estimates of home-based work (to and from the workplace), home-based shopping, home-based school, home-based other, and non-home based. While these trips are estimated to begin or end in certain TAZs, the trip type definitions imply that they can be tied to land use. For example, a home-based-work trip consists of a trip originating from home (residential) going to work (non-residential), or a trip originating from work going home. Likewise, home-based shopping trips imply trips to or from a commercial land use.

The US Census Bureau maintains zonal databases developed for the decennial census. The smallest zonal designation is a block, usually an area bounded by roads or other lineal features (cadastral, hydrologic. etc.). At the census block level, 1990 estimates of the number of households are available. While the estimates are out-of-date, they can possibly provide clues to housing density within the TAZ and land use designations. This information can be used to further spatially disaggregate trips originating from residential areas.

By having good land use data and socioeconomic data, various trips can be disaggregated to smaller zones. Even if the land use designations are as broad as "residential" and "non-residential", the spatial resolution of trip generation estimates can be improved, allowing an improved spatial resolution for engine start estimates.

3.4.3.2. Intra-zonal Running Exhaust Activity

The road network used by travel demand forecasting models usually consists of major roads only. Travel on other roads is either not considered or predicted on an aggregate zonal (TAZ) basis. A key variable in predicting running exhaust emissions is the amount of travel time (preferably broken down by operating mode) because the longer a vehicle is operating, the more pollutant is produced. Travel times for intra-zonal trips (and inter-zonal travel off the major roads) are unaccounted for, other than looking at the size of the zone. However, information exists that allows the development of travel time estimates using the previously mentioned disaggregate trip generation estimates, a digital road network, and spatial analysis tools provided by the geographic information systems (GIS).

Many GISs provide tools that allow the determination of the shortest network path between two points. The disaggregated trip generation estimates provide a trip origin location. The closest intersection of local roads and major roads provides a destination location, representing the point during the vehicle trip when the travel demand models have assigned trips to the network. The shortest network path between the two points provides an estimate of the travel distance. Averaging all the distances within a TAZ provides an estimate of the typical intra-zonal and inter-zonal travel distance that occurs before vehicles reach the modeled network. Assuming an average speed for the local road travel provides an estimate of the average travel time. Although the strategy described above is crude and unvalidated, the method is better than the alternatives of leaving the estimates out, or assuming travel times based on zone area.

3.4.3.3. Modal Activity

Modal activity is a vehicle activity characterized by cruise, idle, acceleration or deceleration operation. Research has clearly identified that modal operation is a better indicator of emission rates than average speed [see section 2.2]. Determining regional modal operation is not possible using current travel demand forecasting models alone. All travel models can provide is road volume ($\pm$ 15%) and average speed ($\pm$ 30%). Because the accuracy of the average speed is poor, it should not be used in emission rate evaluation. However, the average speed could be accurate enough to determine differences in levels of service (LOS) E and F where volume to capacity (v/c) ratios approach or surpass 1.

Research by Grant et al. is attempting to characterize speed and acceleration profiles (Watson plots) by collecting data on major roads around Atlanta with a Laser Rangefinder [Grant, 1996]. The research has produced results for freeway and ramp sections by grade, LOS, and vehicle type. Using these results, speed and acceleration profiles can be identified for prevailing conditions predicted by the travel demand forecasting model. An example of a speed / acceleration profile is provided in Figure 3.3. The figure shows a graph where the x variable is speed in 5 mph increments (0-80), the y variable is acceleration in 0.5 mph/sec (+10 to -10) increments, and the z variable is the fraction of activity.
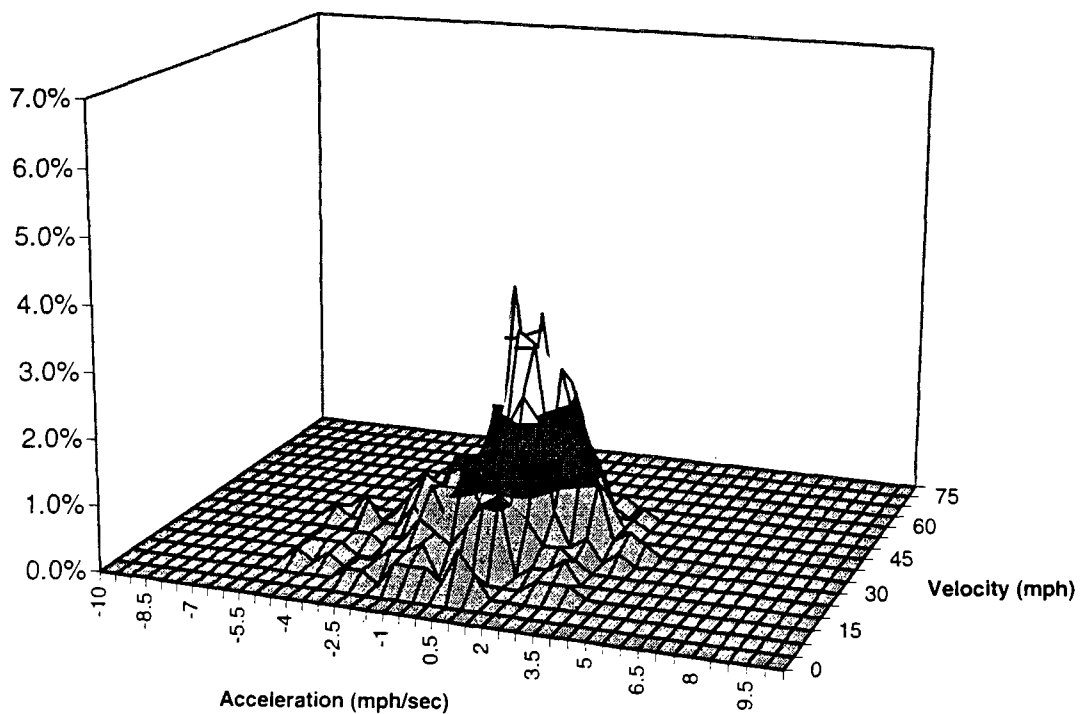
Figure 3.3 - Speed / Acceleration Profile, Interstate Ramp, LOS D

3.4.3.4. Road Grade

The impacts of road grade on emissions are included in the model design. Road grade affects vehicle emissions by impacting the load on the engine. Gravity exerts a force on a vehicle that must be counteracted to maintain a constant speed. Road grade is not included in mandated emission models because tests on the actual effects have not been completed and because metropolitan areas do not maintain spatially defined road grade estimates. Although grade impacts on emission rates are

being researched, results are not available at this time. However, the effects of acceleration on emissions have been quantified. Therefore, the secondary effects of grade on acceleration can be included in the conceptual design.

The effects of grade on acceleration can be quantified by the equation:

$$\text{Acceleration}_{(mph/sec)} = 22.15_{(mph/sec)} \times (\text{Gradient}_{(road)})$$

where $22.15_{(mph/sec)}$ represents acceleration due to gravity. For example, a vehicle wishing to maintain a constant speed along a 5% road grade must accelerate $1.11_{(mph/sec)}$ to counteract deceleration due to gravity.

Road grade data, while not currently comprehensively available for urban areas, is information that can be collected using global positioning systems (GPS) [Awuah-Baffour, 1997]. Given the expected importance of grade in affecting running exhaust emission rates, it is likely that the new GPS strategies will be employed by metropolitan areas in the next few years.

Including vehicle activity impacts resulting from road grade, even if not fully developed, provides an important step in emission model development. Strategies that are used for developing connections between road grade data and other road characteristics will act as guides in the development of future load-based models.

3.4.3.5. Temporal Variability

The temporal variability in estimates of vehicle activity is highly inaccurate because it relies on traditional travel demand forecasting models [see section 2.3.1]. Travel demand forecasting models are designed and operated to predict peak hour travel or daily travel. These are the primary temporal aggregation levels used by transportation planners and traffic engineers. The Travel Model Improvement Program (TMIP) administered by the US Department of Transportation is researching strategies for travel models to better predict activity during off-peak hours.

The ability of the research model to predict hourly emissions will rely heavily on accurate vehicle activity measures throughout the day. Until progress is made in the TMIP research, this emissions model will be unable to accurately incorporate off-peak travel. However, an intermediate step between existing and future models is possible. Many MPOs have developed estimates of hourly or subhourly travel demand factors based on travel survey data. These regional factors by trip type can be used to disaggregate daily trip generation into hourly intervals. Data on the variability of road volume throughout the day are available from departments of transportation for many major roads. Although average speed cannot be predicted to determine LOS F during

The 'conceptual correctness' of the vehicle activity refers to the accurate portrayal of the actions of an aggregate group of vehicles, or, in other words, the ability to predict the distribution of activity in a zone or on a link. The largest source of error comes from the travel demand forecasting model. Heavy reliance on the model transfers errors in trip generation estimates, road volumes, and road speeds to the emission models. The research approach attempts to lessen the impact of these errors by using modal activity measures when possible, and by disaggregating trip origins to appropriate land uses. The use of temporal factors causes substantial error in the activity estimates because the factors are used region-wide and lack spatial variability. While better than current practice, the approach results in significant problems with accurate representation of emission-specific vehicle activity.

- *Vehicle Activity Conceptual Completeness*

The 'conceptual completeness' of the representation of emission-specific vehicle activity refers to the ability of the modeling approach to capture all of the important activities. The largest gap in the completeness of the representation occurs on non-highway or ramp roads. Few speed and acceleration profiles are available for major and minor arterials. An enhancement to the travel model is a linkage between trip origins and the major road network. Minor road shortest paths allow vehicle activity to be estimated between the zonal-based starts and the lineal-based running exhaust. Further, inadequate representation of activity around signalized and unsignalized intersections may cause the exclusion of a large source of emission-specific vehicle activity. Until other data are available that can help determine these operations, the completeness of the vehicle activity estimates will be poor.

- *Vehicle Activity Enterprise Awareness*

The estimates of vehicle activity can be tied to other enterprises through the zonal aggregations and road network. The road network maintains a variety of locational parameters including street address and travel model identifiers.

3.4.4. Facility Emissions

Facilities are divided into zones and lines corresponding to the previously mentioned emission modes of engine starts and running exhaust (respectively). Facility estimates are used to allocate emission production to those vector spatial data structures currently used by transportation planners. By tying emission production estimates to facilities, tasks regarding research, reporting, validation, or control strategy development are made easier.

3.4.4.1. Engine Start Zonal Facility Estimates

Zonal facilities include the zonal representations of TAZs, land use, and Census blocks. The model design allows for other zonal designations to be included, but only the three mentioned have been required. The zones have been included in the definition of facilities because they are used by planners to aggregate socioeconomic information. While running exhaust emissions occur within zones, they are better tied to modal activity that occurs on the road. Engine starts, however, occur at trip origins, generally characterized with point or zonal information.

Figure 3.4 schematically represents the portion of a vehicle's emission profile represented by zonal facilities. The exhaust engine start estimates are modeled as a 'puff' (all start emissions allocated to the trip origin). While start emissions are actually dispersed through the network as a vehicle travels, research has not identified a strategy for correct spatial allocation. However, the role of this model is the study of emission production by automobiles, not air quality. It may be more useful for planners and/or researchers to have start emissions tied to the point of origin, thus allowing linkages to other zonal information.

Engine start emission rates are included in the research model based on results from the statistical model [see section 2.2.3]. Emissions in grams per start are estimated using the regression tree mentioned in sections 3.4.2.3 and 3.4.2.4. Six technology group trees exist for engine start emissions, all based on vehicle technology characteristics. Each emission estimate has established confidence bounds that can be translated back through the model to assess accuracy. The technology characteristics used in the tree process are listed below:

- *MY = Model Year*
- *EMM = Emission Control Equipment, 1-none, 2-oxi, 3-cat, 4-oxi&cat*
- *FINJ = Fuel injection equipment, 1-port, 2-carb, 3-throt*
- *CID = Engine Size, Cubic Inch Displacement*
- *TWT = CERT test weight, lbs.*
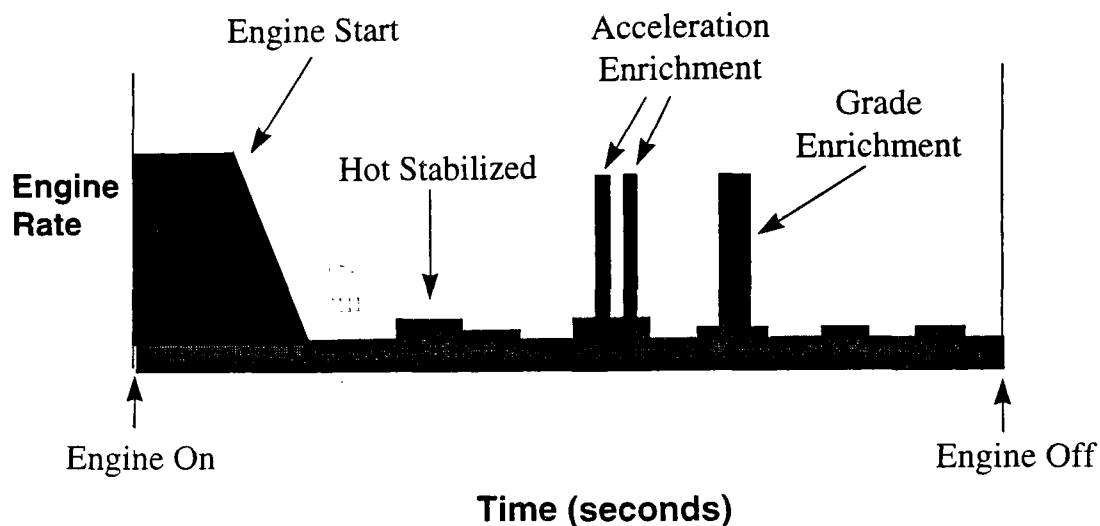
Figure 3.4 - Engine Start Emission Portion

The resulting technology groups are mutually exclusive and listed below:

- *CO Normal:*
1. MY < 1981, TWT < 3250
2. MY < 1980, TWT >= 3250, TWT < 4375
3. MY < 1980, TWT >= 4375, CID < 351
4. MY < 1980, TWT < 4375, CID >= 351
5. MY >= 1980, TWT >= 3250
6. MY = 1981, TWT >= 3688, CID < 131
7. MY = 1981, TWT < 2938, CID < 131
8. MY = 1981, TWT >= 2938, TWT < 368^ CID < 131
9. MY >= 1982, MY < 1987, TWT < 368:
10. MY >= 1982, MY < 1987, TWT >= 36&.
11. MY >= 1987

- *CO High:*
1. CID < 116, FINJ < 2
2. CID < 116, FINJ >= 2
3. CID >= 116, CID < 134
4. CID >= 134, CID < 258, FINJ < 2, MY < 1986
5. CID >= 134, CID < 258, FINJ = 2, TWT < 3563, MY < 1986
6. CID >= 134, CID < 258, FINJ = 2, TWT >= 3563, MY < 1986
7. CID >= 134, CID < 258, FINJ = 2, MY >= 1986
8. CID >= 134, CID < 258, FINJ >= 3
9. CID >= 134, CID >= 258

- *HC Normal:*
1. MY < 1980, TWT < 4125, CID < 154
2. MY < 1980, TWT < 4125, CID >= 154, CID < 241
3. MY < 1980, TWT >= 4125, CID >= 241
4. MY < 1978, TWT >=4125
5. MY >= 1978, MY < 1980, TWT >= 4125
6. MY >= 1980, EMM < 4, CID < 171
7. MY >= 1980, EMM < 4, CID >= 171
8. MY >= 1980, MY < 1988, EMM >= 4, CID < 98
9. MY >= 1980, MY < 1988, EMM >= 4, CID >= 98, CID < 102
10. MY >= 1980, MY < 1988, EMM >= 4, CID >= 102
11. MY >= 1988

- *HC High:*
1. MY < 1980
2. MY >= 1980, FINJ < 3, CID < 196
3. MY >= 1980, FINJ < 3, CID >= 196, CID < 258, MY < 1987
4. MY >= 1980, FINJ < 3, CID >= 196, CID < 258, MY >=1987
5. MY >= 1980, FINJ < 3, CID >= 258, MY < 1983
6. MY >= 1980, FINJ < 3, CID >= 258, MY >= 1983
7. MY >= 1980, FINJ >= 3, TWT < 2688
8. MY >= 1980, FINJ >= 3, TWT >= 2688, MY < 1988, CID < 192, TWT < 3063
9. MY >= 1980, FINJ >= 3, TWT >= 3063, MY < 1988, CID < 192
10. MY >= 1980, FINJ >= 3, TWT >= 2688, MY < 1988, CID >= 192
11. MY >= 1980, FINJ >= 3

- *NOx Normal:*
1. EMM < 3
2. EMM >= 3, EMM < 4, CID < 230
3. EMM >= 3, EMM < 4, CID >= 230, CID < 245, FINJ < 2
4. EMM >= 3, EMM < 4, CID >= 230, CID < 245, FINJ >= 2

5. EMM >= 3, EMM < 4, CID >= 245
6. EMM >= 4, CID < 122
7. EMM >= 4, CID >= 122, CID < 138
8. EMM >= 4, CID >= 138, CID < 146
9. EMM >= 4, CID >= 146, CID < 152
10. EMM >= 4, CID >= 152, CID < 213
11. EMM >= 4, CID >= 213, CID < 288
12. EMM >= 4, CID >= 288

- *NOx High:*
1. EMM < 3, CID < 334, MY < 1980
2. EMM < 3, CID < 334, MY >= 1980
3. EMM < 3, CID > 334
4. EMM >= 3, CID < 137, MY < 1987
5. EMM >= 3, CID < 137, MY >= 1987
6. EMM >= 3, CID >= 137, CID < 152
7. EMM >= 3, CID >= 152, CID < 230
8. EMM >= 3, CID >= 230, CID < 232
9. EMM >= 3, CID >= 232

Zonal estimates of fleet characteristics are divided into the previous technology groups. Each technology group fraction is multiplied by the number of trip origins that occur in the zone. The resulting number of trip origins by technology group is multiplied by the associated gram per start emission rate. The resulting emissions of CO, HC, and NOx are reported for a typical weekday (Tuesday - Thursday) on an hourly basis. The typical weekday limitation is a result of the travel demand modeling process, as few models predict weekend or Friday travel.

3.4.4.2. Minor Road Zonal Facility Estimates

Minor road zones [see section 3.4.1.2] are used to spatially represent the portion of running exhaust emissions that occur between the trip origin and the roads modeled by the travel demand forecasting network. Available minor road vehicle activity information is limited because it is not explicitly modeled in the travel forecasting process, and there are no existing measures of modal activity available for local roads. Lower traffic densities and lower average speeds suggest that the actual portion of running exhaust err. ;ions occurring on local roads may be small. However, little evidence is avai, ·.e to draw conclusions about the impacts of local road driving. This limitation fu.~es a scaled back version of local road emissions modeling.

3.4.4.3. · Lineal Facility Estimates

Lineal facilities are roads that are modeled in the travel demand forecasting model. On-road fleet distributions and predicted traffic flow parameters are used to generate road segment specific estimates of CO, HC, and NOx. Figure 3.5 shows the portion of the emission spectrum represented by linear features. Minor road running exhaust emissions and major road running exhaust emissions estimate the same pollutant and, combined, predict total running exhaust emissions. Network characteristics determine the amount of the running exhaust portion that is allocated to minor zones.
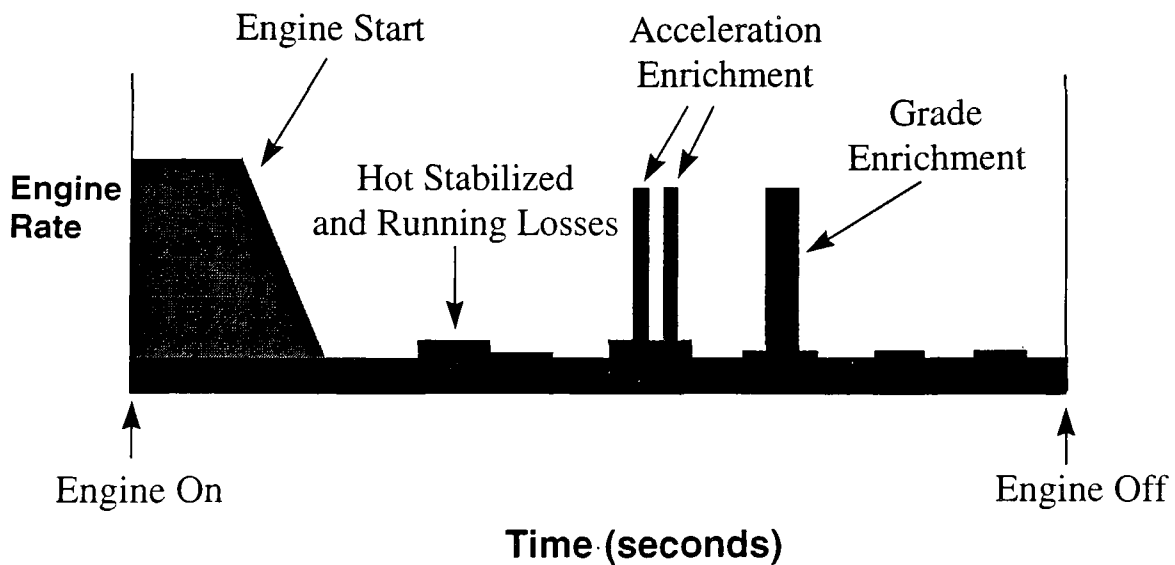
Figure 3.5 - Running Exhaust Emission Portion

Emission rates for running exhaust come from two sources, the statistical approach used for start emissions, and the SCF approach, used by currently mandated models. The purpose for including both approaches is to allow user flexibility and to provide a platform for comparison. Vehicle activity measures for both emission rate approaches come from the same source, although different variables are needed. The SCF approach needs average speed, while the statistical approach uses a variety of

other modal parameters. The same vehicles are aggregated and used for the estimated fleet distribution, although different technology group definitions exist.

The technology groups (see section 3.4.4.2 for variable definitions) for the statistical model were determined similarly to those described in the engine start section. The definitions are as follows:

- *CO Normal:*
1. EMM < 4, MY < 1979
2. EMM < 4, MY >= 1979
3. EMM >= 4, CID < 146, MY < 1979
4. EMM >= 4, CID < 146, MY >= 1979
5. EMM >= 4, CID >= 146, MY < 1979
6. EMM >= 4, CID >= 146, MY >= 1979, MY < 1985
7. EMM >= 4, CID >= 146, MY >= 1985, MY < 1987
8. EMM >= 4, CID >= 146, MY >= 1987

- *CO High*
1. EMM < 4, FINJ < 3, TWT < 3375
2. EMM < 4, FINJ < 3, TWT >= 3375
3. EMM < 4, FINJ >= 3
4. EMM >= 4, TWT < 3313
5. EMM >= 4, TWT >= 3313

- *HC Normal:*
1. MY < 1985
2. MY >= 1985, MY < 1987, TWT < 3188
3. MY >= 1985, MY < 1987, TWT >= 3188
4. MY >= 1987, MY < 1990, CID < 143
5. MY >= 1987, MY < 1990, CID >= 143, CID < 196
6. MY >= 1987, MY < 1990, CID >= 196
7. MY >= 1990, TWT < 3563, CID < 143
8. MY >= 1990, TWT < 3563, CID >= 143, CID < 186
9. MY >= 1990, TWT >= 3563, CID < 143
10. MY >= 1990, TWT >= 3563, CID >= 143, CID < 186
11. MY >= 1990, CID >= 186, CID < 196
12. MY >= 1990, CID >= 196

- *HC High:*
1. MY < 1984, EMM < 4
2. MY >= 1984, MY < 1985, EMM < 4
3. MY >= 1985, MY < 1988, EMM < 4

4. MY < 1985, EMM >= 4, TWT < 4000
5. MY < 1985, EMM >= 4, TWT >= 4000
6. MY >= 1985, MY < 1988, EMM >= 4
7. MY >= 1988

- *NOx Normal:*
1. MY < 1989, CID < 154
2. MY >= 1989, MY < 1994, CID < 154
3. MY >= 1994, MY < 1995, CID < 154
4. MY >= 1995, CID < 154
5. MY < 1994, CID >= 154
6. MY >= 1994, MY < 1995, CID >= 154
7. MY >= 1995, CID >= 154

Each engine start technology group will have a gram per start emission rate. Running exhaust emissions rates depend on modal operation. The regression tree used to determine running exhaust emissions rates includes modal parameters. The modal parameters are:

- *AVGSPD*   *The average speeding miles per hour.*
- *PKE>X*   *The fraction of activity with positive kinetic energy (speed*acceleration) greater than X mph^2/sec.*
- *POW>X*   *The fraction of activity with power ($speed^2$*acceleration) greater than X mph^3/sec.*
- *ACC>X*   *The fraction of activity with acceleration greater than X mph/sec.*
- *DEC>X*   *The fraction of activity with deceleration greater than X mph/sec.*
- *CRZ>X*   *The fraction of activity with zero acceleration and speed greater than x mph.*
- *IDLE*   *The fraction of the activity with zero acceleration and zero speed.*

For each road segment and each hour, modal variables are determined. Road segment-specific technology groups and modal variables are combined to develop the fraction of activity with specific emission rates (grams per second). Total hourly travel time is calculated and segmented by the fraction of the vehicles with each emission rate.

The SCF emission rate approach uses a 'look-up' table created from running MOBILE5a for a series of model year distributions and average speed distributions. The vehicle characteristics developed for the model include model year as a variable, allowing the creation of model year technology groups. The vehicle activity component of the model allows the estimate of average speed to be available for every

road segment and every hour. By running MOBILE5a for each combination of average speed and model year distribution, gram per second emission rates will be developed.

- *Facility Emissions Estimates Conceptual Correctness*

Conceptual correctness for facility emission estimates is the ability of the approach to accurately estimate actual emission production. Sources of inaccuracies come from three places; the quality of the input data, the model's manipulation of the data, and the errors in the development of emission rates. The quality of the input data and the ability of the model to manipulate them into a usable form result in large errors. All of the problems mentioned in the previous sections culminate in substantial error. Errors in the development of the emission rates are the result of incomprehensive test datasets and unrepresentative operating profiles. Emission rates have confidence bounds associated with them, allowing some measure of the variability of the errors. As in the representation of vehicle activity, the conceptual correctness of the facility emission rates is defined by the aggregate level of the estimate. While the accurate assessment of a single vehicle may be poor, the aggregate estimate may prove better. The SCF emission rates have been strongly criticized as being insensitive to important activity. Thus, the SCF emission rates have less 'correctness' than the statistically based approach that includes modal parameters.

- *Facility Emissions Estimates Conceptual Completeness*

The 'conceptual completeness' of the exhaust emissions estimates is fairly strong. All automobiles will fit into engine start, running exhaust, and SCF technology groups. All operating vehicles will fall in the speed and acceleration profile identified for the particular conditions. The only source of incomplete representation of emission rates is due to problems that have already been mentioned regarding the emission test dataset and the cycle test design. All represented vehicles can be assigned an emission rate.

- *Facility Emissions Estimates Syntactic Correctness and Completeness*

The communication of the facility emission estimation component relies on clear definitions and visual representation. The terminology used to describe the technology groups and other input parameters can be clearly communicated if the user is given concise explanations of all the parameters. The spatial component of the facility estimates can be clearly communicated using GIS and by using the input data spatial structures.

- *Facility Emissions Estimates Enterprise Awareness*

The estimates have strong 'enterprise awareness' due to the ability of the estimates to be aggregated to any of the original input spatial structures (TAZs, Census, etc.). If needed, the estimates can also be presented as emissions per unit of

54

distance or area to aid in translating the other locational parameters. This step is improved through the translation to raster structure in the next section; however, the base information needed is available at this level.

3.4.5. Emissions Inventory

The role of the emissions inventory module is to prepare the facility-based emission estimates for input into gridded photochemical models. An important component of the entire modeling process is the ability to aggregate estimates to a user-defined grid cell size. The most efficient technique for accomplishing this task is to convert the engine start, minor road running exhaust, major road running exhaust, and major road SCF emission estimates to raster data structures. Once in a raster structure, developing gridded estimates for inventory reporting is fairly easy. Conversion of the data from vector to raster is a tool available in many of the larger GIS software tools. After conversion, total mobile source emission estimates are calculated for the entire area. Engine starts, minor road running exhaust, and major road running exhaust emissions are used to develop totals.

The tools available in the GIS for conversion make some assumptions about the vector data that may not be desired. Problems occur with direct conversion, especially for linear structure. Straight conversion is possible, but grid cells take on the value of the largest feature, or largest portion within its boundaries. For a linear feature, this means that all cells that represent the line will have the same emission value or rate. However, the line can bisect the cell at any point, resulting in variations in the cell's ability to properly identify the portion of the road that falls within its boundaries. Similar issues occur with polygons along their boundaries. The smaller the grid cell, the lesser the problem. However, gridded inventories require grid sizes as large as 4-5 square kilometers, much larger than the anticipated zonal and lineal structures.

One way around the problem is to intersect the zonal and lineal structures with a vector grid. Once intersected, all emission values falling within the grid cell boundaries can be weighted by area or length and summed. The resulting vector grid data are then converted to raster cells with cell sizes equivalent to the vector grid size.

The final raster datasets are individual 'layers' of each pollutant by hour and emission 'mode' (totals, engine starts, etc.). Tools available in the GIS allow for the development of special visualization interfaces that can create and query two and three dimensional images of the various databases.

55

- *Gridded Emissions Conceptual Correctness*

For the final module, 'conceptual correctness' refers to the ability to maintain data integrity while converting from raster to vector. The deterministic, vector approach to developing gridded estimates ensures that few errors are introduced during the process. There are also problems with grid cells that fall on the boundaries of the study area. Cells that overlap study area boundaries will only have values for those emissions in the study area, causing the cell value to be underestimated. As long as cells lie completely within the boundaries, this is not a problem. The accuracy of the gridded estimates is a function of all of the previously discussed problems.

- *Gridded Emissions Conceptual Completeness*

The completeness of the process is only limited when grid structures do not encompass or lie within the entire study area.

- *Gridded Emissions Enterprise Awareness*

Aggregating and rasterizing the emission estimates allow the storage and communication of specific emission production intensities. Previously, a zone would represent only emission produced by 'zonal' information like trip generation estimates. With the gridded emissions, every location has an estimate of the total emissions as well as the specific modes. The flexible locational parameters allow the estimates to be translated to other areas.

3.5. Conclusion

This chapter defined the parameters and quality assurance measures for the design of a comprehensive modal emissions model. Initially, a detailed list of issues to be addressed by the model was compiled from background research. This list of issues led to a model design that incorporated those parameters while maintaining a flexible, comprehensive and accurate account of the science being modeled. In developing the design, GIS became a powerful tool for data preparation, storage, and analysis.

4. MODEL DEVELOPMENT

This chapter discusses the design of the research model, hereafter referred to as the Mobile Emission Assessment System for Urban and Regional Evaluation (MEASURE). The chapter provides descriptions of the data, processes, algorithms, and files. The chapter will review each of the input files, pre-processing steps, program modules, and program code. By the end of the chapter, the reader should have a clear vision of the model scope and process.

The purpose of the MEASURE is to provide researchers with a tool for measuring the air quality impacts of urban and regional transportation policy and development changes. This model is not intended to be used directly for conformity or inventory reporting, but for use in a research environment by scientists exploring the transportation and air quality relationship.

MEASURE will produce hourly transportation facility-level and gridded automobile exhaust emission estimates of carbon monoxide (CO), hydrocarbons (HC), and oxides of nitrogen (NOx). The model develops these estimates based on a geographic area's vehicle registration data, accurate digital road dataset, TRANPLAN (travel demand forecasting) model output, and zone-based socioeconomic data (i.e., US Census Blocks, Land Use). The outputs of the model include dbase data files and gridded inventories in ARC/INFO raster 'grids'. ARC/INFO software features allow the user to develop maps and other visualization tools.

MEASURE currently includes:

- *Modal (cruise, acceleration, deceleration, idle) activity*
- *Estimates of CO, HC, and NOx*
- *~100 emission specific technology groups*
- *Separate tree-based regression engine start and running exhaust emission rates*
- *Comparative MOBILE5a SCF running exhaust estimates*
- *Impacts of grade on acceleration*

MEASURE currently does not include:

- *Non-automobiles*
- *PM10 or PM2.5*
- *Evaporative emission estimates*
- *'Vehicle deterioration effects*
- *'Impacts of grade on engine load*

- *Impacts of accessory load*
- *Intersection specific estimates*

Included, but technically weak components:

- *Hourly, on-road traffic volume estimates*
- *Hourly, on-road traffic flow estimates*

Running and managing the model require technical skills in geographic information systems (ARC/INFO 7.0), transportation planning practice and software (TRANPLAN), air quality modeling (MOBILE 5a), C programming, and UNIX and MS-DOS operating systems. Substantial data collection and processing activity is required prior to model operation and is discussed in detail in the next section.

MEASURE is divided into 12 modules (see Figure 4.1). The modules and their associated input and output data files are managed by a single ASCII description file called 'Makefile'. The 'Makefile' is executed by a UNIX system utility called *make* (discussed in detail later). The modules roughly follow a tree structure in that one module may depend on the outputs of other modules. The modular structure allows individual components to be executed independently or in a fully connected process. The modules are grouped into five tiers: the spatial environment, the fleet characteristics, the vehicle activity, the facility emissions, and the emission inventory.

GIS-Based Model of Automobile Exhaust Emissions

Input Data:	• Census Block Data • Land Use Data • Traffic Analysis Zone Data • Vehicle Registration Data	• Road Network Data • Travel Demand Forecasting Network • Speed / Acceleration Profiles • Road Grade Data

Spatial Environment

1 Zonal Module — 2 Road Module

Fleet Characteristics

3 Zonal TG Module — 4 On-road TG Module

Vehicle Activity

5 Engine Start Activity Module — 6 Minor Road Activity Module — 7 Running Exhaust Activity Module

Facility Emission

8 Engine Start Emissions Module — 9 Minor Road Running Exhaust Module — 10a Agg. Modal Running Exhaust Module — 10b SCF Running Exhaust Module

Emission Inventory

11 Hourly Gridded Starts, Hot Stabilized, Enrichment, and Total Emissions Module

Figure 4.1 - Model Design

If all input files and associated software are in place, the modules are executed in the appropriate order by typing 'make emissions' at the UNIX command line. The test dataset (100 sq. km in northeast Atlanta) should take approximately four hours to complete on a Sun SPARCstation 10. A larger urban area may take as long as 24 to 36 hours to complete a full execution.

4.1. Input Files

This section will describe the datasets, fields, and directories needed to be resident in the system. A substantial amount of data collection and processing effort needs to be conducted prior to model operation. Several tools and guidelines are included to aid this process. As a general note, all ARC/INFO coverages and files containing coordinates should use the same geographic projection and use meters as the base unit.

4.1.1. Directory structure

The directory structure consists of one home directory and several data directories. The data directories are:

- zone: *stores all zonal data and coverages*
- road: *stores all lineal data and coverages*
- grid: *stores all vector grid data*
- em: *stores all emission estimate data*
- tg: *stores all technology group data*
- grade: *stores all road grade related data*
- raster: *stores all raster data*
- raw: *place to store backup copies of data and programs*
- aml: *stores all AML code*
- code: *stores all C code*
- templates: *stores a number of INFO file templates*
- sa: *stores all speed / acceleration profiles*
- modalmats: *stores all modal matrices*
- lookup: *stores all ASCII lookup files*
- temp: *stores temporary files used during program runs*

4.1.2. Zone.twt and zip.twt

Two ASCII files, zone.twt, and zip.twt, need to be created and stored in the *tg* directory. These files represent the area's vehicle characteristics. *Zone.twt* is a list of successfully addressed geocoded vehicles and an ID number representing the registered zonal location. *Zip.twt* are those vehicles that were not matched and therefore only the registered ZIP code location is included. The ASCII files have the same comma-delimited structure of:

- *zone or ZIPcode*
- *vehicle identification number (VIN)*
- *model year*
- *emission control code*
- *fuel delivery system code*
- *engine size (cubic inch displacement)*
- *vehicle dynamometer test weight*
- *CO high emitter flag*
- *HC high emitter flag*
- *NO high emitter flag*

Creating these files from an area's vehicle registration dataset is a lengthy, one-time process and therefore removed from the formal model. The first step in the process is to acquire the area's registration datafile that contains VIN, street address, and ZIP code. For a large urban area, this initial file can consist of millions of records (2.2 million for Atlanta). The file needs to be address geocoded (with offset) using tools available in ARC/INFO or other software, and an accurate, up-to-date, road database. Successful matches are stored in an ARC/INFO point coverage with the VIN. Unsuccessful matches are stored in an ASCII file of ZIP code and VIN. Keep track of the match rate, as it will be important later in the model. The point coverage is then overlaid with any zone coverage, preferably of census blocks, using ARC/INFO's *identity* command. The resulting zone-id and VIN are written to an ASCII file. The two ASCII files are processed through a series of programs written by individuals at Georgia Tech. These programs read these files, decode the VINs using vendor software, flag a random sample as high emitter for the three pollutants, and write the results to ASCII files called *zone.hne* and *zip.hne*. A separate program, *hne2twt.c*, reads these files and uses a lookup table to add vehicle dynamometer test weight.

4.1.3. Allroads (ARC/INFO coverage)

The *allroads* coverage is a line database of all roads in the area and should be as accurate as possible. It should be stored in the *road* directory. The database contains, in addition to standard ARC/INFO fields, the key fields called *arid* and *tfid*.

These fields are identifier fields and are used to link datasets back to the individual road segment locations. *Arid* is unique for each road segment. *Tfid* identifies the corresponding travel demand model link. Because travel demand models usually model major roads only, not all lines in *allroads* will have a non-zero *tfid*. Further, some travel model links will span a number of lines in *allroads* resulting in lines that have the same *tfid.* Establishing the *tfid* on the lines is completed through a manual process called conflation. The process of conflation involves selecting each travel model link, selecting all corresponding road segments, and joining the ids.

4.1.4. Tdfn.dat (INFO file)

The *tdfn.dat* INFO data file is a table stored in the road directory with the following items:

- tfid - *the key field to link to* allroads
- assign_gr - *the road classification code (1-interstates, 2-ramps, 3-major arteries, 4-minor arteries)*
- abvolume - *the oneway 24 hour volume*
- abspeed - *the oneway 24 hour average speed*
- abcap - *the oneway road capacity*
- bavolume - *the otherway 24 hour volume (blank if divided road)*
- baspeed - *the otherway 24 hour average speed*
- bacap - *the otherway road capacity*

This file is created by converting the area's travel demand forecasting model outputs into an INFO format. Programs by Georgia Tech are available that can complete this process for TRANPLAN models. If other travel modeling software is used, one may have to contact the vendor for a conversion package, or develop customized software that creates datafiles compatible with ARC/INFO's 'generate' and 'add from' commands.

4.1.5. Census (ARC/INFO coverage)

The *census* polygon coverage (stored in the *zone* directory) can be any zonal structure that contains a 'household' field. US Census blocks are preferred because they are available around the country and provide substantial detail with regard to population and household information. The *census* coverage contains the following fields:

- cbid - *the key field to link associated databases together*
- hu90 - *the household field*

Census data can be acquired from the US Census Bureau.

4.1.6. Land use (ARC/INFO coverage)

The *landuse* polygon coverage (stored in the *zone* directory) is a database of residential, non-residential, and commercial land uses. The coverage needs to contain only a single attribute field of *lu*. *Lu* is a character field that consists of 'RES' if the zone is residential, 'COM' if the zone is commercial, and 'NONRES' if the zone is non-residential. Areas without identifiers will be considered non-residential.

4.1.7. ZIP code (ARC/INFO coverage)

The *Zipcode* polygon coverage (stored in the *zone* directory) is a database of postal ZIP codes. The coverage needs only a single field called *Zipcode* and should be populated with current five-digit ZIP codes.

4.1.8. TAZ (ARC/INFO coverage)

The *TAZ* polygon coverage (stored in the *zone* directory) is a database of traffic analysis zones used in the travel demand forecasting process. The coverage needs a single key field called *tzid* and should be unique for every zone. Developing this datafile will probably require assistance from the local planning organization.

4.1.9. TAZ.dat (INFO file)

The *TAZ.dat* file stores all the trip generation information from the four step travel demand model. The fields are as follows:

- tzid   *the key field used to link the data to other zones*
- hbwprd   *home-based work productions*
- hbshprd   *home-based shopping productions*
- hbgsprd - *home-based grade school productions*
- hbuprd - *home-based university productions*
- nhbprd - *non-home-based productions*
- hbwatt   *home-based work attractions*
- hbshatt - *home-based shopping attractions*
- hbgsatt   *home-based grade school attractions*
- hbuatt   *home-based university attractions*
- nhbatt - *non-home-based attractions*

The productions and attractions fields listed above represent standard travel demand forecasting terminology for trip type. The fields should represent the 24-hour trip type quantities by TAZ. This file, like the *TAZ* coverage, needs to be converted to INFO from whatever format the planning organization provides.

4.1.10. Landmarks(ARC/INFO coverage)

The *landmarks* coverage is a point database of educational institutions. The points should have an identifier field that shows whether or not it is a university or grade school. The determination of which schools to include depends on the definition of the trip types from the travel demand forecasting model. They are used primarily to spatially allocate home-based-university trips and home-based-grade school trips.

4.1.11. Grid (ARC/INFO coverage)

The *grid* coverage will represent the spatial structure of the inventory estimates. Grid cells of any size can be used; however, they should not be smaller than the published accuracy of any of the input coverages. The coverage should contain one attribute called *gdid*. *Gdid* will be used as a key item to aggregate estimates from the zones and lines.

4.1.12. Grade.xy and grade.gr

The ASCII files *grade.xy* and *grade.gr* store road grade information collected from GPS units. The *grade.xy* file contains comma delimited fields of grade-id, x_coordinate, and y_coordinate. The coordinates must match the geographic projection system and units used by the coverages. The *grade.gr* file contains comma delimited fields of grade-id and grade. The files must be separate in order to read them into ARC/INFO as a point coverage. The model joins them in the *mr-act.aml* process.

4.1.13. Lookup files

Three lookup files are provided, *temporal.factors* (INFO file), *scf.csv* (ASCII file), and *twt.lu* (ASCII file). These files can be used in any model run, but may be updated. *Temporal.factors* provides the breakdown of vehicle volumes and trips by hour. *Scf.csv* provides MOBILE 5a speed-corrected emission factors by 5 mph speed increment and model year. It is used in developing the SCF running exhaust emission estimates. *Twtlu.asc* is a file of vehicle make, model, model year, and test weight. It is used outside MEASURE to add the test weight field to the vehicle characteristics files.

4.2. The Makefile

The *Makefile* is the most important file in the system to become familiar with. It manages the entire modeling process by checking file and code dependencies. The *Makefile* is interpreted by a system utility in UNIX operating system software (*make*). *Make* identifies file dates and times to determine when updates have been made, and, if needed, calls a series of actions. For example, the first dependency relationship in the *Makefile* is listed as follows:

```
$(em)/grid-em.dbf      :  $(em)/scf-em.dbf \
                          $(em)/sz-em.dbf \
                          $(em)/mr-em.dbf \
                          $(em)/mz-em.dbf \
                          $(aml)/grid-em.aml
                /bin/rm -r $(raster)
                /bin/mkdir $(raster)
                /bin/rm -r $(temp)
                /bin/mkdir $(temp)
                $(arc) "&r $(aml)/grid-em.aml"
```

Make first checks to see if a file `grid-em.dbf` exists, and if it does, it compares it with the last update of the other files. If one or more of the other files has a newer date than 'grid-em.dbf', the last five lines are executed. If `grid-em.dbf` is current with respect to the other files, the code is not executed. The value of this is that it saves time in large complicated multi-program processes because only those portions that have been updated are executed.

The *Makefile* is segmented into twelve parts that represent the twelve modules listed in Figure 4.1. The twelve parts are called:

- *gridded_emissions*
- *scf_emissions*
- *eng_starts_emissions*
- *maj_rds_run_exh_emissions*
- *min_rds_run_exh_emissions*
- *min_rds_activity*
- *eng_starts_activity*
- *maj_rds_activity*
- *run_exh_tech_groups*
- *eng_starts_tech_groups*
- *rds_spatial_environment*
- *zonal_spatial_environment*

65

Any of the above parts can be executed by typing 'make <part>' and that particular module, and any non-up-to-date module it depends on, will be executed. By typing 'make gridded_emissions' all the parts are checked and executed if needed because the final gridded emission estimates depend on all of the other components.

In addition, two other parts have been added to aid in modeling. 'Make clean' will remove all data files except the original input files. This can be handy when one wishes to start with a clean slate. 'Make programs' will compile all the 'C' programs. This is handy when editing code and you want to make sure it works before running the Model.

Prior to running the Model on a new system, edits to the *Makefile* must be made. The first section entitled 'Variable Definitions' is as follows:

```
dir   = /proteus/home/wbachman/gismodel2
tg    = $(dir)/tg
zone  = $(dir)/zone
road  = $(dir)/road
lm    = $(dir)/landmarks
em    = $(dir)/em
grid  = $(dir)/grid
c     = $(dir)/code
aml   = $(dir)/aml
raster   $(dir)/raster
temp  = $(dir)/temp
arc   = /miranda/ceesri/arcexe70/programs/arc
cc    = /opt/SUNWspro/bin/cc
```

The variable 'dir' in the first line must be updated to the current path where the *Makefile* and other directories reside. Also, the 'arc' variable must be updated to the correct executable path for ARC/INFO. The same update must be made for the 'cc' variable, identifying the location of the C compiler.

4.3. The Modules

The modules, although separate, are linked together through dependent and co-dependent files and programs. Even though they are discussed separately, and function independently, their design and operation is affected by inputs and outputs of other modules.

4.3.1. Zonal Environment Module

The zonal environment module's purpose is to establish a polygon database called *sz* that will be used to develop engine start emission estimates. Figure 4.2 represents the entities involved in MEASURE. Figure 4.3 represents the flow of the code. The polygon database is created by spatially joining four input coverages, *census, taz, landuse,* and *ZIPcode.* The process of joining the databases involves the use of the ARC/INFO command 'identity'. The input coverages may have relative inaccuracies that can cause 'sliver' polygons to be created. Operationally, this is not important, but it does affect relative accuracy. A line on one coverage may represent the same feature as the line on another coverage; however, the spatial representations of the line may differ. To lessen the impact of this problem, a 'fuzzy' tolerance is designated. Any lines that fall within the tolerance distance will be considered the same feature. The fuzzy tolerance should be set to the size of the worst reported absolute accuracy. In most instances, this will be the Census Bureau files that maintain an accuracy of 30-100 meters.

The module also creates a data file that maintains various attributes accumulated through the integration of the four input databases. These fields are land use (RES, NON-RES, and COM), housing unit per square kilometer, and all the TAZ trip generation estimates. The housing unit rate is multiplied by the *sz* area to predict the number of households. The amount of error introduced by disaggregating household data is a function of the size and accuracy of the original household database. If the Census blocks are used, a common source of large scale household data, the disaggregation errors will be small because the new zones will be similar to the blocks.

4.3.2. Road Environment Module

The purpose of the road environment module is to create the spatial structure used to represent running exhaust emissions. The road environment module divides the *allroads* coverage into roads modeled by the area's travel demand forecasting model, and those that are not modeled. The modeled road segments are used to create the *mr* (major road) line coverage. The *mr* lines form the boundaries for *mz* (minor zone) polygons that represent a zonal aggregation of minor roads. Figures 4.4 and 4.5 represent the entities and program code flow, respectively.

The module does not result in any loss of accuracy. Some minor zone polygons become insignificant due to the techniques used for generation. Median areas bounded by two parallel roads become minor zones. They bias the network travel distances slightly (determined later).

4.3.3. Zonal Technology Groups Module

The purpose of the zonal technology group module is to convert the outputs of the vehicle characteristics process into zonal technology groups. The vehicle characteristics process and zonal TG modules are displayed in Figures 4.6 and 4.7. Figures 4.8 and 4.9 describe the pre-processing steps. The module reads the two ASCII input files of *zone.twt* and *zip.twt* and creates four dbase files; *sztg.dbf, retg.dbf, scftg.dbf,* and *regiontg.dbf.* The module contains four programs; *techgr.c, regiontg.c, jointg.c,* and *sz-tg.aml.* *Techgr.c* (Figure 4.10) is executed first and it assigns each vehicle into a technology group for each pollutant and summarizes by zone or ZIP code. The resulting distributions are written to a series of ASCII files. The *regiontg.c* program is similar in structure to *techgr.c,* but it reads all the vehicles and determines a regional technology group distribution. The *jointg.c* (Figure 4.11) program brings the separate zonal and ZIP code distributions using a control file (*szzp.asc*) containing zonal ids and corresponding ZIP codes. It is important to adjust the code to the address match success rate developed in the preprocessing. Currently the rate is set to 81% zonal and 19% ZIP code, suggesting that 81% of the vehicles were successfully geocoded (the match rate for the test dataset). *Jointg.c* writes three ASCII files, one for engine starts, one for running exhaust, and one for the SCF approach. *Sz-tg.aml* reads the outputs of the C programs and converts the files to dbase files.

Few errors are generated with this module; however, many errors occur in the pre-processing data development stage. These and other errors are discussed in Chapter 4. As much as 40% of vehicle representation is lost due to decoding errors. Errors resulting from the address matching process further degrade the spatial quality; however, the use of distributions of groups may lessen the impact.

4.3.4. Major Road Technology Groups Module

The purpose of the major road technology group module is to estimate on-road fleet distributions. The major road technology groups module reads the zone-based *retg.dbf, regiontg.dbf* and *scftg.dbf* (Figure 4.12 shows the module). By utilizing the *sz* and *mr* coverages, it develops road segment specific fleet distribution estimates. This is accomplished in the *mr-tg.aml* process by determining a 'local' fleet distribution and combining that with a regional fleet distribution (Figure 4.13). The local fleet is defined as an aggregation of vehicles registered within a 3 km radius of the road segment. This process takes the longest of all of the modules because it has to develop the combined distribution for each of multiple thousands of road segments individually. The output of this process is a single dbase file called *mrtg.dbf* containing the road segment id and the technology group percentages.

68

Significant errors can occur in this process due to the unvalidated nature of the assumptions. The assumption that the on-road fleet can be predicted by combining distributions from a region and a 'user-defined' local fleet is loosely based on research completed by Tomeh, 1996. Some evidence described in Chapter 6 is available that defends the estimate.

4.3.5. Engine Start Activity Module

˙ The engine start activity module predicts the number of engine starts by *sz* poly and by hour of the day. Figures 4.14 and 4.15 represent the flow of the engine start activity estimation process. This is accomplished by reading the TAZ trip generation results file (*taz.dat*) and disaggregating the trips to *sz* zones using landuse, housing units, zone size, and school / university landmarks. All home-based trip origins are assigned to residential land uses based on household density. Shopping return trip origins are allocated to commercial land uses based on area. University return trip origins are allocated to zones that contain university landmarks based on enrollment. Grade school return trip origins are allocated to zones that contain grade school landmarks based on enrollment. Work return trip origins are allocated to non-residential zones based on area. If conditions exist that do not allow allocation of known trips using the above rules, trips are allocated to all zones based on area. Once trips are allocated to *sz* polys, they are disaggregated to hours of the day using hourly factors developed from Atlanta surveys for each trip purpose. The output of the module is a dbase file of total trip origins by zone and time of day.

Error in this process is largely the result of poor input data quality. The use of the travel demand forecasting values defines the highly aggregate original form of the data. A major missing component is the inclusion of intra-zonal and external travel patterns. Engine starts that result from individuals that go to destinations within the TAZ will not be accounted for. Intra-zonal / external trips that originate outside the study area and stop inside the site, will also not be represented.

4.3.6. Major Road Running Exhaust Activity Module

The running exhaust activity module determines hourly traffic conditions on all major road segments found in the *mr* line coverage. Figures 4.16 and 4.17 illustrate the process. Predicted conditions from the travel demand forecasting model output file (*tdfn.dat*) are combined with an hourly factor to predict road segment specific hourly volume, average speed, and LOS. The AML process also reads the grade ASCII files, joins them into a point coverage, assigns each point to the closest major road segment, and summarizes the grade points into five intervals. The data summaries of grade, static conditions, and dynamic conditions are written to a dbase file called *mr-act.dbf* storing results by the *arid* key field.

Besides input data error, the incorporation of road grade causes some significant problems. While the anticipated technique for collecting road grade (GPS) results in absolute positional accuracy that exceeds the road database, it is the relational accuracy that allows useful locational data to be collected. The process of snapping points to the lines in order to develop the relational structure can produce poor results around intersections and close parallel roads. Further analysis on this subject is available in Chapter 6.

4.3.7. Minor Road Activity Module

The minor roads activity module develops estimates of the mean travel time and hourly trips occurring within each of the minor zones. Figures 4.18 and 4.19 illustrate the process. The shortest path is determined between the *sz* polygon centroids and the closest node of the major road network. The aggregate travel time is developed along with summaries of the hourly trip production at the zone (from the *sz-act.dbf* file). The final output is a dbase file called *mz-act.dbf* containing an aggregate travel time value and hourly trip production for each minor zone.

The potential errors in estimating the travel time are substantial. However, the alternative use of centroid connectors or zonal surface area would not allow a measure of network configuration within the zone.

4.3.8. Engine Start Emissions Module

The engine start emissions module predicts hourly CO, HC and NOx emissions or engine starts in *sz* polygons. Figures 4.20 and 4.21 illustrate the process flow. Generally, engine start technology groups are combined with emission rates and estimates of the number of hourly trips to develop the emission estimates. Output is written to a dbase file and stored in the *em* directory.

Emissions are elevated at the start because catalyst control equipment needs to operate at high temperatures. While the actual emissions are dispersed as the vehicle travels for the first few minutes, the model allocates the entire start portion to the origin *sz*. Therefore, spikes of high emissions may be estimated at high population density locations when in actuality the emission production at that location would be lower. This source of error in position is significant. Although the 'puff' allocation is significant in identifying the original sources of high start emissions, the actual location that the emissions entered the atmosphere is misrepresented.

70

4.3.9. Minor Road Running Exhaust Emissions

The minor road running exhaust module has the task of predicting hourly emissions of CO, HC, and NOx, given the activity conditions provided by *mz-act.dbf*. Figures 4.22 and 4.23 show the module flow. While information regarding technology groups is available for minor zones, traffic flow and modal conditions are not. Speed and acceleration data have not been collected to provide clues for determining local road profiles. For this reason, the module uses an average high and normal emitter running exhaust emission rate for the three pollutants of concern. As soon as data become available, the minor zone emissions can follow a similar track as the major road emissions model.

4.3.10.   Major Road Running Exhaust Emissions

The major road running exhaust emission's module calculates an aggregate hot-stabilized and enrichment emission estimate for all major roads. The module is represented in Figures 4.24 and 4.25. Estimates of vehicle activity and on-road technology groups are combined with technology and operating mode specific emission rates to develop hourly, road segment estimates of CO, HC, and NOx. Matrices of modal variables are available as lookup files. These files match the format of the speed / acceleration profiles, and consist of zeros and ones. The two files are multiplied together, and the resulting speed and acceleration file bins are summed. The values represent the fraction of activity in that particular mode.

4.3.11.   Gridded Emissions

The gridded emissions module develops the total emission estimates by grid cell in a raster format. It has the task of overlaying a user-defined grid polygon coverage with the *sz, mr,* and *mz* coverages, and aggregating all the emission estimates (weighted by the area or length). All of the emission mode estimates (engine start, minor zone, major road, and SCF) are converted to raster datasets. The engine start, minor zone running exhaust, and major road running exhaust are summed together to develop estimates of total hourly CO, HC, and NOx. The process for obtaining gridded emissions is shown in Figures 4.26 and 4.27.

4.4.   Conclusion

Chapter 4 described how the applied model design from Chapter 3 was translated into a functional computer model. Detailed descriptions of the input files and system structure are provided. Flow charts for the various modules and programs

71

are provided as well. The model includes both SCF emission rates and modal emission rates. Engine starts and running exhaust are calculated separately for normal and high emitting vehicles. Gridded, hourly emission estimates are produced.
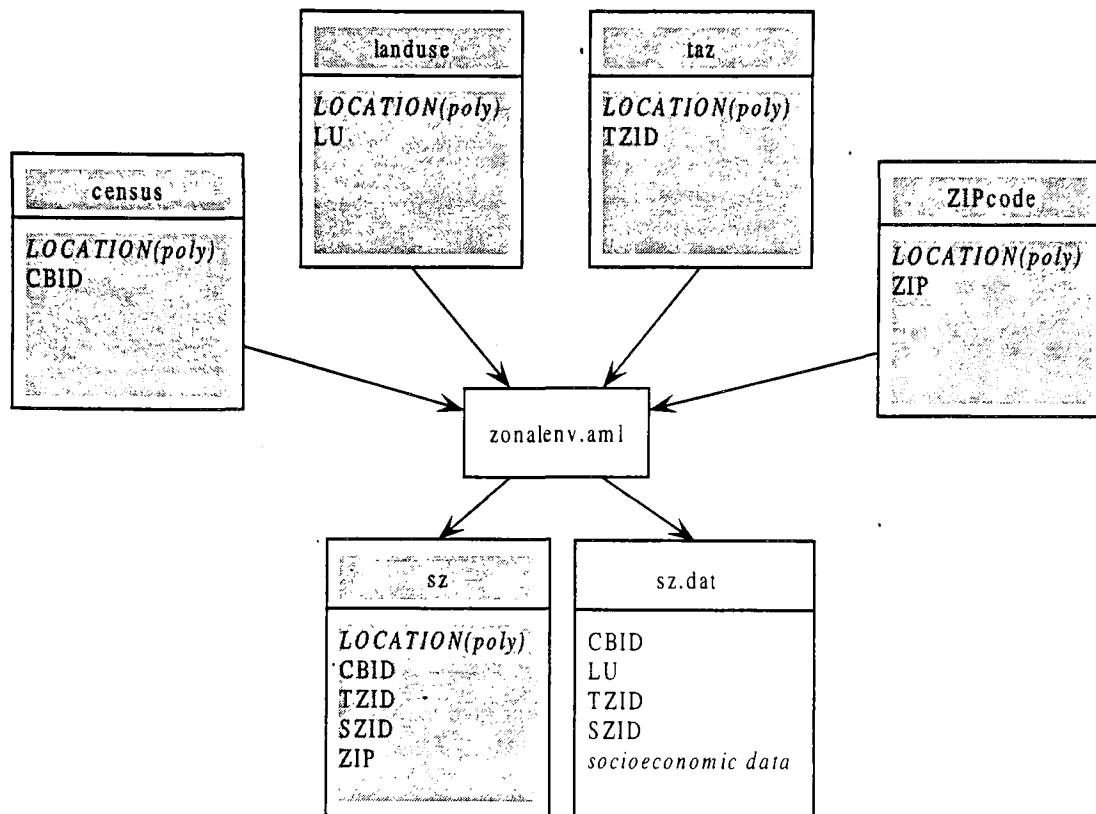
Figure 4.2 - Zonal Environment Entities
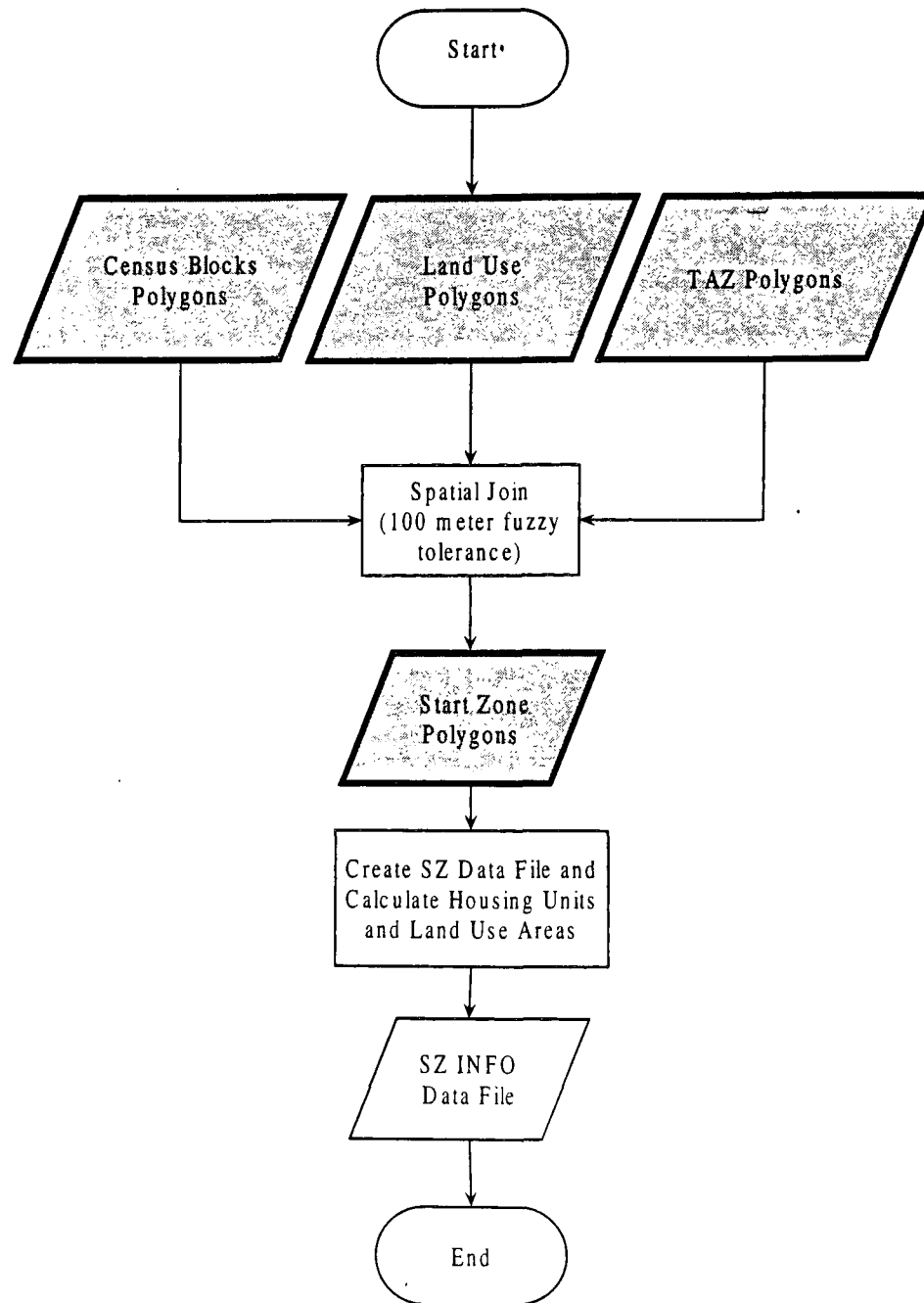
Figure 4.3 - *Zonalenv.aml* Flow Chart
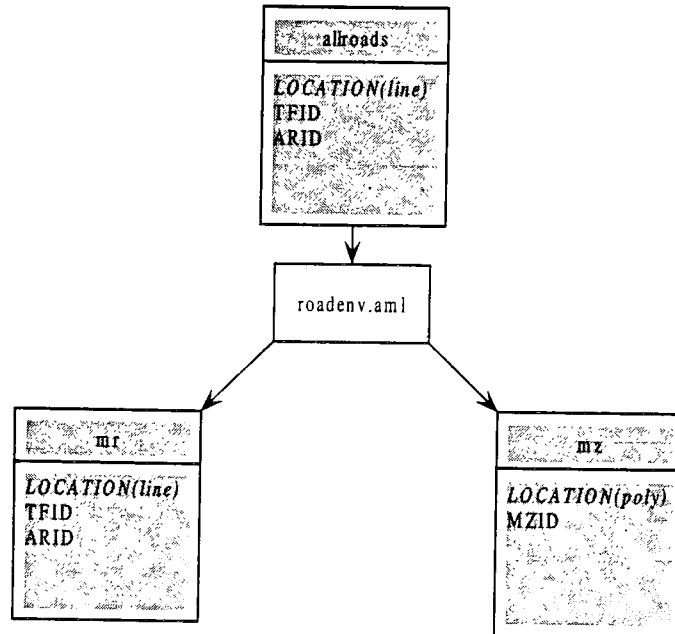
Figure 4.4 - Road Environment Entities
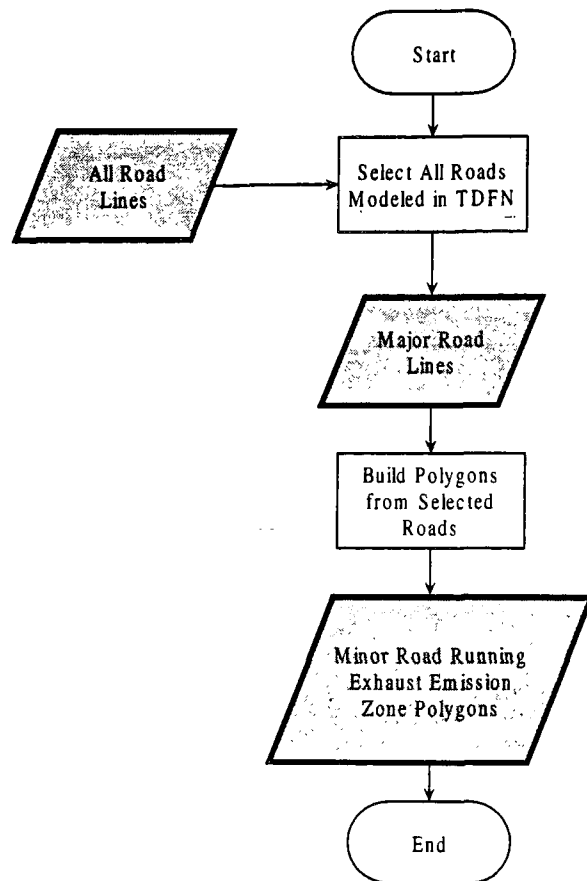
Figure 4.5 - *Roadenv.aml* **Flow Chart**
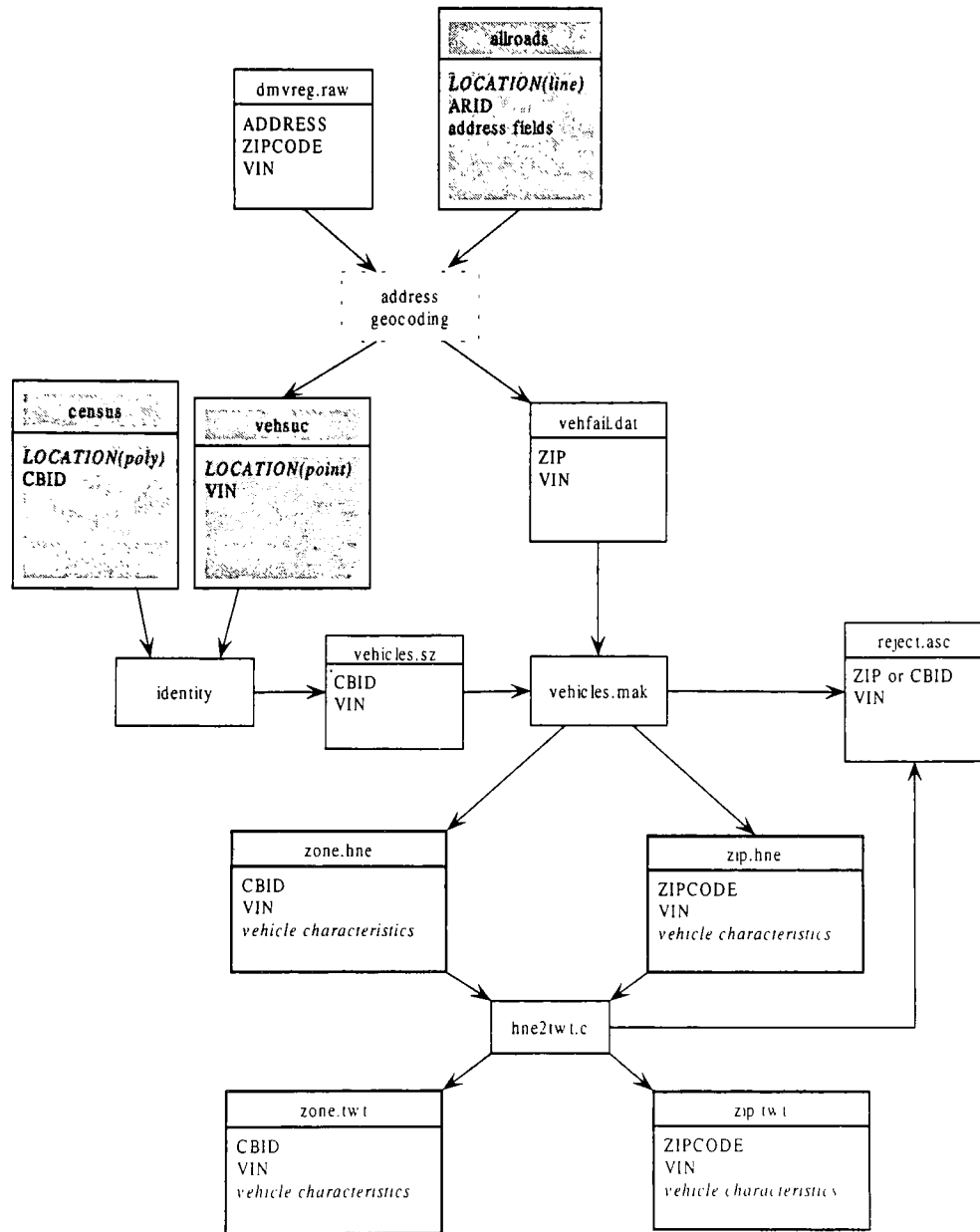
Figure 4.6 - Vehicle Characteristic Entities

zone.twt

CBID
VIN
vehicle characteristics

ZIP.twt

ZIPCODE
VIN
vehicle characteristics

techgr.c

zones.tg	zonre.tg	ZIPes.tg	ZIPre.tg	zonscf.tg	ZIPscf.tg
CBID	CBID	ZIPCODE	ZIPCODE	CBID	ZIPCODE
frequency	frequency	frequency	frequency	frequency	frequency
con%	con%	con%	con%	my%	my%
coh%	coh%	coh%	coh%		
hcn%	hcn%	hcn%	hcn%		
hch%	hch%	hch%	hch%		
non%	non%	non%	non%		
noh%	noh%	noh%	noh%		

szzp.asc

szid
zpid
szarea
zparea
match rate

jointg.c

es.tg	re.tg	scf.tg.
CBID	CBID	CBID
frequency	frequency	frequency
con%	con%	my%
coh%	coh%	
hcn%	hcn%	
hch%	hch%	
non%	non%	
noh%	noh%	

Figure 4. 7 - Technology Group Entities

78
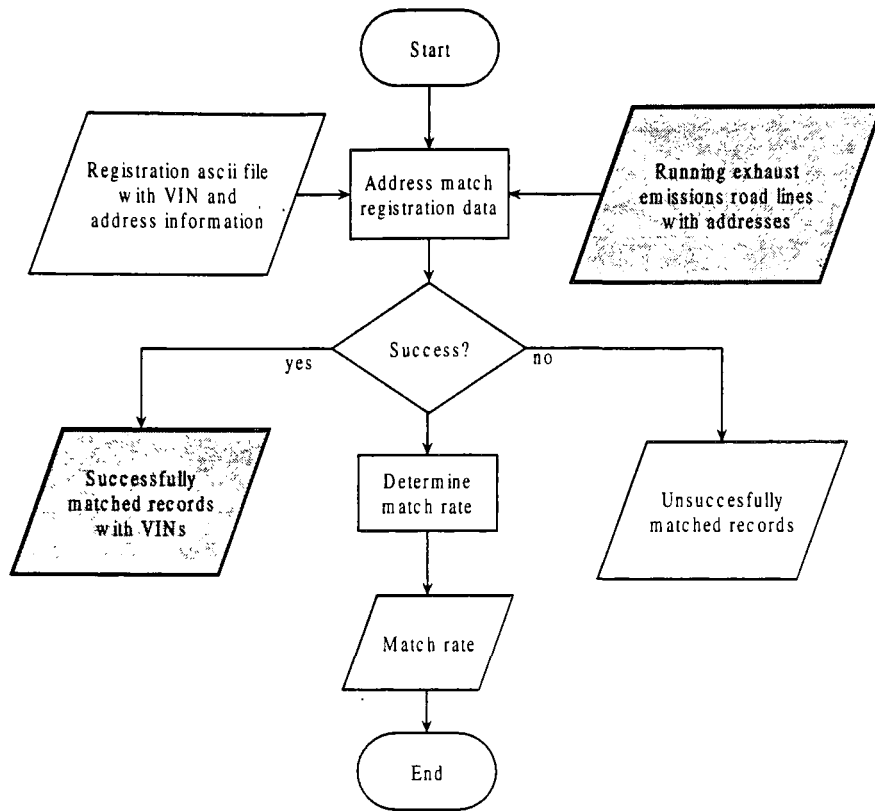
Figure 4. 8 - Address Matching Flow Chart

```
                            ┌─────────┐
                            │  Start  │
                            └─────────┘

┌──────────┐   ┌──────────┐   ┌──────────┐   ┌──────────┐   ┌──────────┐
│Start Zone│→  │  Decode  │ ← │   VIN    │ → │  Decode  │ ← │ ZIPcode  │
│Vehicle   │   │ Matched  │   │ Decoder  │   │Unmatched │   │Vehicle   │
│File      │   │ Vehicles │   │          │   │Vehicles  │   │File      │
└──────────┘   └──────────┘   └──────────┘   └──────────┘   └──────────┘

              Success? ──no──→ Reject File ←──no── Success?

              yes                                   yes

         Temporary Start                      Temporary
         Zone Vehicle File                    ZIPcode Vehicle File

         Lookup Test  ←── CERT        ──→    Lookup Test
         Weight           Lookup File        Weight

         Success? ──no──→ Reject File ←──no── Success?

         yes                                  yes

         Start Zone                           ZIPcode
         Vehicle File                         Vehicle File

                            ┌─────────┐
                            │   End   │
                            └─────────┘
```

Figure 4. 9 - *Vehicles.mak* Flow Chart
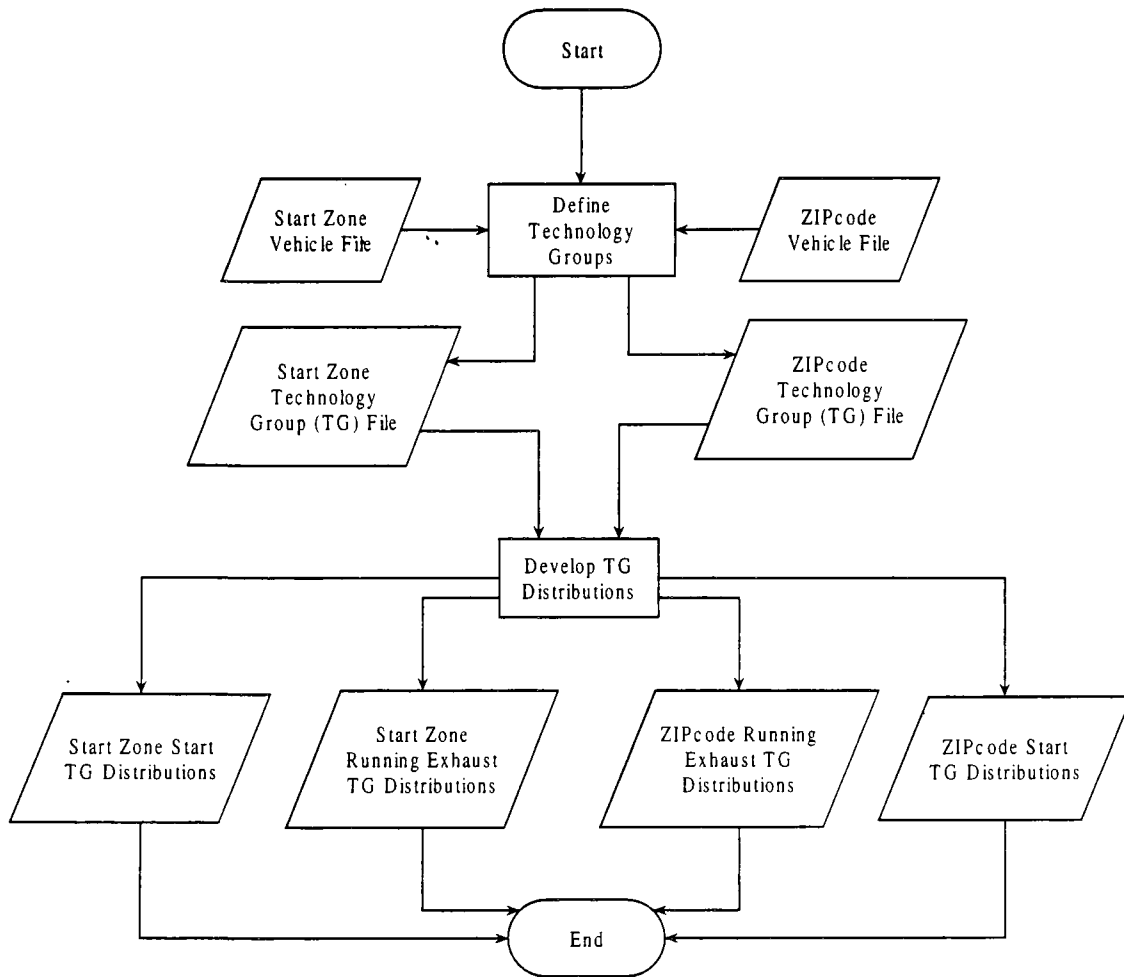
Figure 4. 10 - *Techgr.c* **Flow Chart**
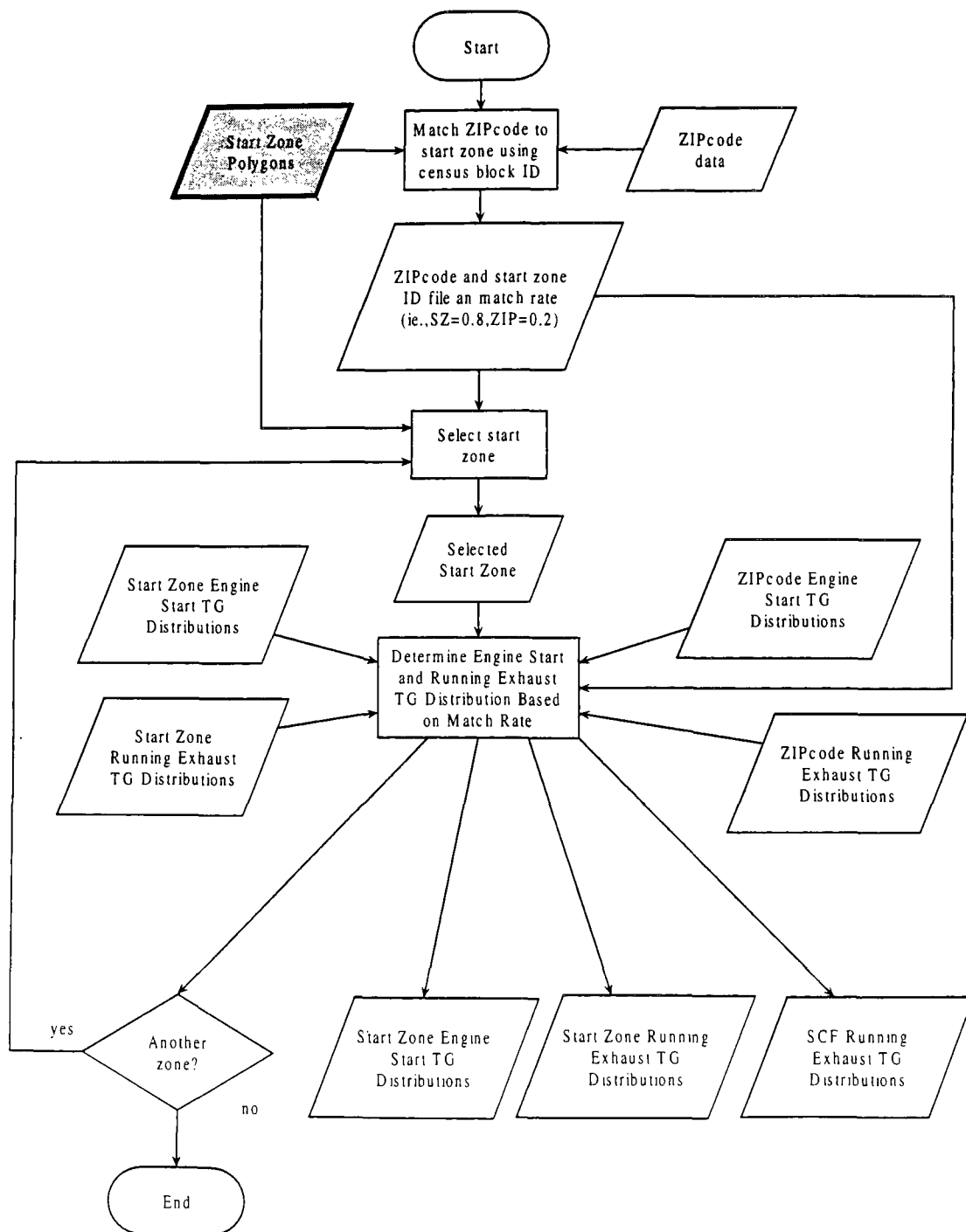
Figure 4.11 - *Jointg.c* **Flow Chart**
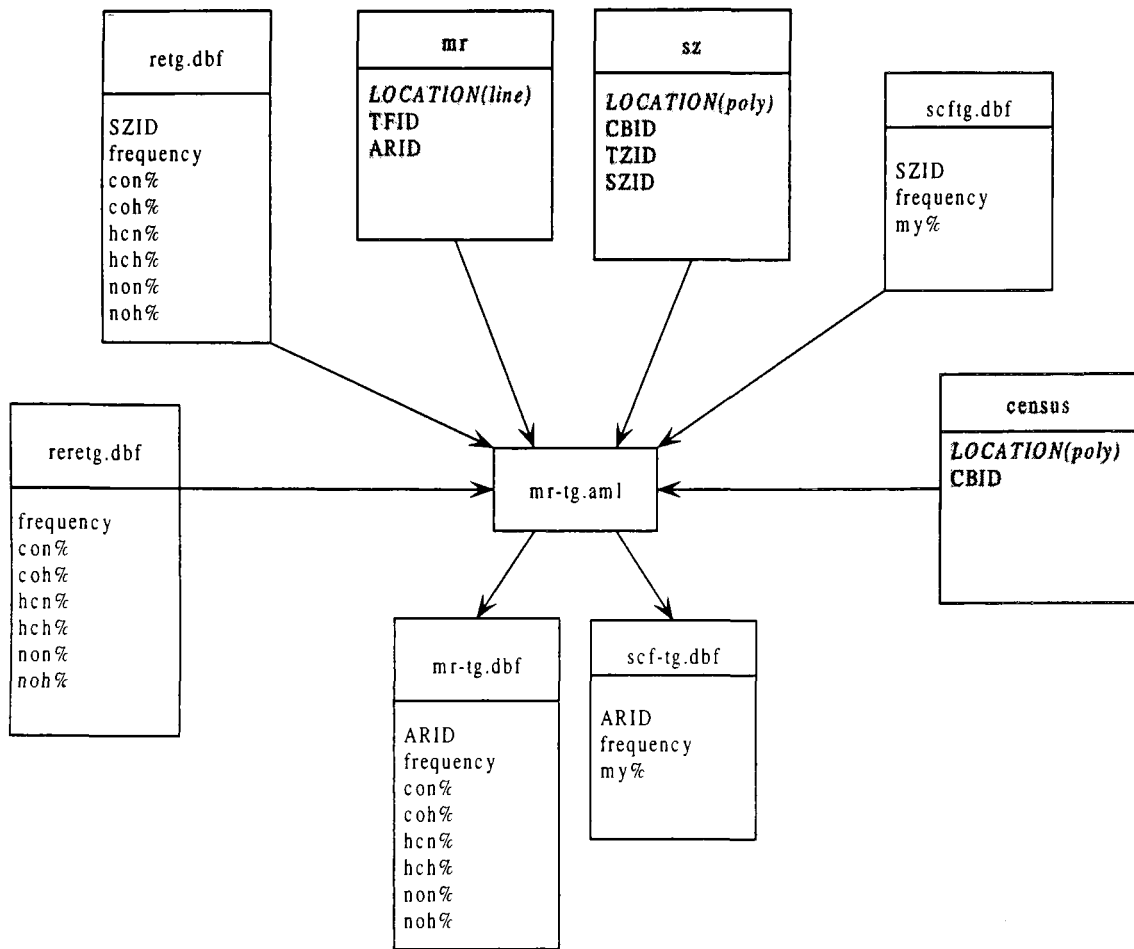
Figure 4.12 - On-Road Technology Group Entities
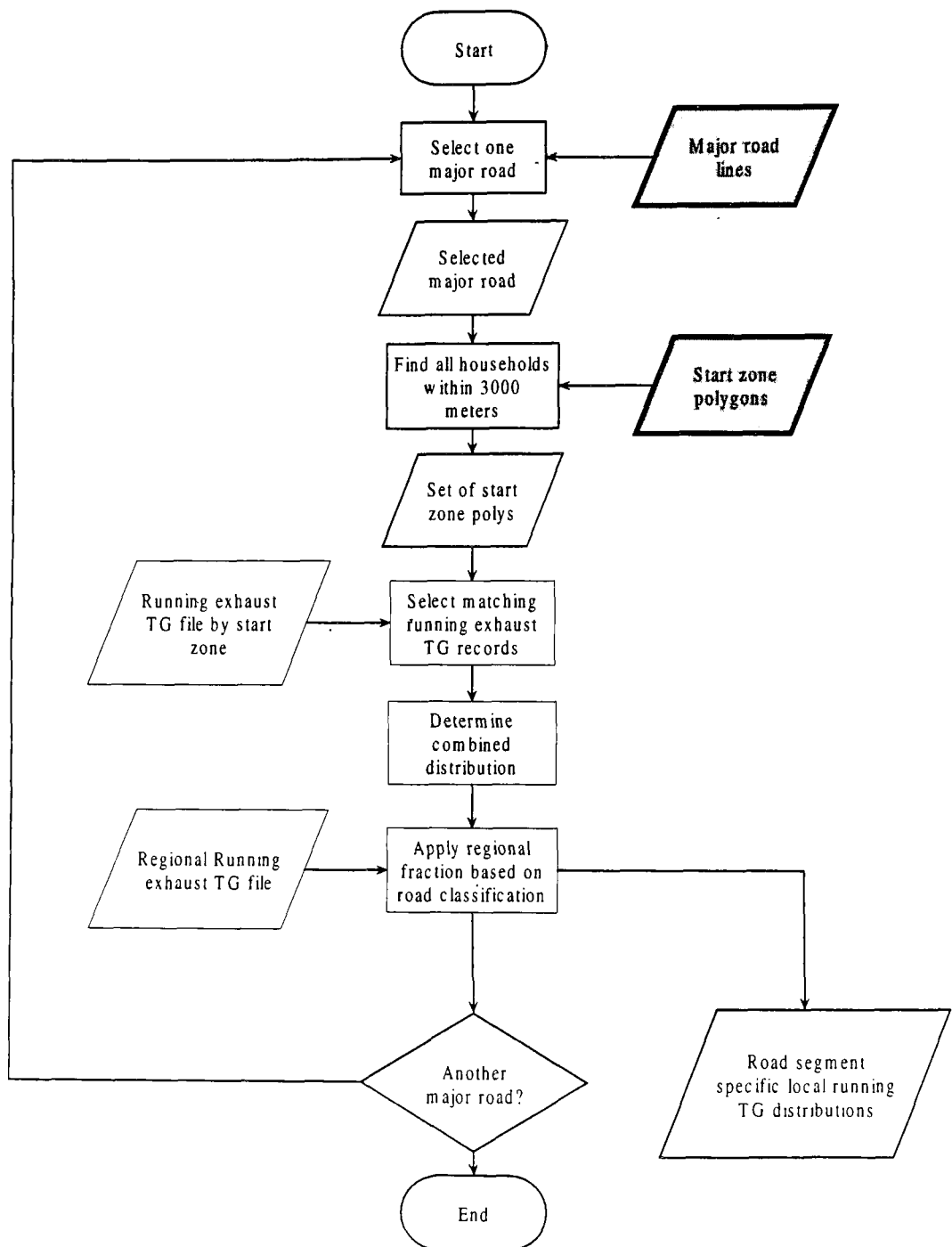
Figure 4.13 - *Mr-tg.aml* Flow Chart

```
        ┌──────────────┐   ┌──────────────┐
        │    sz.dat    │   │  landmarks   │
        ├──────────────┤   ├──────────────┤
        │ SZID         │   │ schools      │
        │ TZID         │   │ universities │
        │ landuse      │   │              │
        │ household data│  │              │
        └──────────────┘   └──────────────┘
┌──────────────┐                              ┌──────────────┐
│     taz      │                              │ temporal.lu  │
├──────────────┤                              ├──────────────┤
│ LOCATION(poly)│                             │ hourly       │
│ TZID         │                              │ distributions by│
│ trips by type │                             │ trip type    │
│              │         ┌──────────┐         └──────────────┘
│              │ ──────→ │sz-act.aml│ ←───────
│              │         └──────────┘
└──────────────┘              │
                              ↓
                       ┌──────────────┐
                       │  sz-act.dbf  │
                       ├──────────────┤
                       │ SZID         │
                       │ hourly trips │
                       └──────────────┘
```
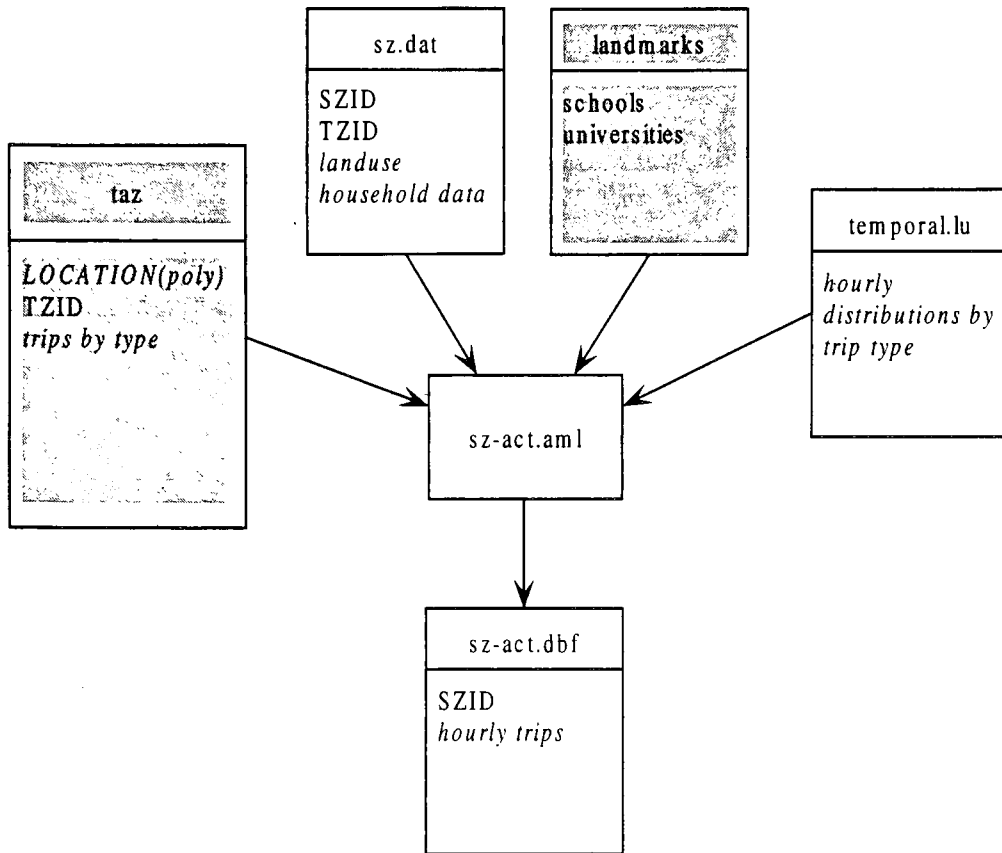
Figure 4.14 - Start Zone Activity Entities
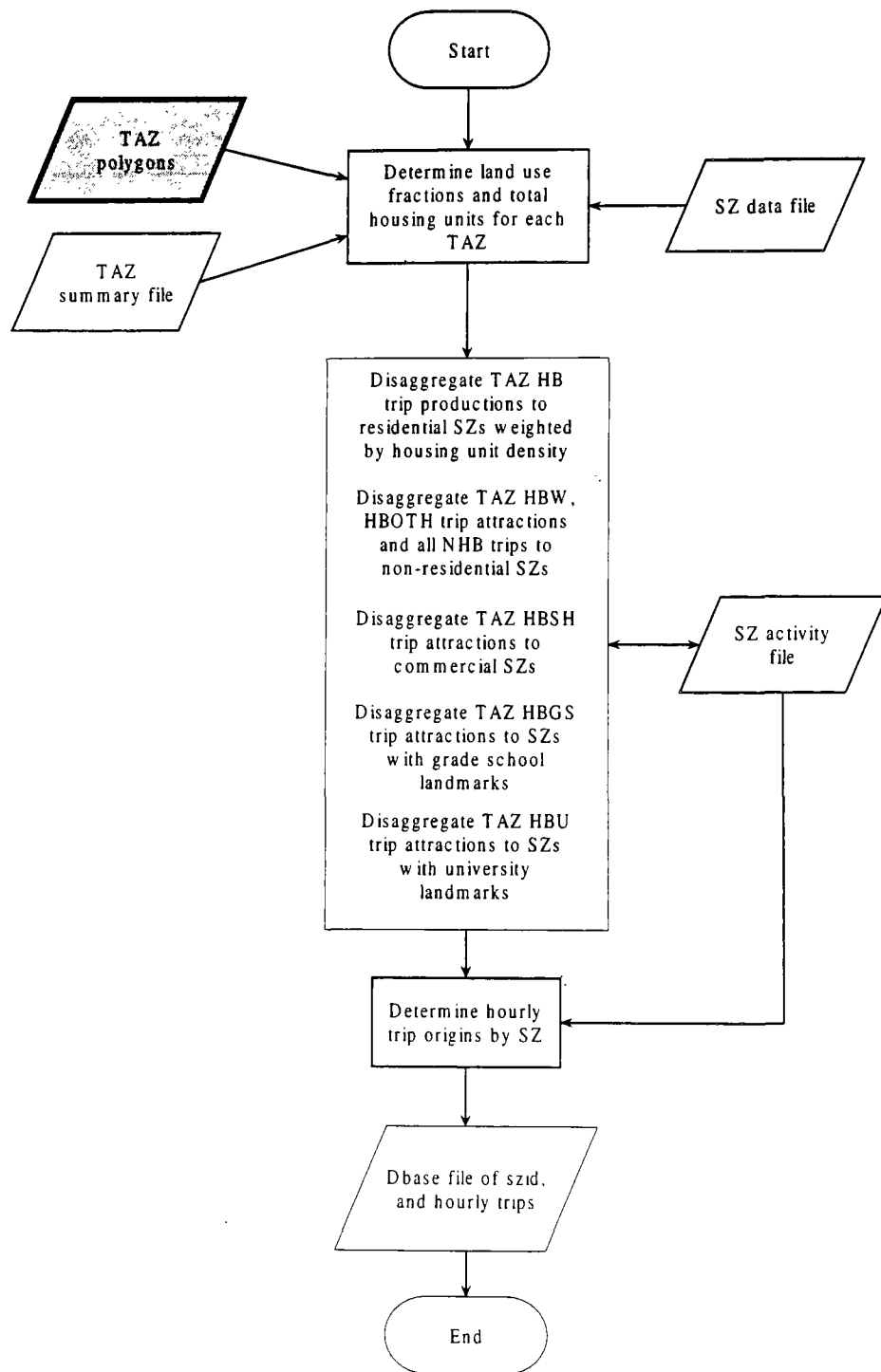
Figure 4.15 - *Sz-act.aml* Flow Chart

```
+--------------+    +--------------+    +--------------+
|  grade.xy    |    |  grade.gr    |    |  tdfn.dat    |
+--------------+    +--------------+    +--------------+
| G-ID         |    | G-ID         |    | TFID         |
| x            |    | grade        |    | predicted link|
| y            |    | time         |    | level traffic|
|              |    | flag         |    | parameters   |
+--------------+    +--------------+    +--------------+

+--------------+                              +-------------------+
|     mr       |                              | temporal.factors  |
+--------------+                              +-------------------+
| LOCATION(line)|                             | hourly on-road    |
| TF-ID        |         +------------+       | factors           |
| RD-ID        |-------->| mr-act.aml |<------ |                   |
| GD-ID        |         +------------+       +-------------------+
+--------------+               |
                               v
                      +--------------+
                      |  mr-act.dbf  |
                      +--------------+
                      | ARID         |
                      | geometrics   |
                      | hourly volume|
                      | hourly v/c   |
                      | hourly avg speed|
                      +--------------+
```
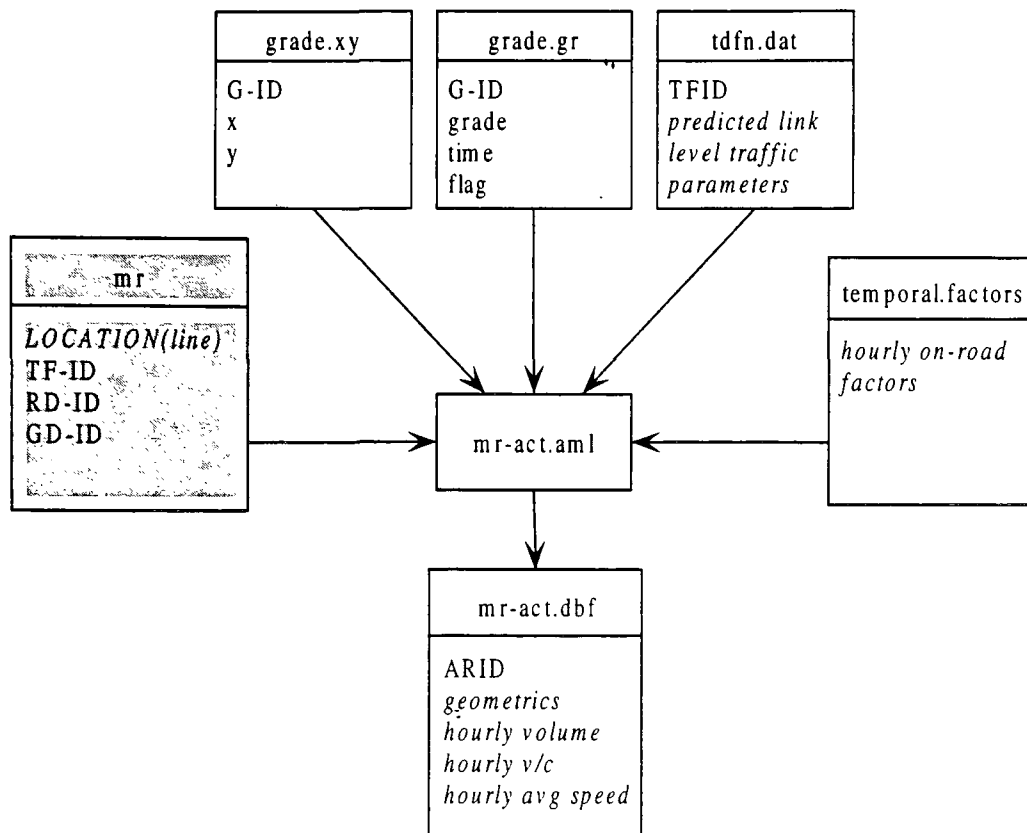
Figure 4.16 - Major Road Activity Entities
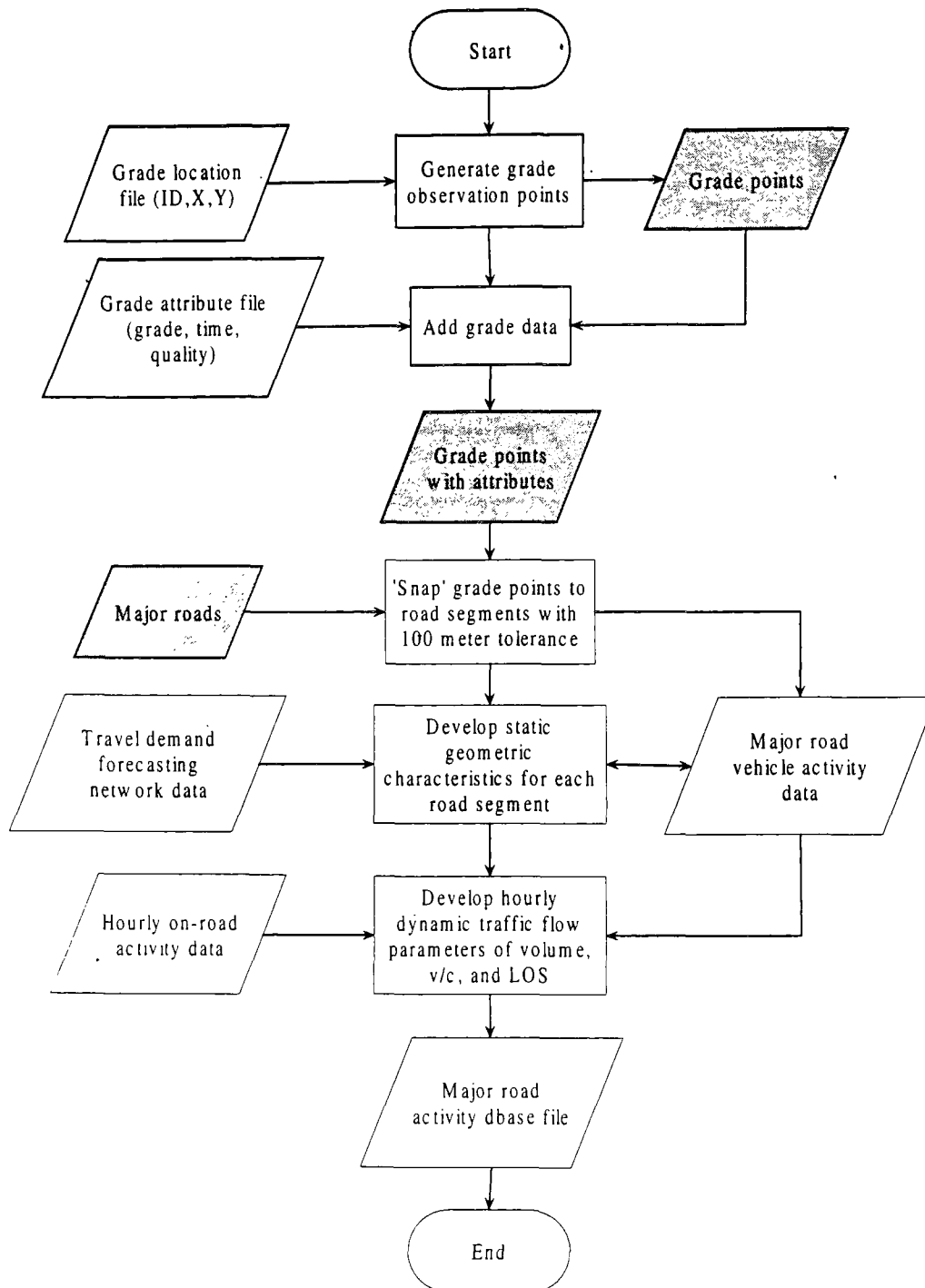
Figure 4.17 - *Mr-act.aml* **Flow Chart**
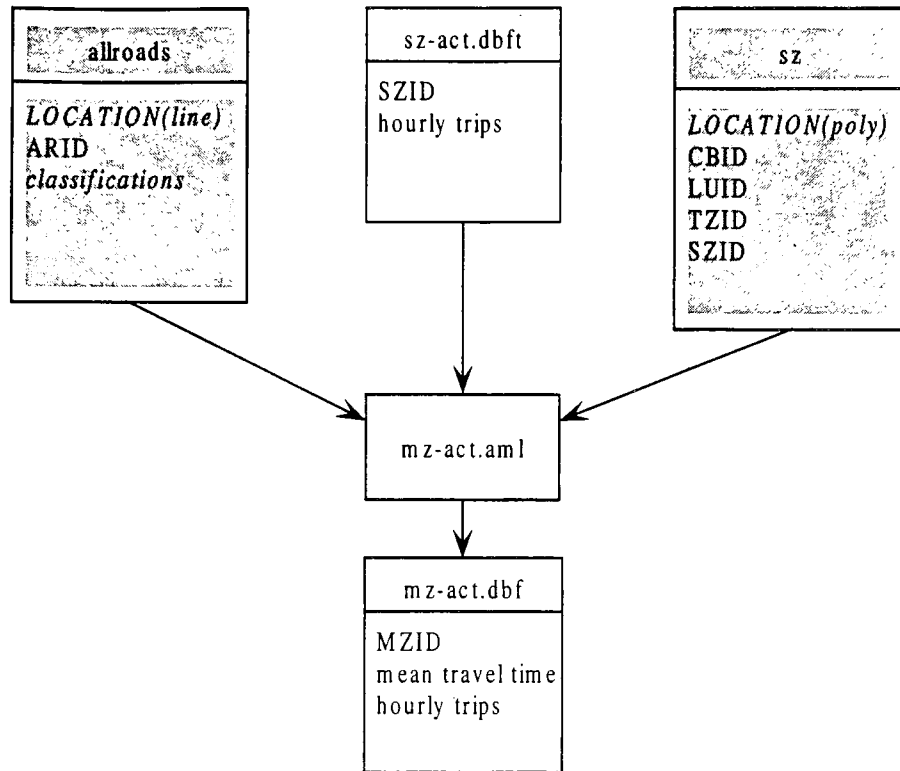
Figure 4.18 - Minor Road Activity Entities
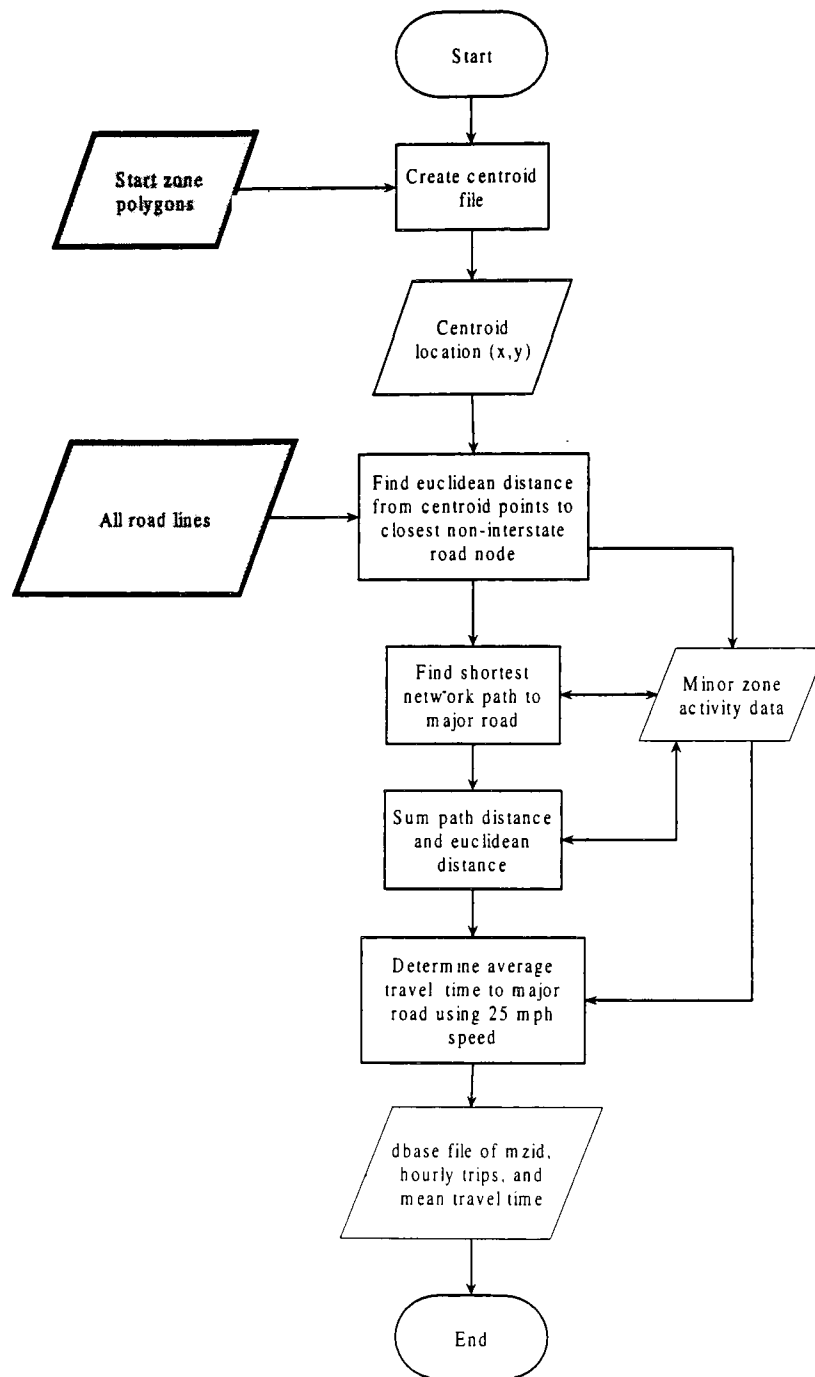
Figure 4.19 - *Mz-act.aml* Flow Chart
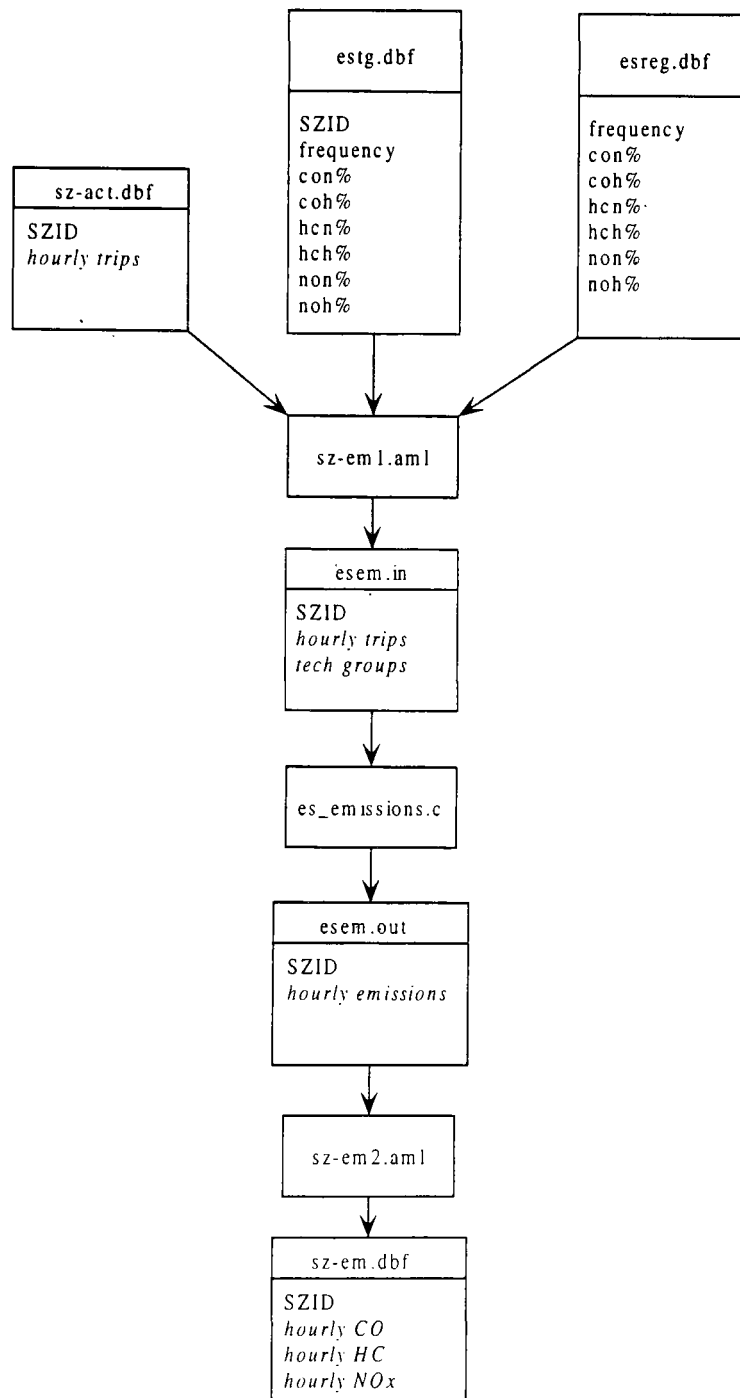
Figure 4.20 - Start Zone Emissions Entities
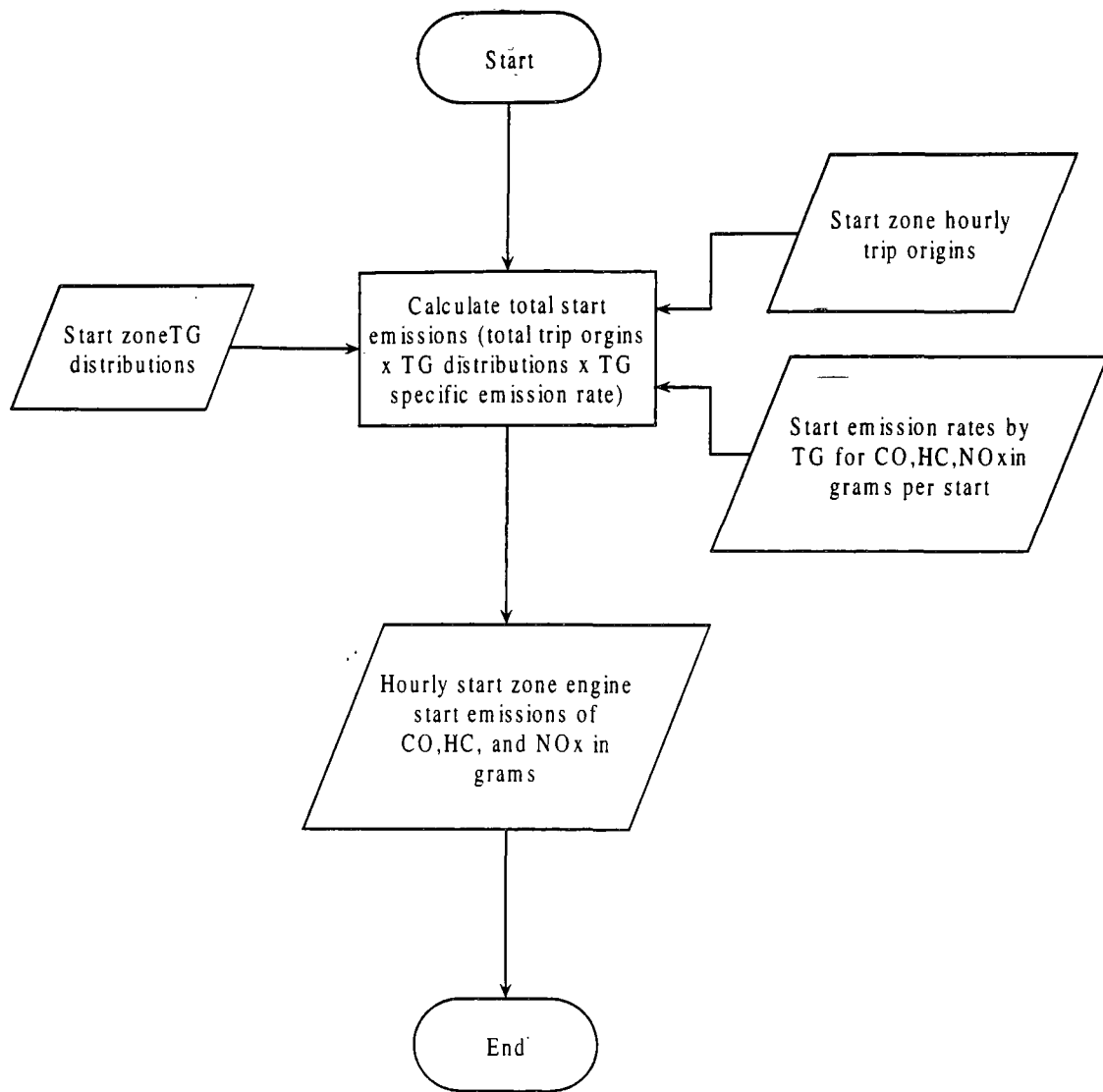
Figure 4.21 - *Es_emission.c* Flow Chart
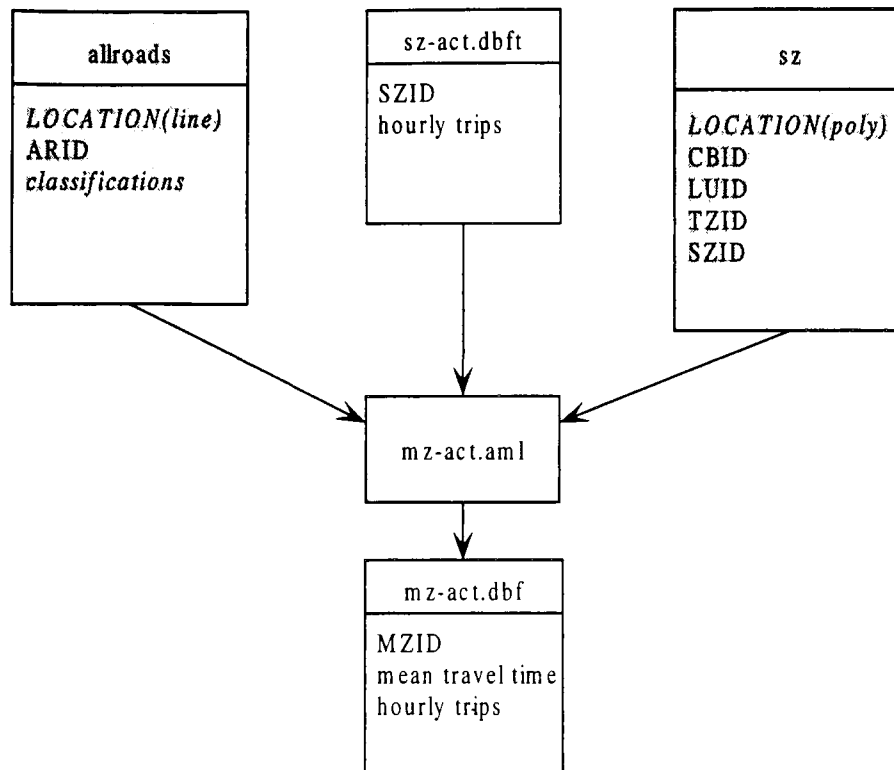
Figure 4.22 - Minor Zone Activity Entities
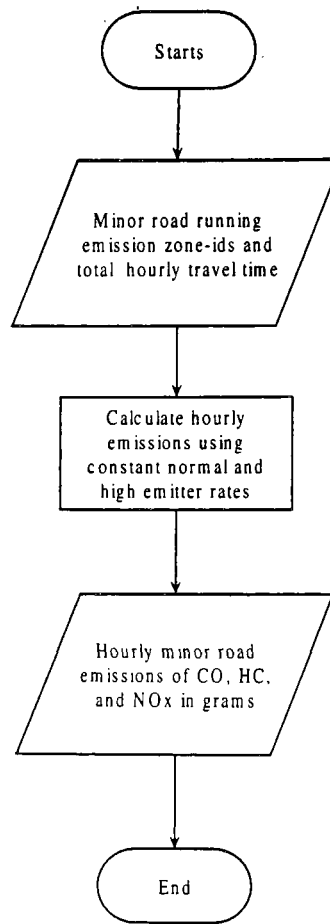
Figure 4.23 - *Mz-em.aml* Flow Chart

rereg.dbf

con%
coh%
hcn%
hch%
non%
noh%
.   .

retg.dbf

TFID
ARID
con%
coh%
hcn%
hch%
non%
noh%

mr-act.dbf

ARID
grade
hourly volume
hourly v/c
hourly ff speed

mr-em1.aml

re-em.in

ARID
length
class
grade
hourly volume
hourly v/c
hourly ff speed

modal(x).mat

Speed by Accel

sa(x).mat

Speed by Accel

re_emissions.c

reem.out

ARID
hourly CO
hourly HC
hourly NOx

mr-em2.aml

mr-em.dbf

ARID
hourly CO
hourly HC
hourly NOx

Figure 4.24 - Major Road Emissions Entities

95
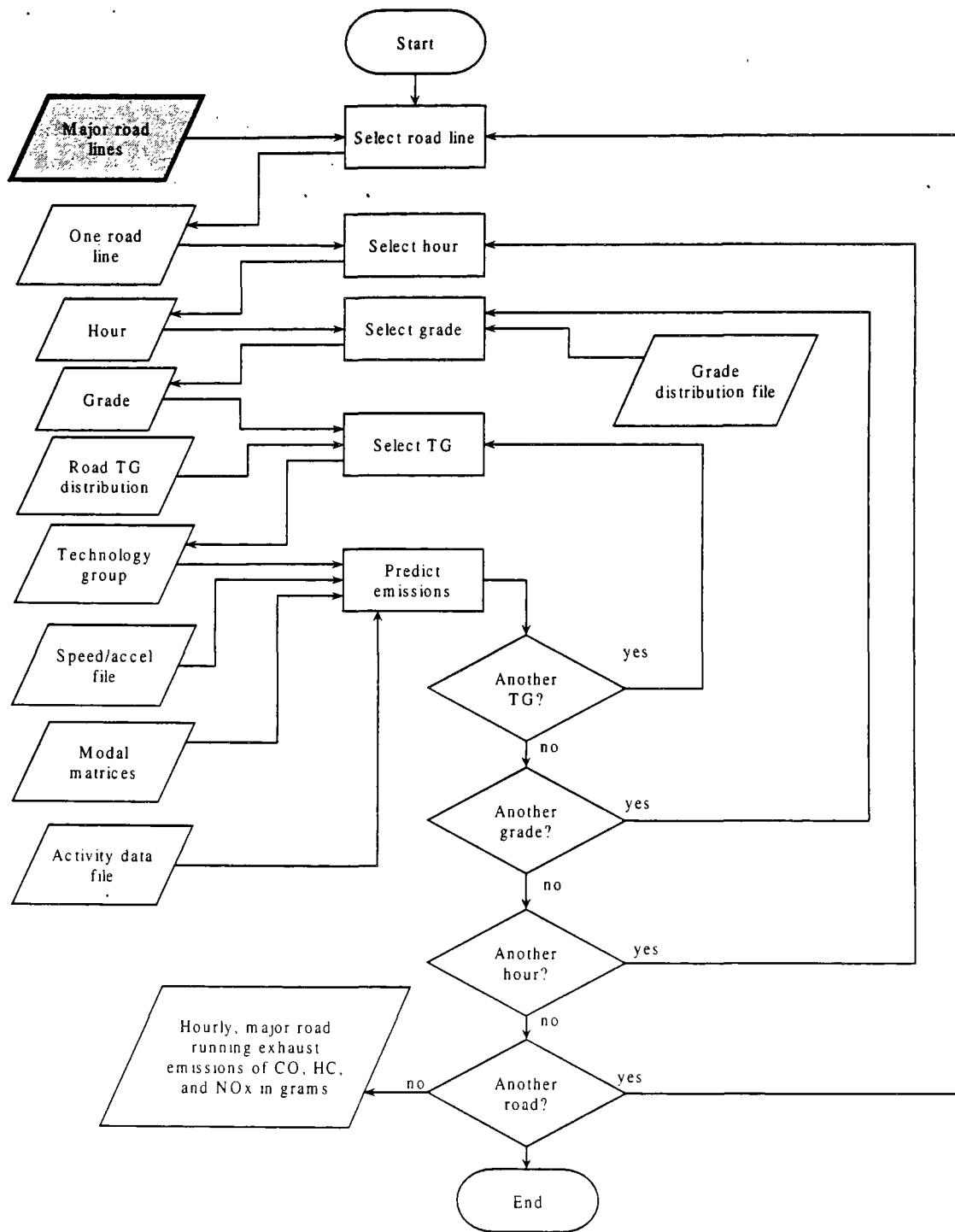
Figure 4.25 - Re_emissions.c Flow Chart
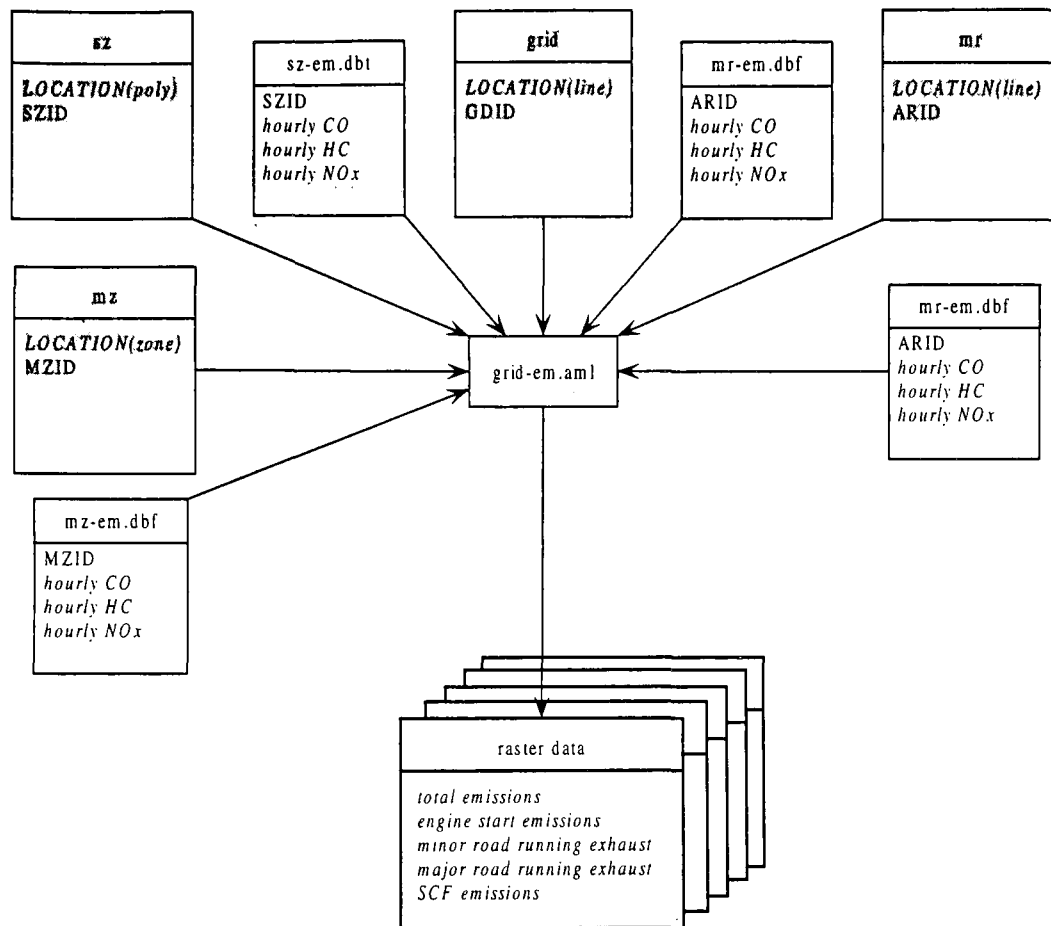
Figure 4.26 - Gridded Emissions Entities
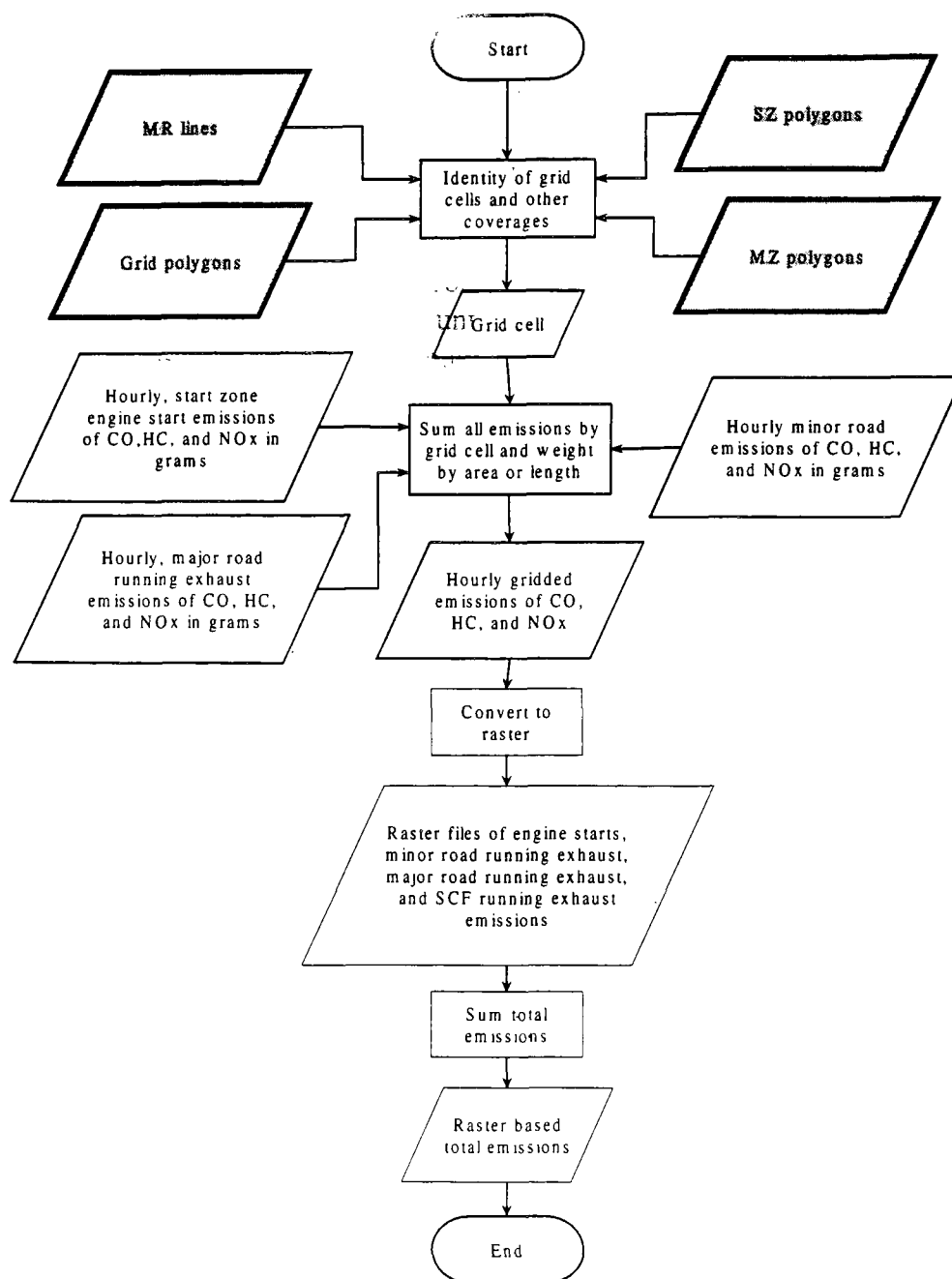
Figure 4.27 - *Grid-em.aml* **Flow Chart**

5. MODEL DEMONSTRATION

This chapter demonstrates model capabilities. By applying the model described in Chapter 4 for a study area, an estimation of emissions will be developed. While the conceptual value of the model is revealed through its design parameters, a demonstration provides insight into the model's practical value. The model will predict grams of CO, HC, and NOx, with a spatial resolution determined by the user. In this case, hourly gridded estimates are provided for 100 meter, 250 meter, 500 meter, and 1km grid cells.

The study area is a 100 square kilometer portion of the Atlanta, Georgia metropolitan area and is shown in figure 5.1. The 10 kilometer by 10 kilometer slice of the northeast suburbs was selected as a sample study area because it contains: diverse landuse, variable densities of development, a major interchange, major north-south arterials (leading to the CBD) and an interstate known for congestion (northern portion of I-285).

The following input datasets were used:

- *1995 Georgia Department of Motor Vehicles Registration Dataset*
- *1990 US Census Summary Tape File 3a*
- *1994 US Census TIGER File*
- *1995 Updated TIGER Road Database*
- *1995 Atlanta Regional Commission's (ARC) Traffic Analysis Zones*
- *1995 ARC's Travel Demand Forecasting Network*
- *1995 ARC's Land Use Data*
- *1996 ARC's ARCMAP Road Database*
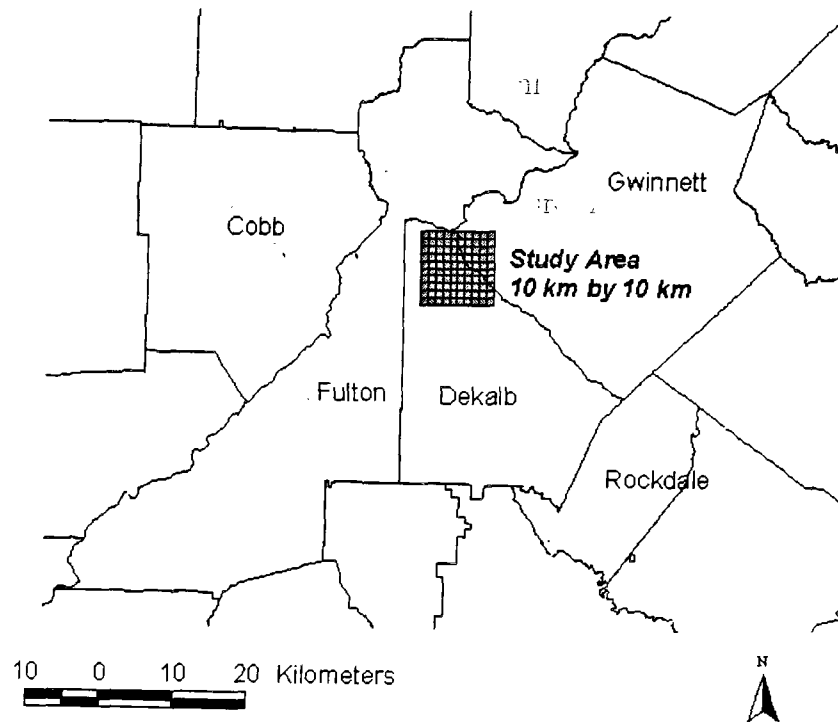
Figure 5.1 - Model Study Area Site Map

5.1. Preprocessing

Several preprocessing steps were completed to prepare model input data. Most model implementation efforts will require substantial preprocessing steps due to the variability of data availability. In this sample case, preprocessing was needed for the road network and vehicle characteristics development. Other data required simple conversion or transformation. Several preprocessing steps were needed.

5.1.1. Vehicle Characteristics

The Georgia DMV Registration Database is protected under privacy regulations. The Georgia Tech Air Quality Laboratory (AQL) has data access permission for research (under contract). To protect the privacy of vehicle owners, a three step 'double-blind' procedure was used to provide vehicle data. First, data consisting of owner address information and a unique identifier were transferred from the AQL. Second, the data were address-matched (in the GIS), and aggregated to Census Block or ZIP code. A file of the unique identifier and the zonal identifier was transferred back to AQL. Third, a file of the zonal identifier and vehicle identification number (VIN) was returned from AQL, thereby providing a spatially-resolved, decodable, file of vehicles.

Address-matching provides the ability to develop vehicle registration information at a better spatial resolution than provided by ZIP codes. The vehicle file was address-matched using two road datafiles, the ARC's ARCMAP Road Database and the road database. The ARCMAP database provided comprehensive coverage for the entire metropolitan Atlanta area. The road database provided higher spatial accuracy, but did not cover the entire area. The road database was used first, to maximize spatial accuracy, and then the ARCMAP was used to maximize comprehensiveness. Actual vehicle locations were offset by 30 meters to ensure that vehicles would not fall on zonal boundaries when aggregated. For a successful match, the ZIP codes must match. Slight errors in spelling were allowed. The road database resulted in a 63% match rate. Another 18% were matched using the ARCMAP road database, for a total of 81%. Therefore, two files were created; one of matched vehicles (81%) and one of unmatched vehicles (19%) (discussed further in chapter 6). The matched vehicles were aggregated to US Census Block (census) polygons as described in the previous paragraph.

The two files were sent to the *vehicles.mak* PC process [see figure 4.6]. This process was developed by John Leonard, William Bachman, and Osama Tomeh· at

Georgia Tech in 1996 [Tomeh, 1996]. The process reads a file consisting of a record identifier and VIN. During the process, the VIN is decoded using software developed by Radian International Corporation, the vehicles emission test weight was added using a lookup table, vehicles are flagged as being high or normal emitters, and emission-specific characteristics are written to an output file (record identifier, vehicle characteristics, emitter types). Each of the files was sent through this process resulting in the two files required as inputs to the emissions model.

5.1.2. Conflation

Conflation is the process of combining two separate line datasets into a single dataset. In the model, the prognostic data (volumes and speeds) provided by the travel demand forecasting network are transferred to a road network that has better spatial accuracy. There is not a one-to-one correspondence between the two datasets' road segment representation, nor are there attribute fields that can create connectivity. The abstract spatial structure of the travel demand forecasting network prevents a clearly-defined locational connectivity. However, enough spatial definition exists that a manual link-by-link assessment can establish connectivity. The process of conflation is frequently used by transportation agencies to bring various linear datasets together.

The sample area's portion of the travel demand forecasting network consisted of 532 links, and the accurate road database consisted of 3602 road segments. Overlaying the datasets in the GIS (ARC/INFO) identified representational similarities. As individual travel model links were 'selected', corresponding accurate road segments were also selected. The NAVTECH roads were assigned an identifier field that could be used to transfer attributes (predicted volume, speed, etc.). Each of the 532 was processed in this manner. The resulting database is one of the required inputs into the model.

5.1.3. Other Steps

Numerous other steps were needed to fully prepare data for model running. The databases are all distributed in different formats, requiring conversion, transformation, and renaming. Atlanta area Census data (STF3A and TIGER) had to be selected, joined, and transformed to develop the zonal database containing detailed household information. The ARC's travel demand forecasting network had to be converted from an ASCII file to an ARC/INFO coverage using programs written by Wayne Sarasua, Xudong Jia, and William Bachman at Georgia Tech. The ARC's TAZs and land use were delivered in an ARC/INFO coverage. Other urban areas may have to develop customized strategies to get the input information in the format described in the previous chapter, and in the data dictionary found in the appendix.

102

5.2. Spatial Environment

The spatial environment consists of the ARC/INFO input coverages: *taz* (Atlanta Regional Commission's (ARC) traffic analysis zones), *census* (US Census blocks), *landuse* (ARC's landuse), ZIP*code*, and *allroads* (conflated road database). The spatial environment output coverages were: *sz* (engine start polygons), *mr* (major road running exhaust lines), and *mz* (minor road running exhaust polygons). Figures 5.2 and 5.3 demonstrate the connectivity.

Census
(CBID)

TAZ
(TZID)

Landuse
(LU)

SZ
(SZID
CBID
TZID
LU)
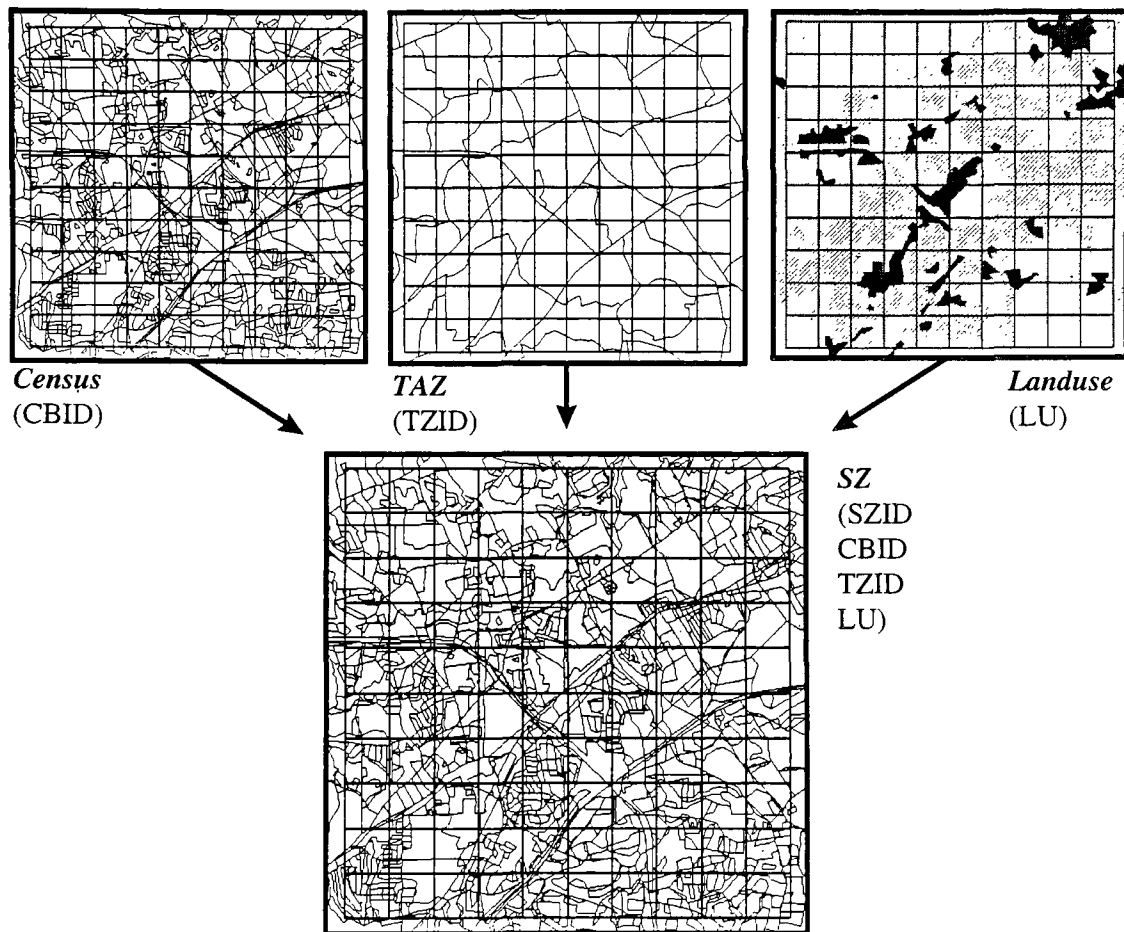
Figure 5.2 - Engine Start Zone Creation
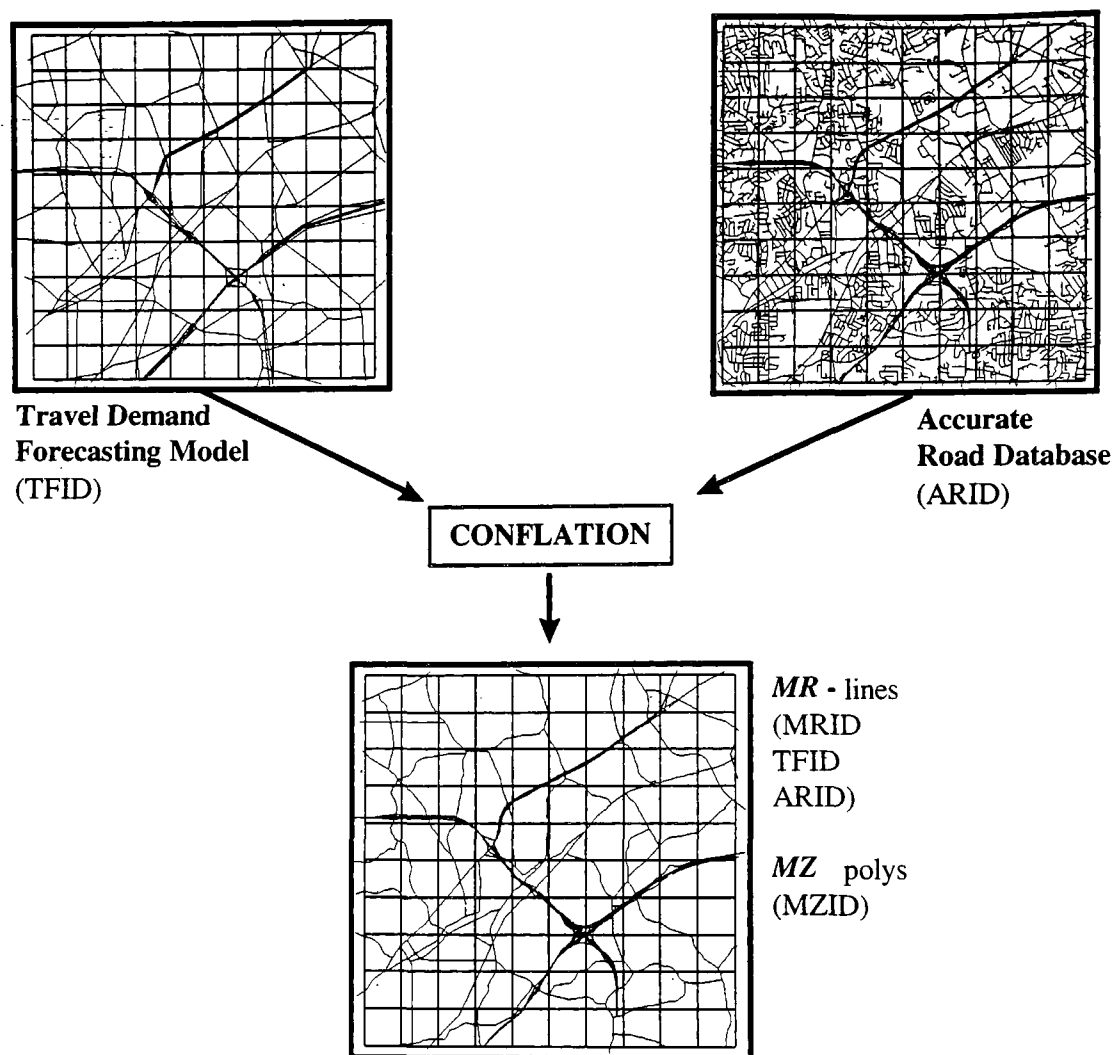
Figure 5.3- Running Exhaust Entity Creation

5.2.1. Engine Start Polygons

Engine start polygons (*SZ*) consist of spatially joined features from *taz, census, landuse,* and *zipcode.* The 1624 SZ polygons had a mean area of 72,746 square meters. Most of the polygons are identical to smallest input polygons, generally, the US Census blocks. Each SZ polygon maintains identifiers to the original databases in a one-to-one or many-to-one relationshir

5.2.2. Running Exhaust Lines and Polygons

Running exhaust lines and polygons (*MR* and *MZ*) consist of conflated travel demand model network segments and areas bounded by those segments. The segments represent roads that have prognostic estimates of travel behavior from the ARC. The polygons represent the roads that have road-specific estimates of travel activity. The road segments had a median length of 202 meters. The median of the 205 minor zones is 487,815 square meters.

5.3. Fleet Characteristics

The fleet characteristics were developed from the two files described in section 5.1.1. Five fleet distribution databases were created by the model from the two files; *es.tg* (zone-based engine start technology groups), *re.tg* (zone-based running exhaust technology groups), *scf.tg* (zone-based SCF technology groups), *rereg.tg* (regional running exhaust technology groups) and *mr.tg* (major road running exhaust technology groups). The files contain identifiers that connect records to spatial entities (*sz* or *mr*).

5.3.1. Model Year Distributions

The distributions of vehicle model years predicted by the model for the sample area are shown in Figure 5.4. The figure shows the entire sample area's mean distribution and two sample *sz* polygon distributions. The sample size for the zone with SZID of 2176 was 109 vehicles and SZID 209 was 126 (zonal frequencies varied from zero to several hundred). The entire area's mean frequency was 49. In Chapter 6, a comparison between observed and estimated model year distributions is presented.
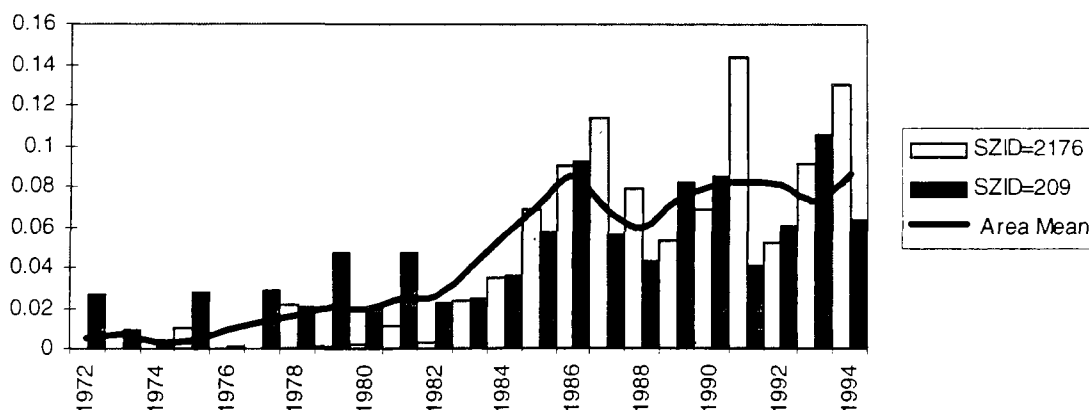
Figure 5.4 - Model Year Distribution

5.3.2. High Emitters

High emitting vehicles were defined in one of the preprocesses [see section 3.4.2.3]. In the process, vehicles are randomly selected from four groups of vehicles, each having different likelihoods of being high emitters. The 'flagged' vehicles are then characterized as high emitters. The resulting sample area high emitting vehicle distributions are shown in figure 5.5. The figure shows a dot density map of all engine starts, CO high emitters, HC high emitters, and NOx high emitters. The two circles identify locations with high numbers of engine starts, but low numbers of high emitter starts, suggesting that the likelihood that a vehicle is a high emitter, varies spatially. Although this can't be validated until other model components are validated, it does suggest a possible value of the model.

5.4. Vehicle Activity

The sample area vehicle activity was developed from the Atlanta Regional Commission's (ARC) Travel Demand Forecasting Model dataset. Supplemental information came from 11 speed and acceleration profiles and temporal factors from Parsons Brinkeroff Inc.

5.4.1. Engine Start Activity

Engine starts were developed from trip generation data at the ARC's traffic analysis zone level. Trips were disaggregated to SZ coverage polygons based on the ARC's 1995 land use data, 1990 US Census STF3A data housing unit densities, 1994 US Census TIGER data, and school and university landmarks developed by Georgia Tech. The AM peak hour engine starts spatial distributions is shown in figure 5.5.

The engine start temporal distributions were developed using half hourly distributions by trip type (home-based-work, home-based-shopping, home-based-other, home-based grade school, home-based-university, and non-home-based). Figures 5.6 through 5.10 show the hourly distributions of the trip origins, directly translated to engine starts. The non-home-based are not disaggregated by origin and destination because there is no information regarding the origins or destinations.

5.4.2. Running Exhaust Activity

Running exhaust activity was estimated for major roads and minor roads. Minor roads consisted of all roads not explicitly modeled in the ARC's travel demand forecasting model. Major road were explicitly modeled by ARC for daily activity using The Urban Analysis Group's TRANPLAN product. The network used in the TRANPLAN model was conflated to accurate roads.

106

Minor road running exhaust activity was predicted using the engine start activity estimates for each *SZ* zone, and the shortest network path from the centroid of that zone to the closest *MR* line. An average travel time for each path was determined using an average travel speed of 30 MPH. The aggregate travel time for activity in each *MZ* became the estimate of minor road vehicle activity.

Major road running exhaust activity was developed directly from the TRANPLAN output, and associated speed and acceleration profiles. Figure 5.11 represents the spatial distribution of peak hour volume density. Figure 5.12 shows the temporal distribution used to divide daily activity into hourly segments. Eleven speed and acceleration profiles were used to predict modal activity; five for each LOS on interstates, five for each LOS on interstate ramps, and one for all other roads. The data for one speed and acceleration profile used for the non-interstate, non-ramp, roads were collected on a major arterial between signalized intersections.

Road grade data were available for approximately 25% of the interstates running through the sample area. All roads without grade information were assumed to have a zero grade. Roads with grade information had the grade distribution segmented into five intervals.

5.5. Facility and Gridded Emissions

The emissions estimates for the sample area were developed for engine starts and running exhaust activity. Engine start emissions were developed for each *sz* polygon. Running exhaust emissions were developed for *mz* polygons and *mr* lines. All estimates are in grams. The gridded estimates were aggregated from the facility entities at 100, 250, 500, 1000 and 2000 meter grid cell sizes (the value refers to the length of one cell side, not the cell's area). The various sizes were developed to explore and demonstrate the impact of grid cell size on the emissions estimate. Figures 5.13 through 5.19 represent the 100 meter grid cell aggregation in a surface (with an overlay of 500 meter grid lines). Figures 5.20   5.22 show temporal distributions of emissions.

5.5.1. Engine Start Emissions

Engine start emissions are shown in figure 5.13. The figure shows a three-dimensional surface where *x* and *y* are geographic coordinates and *z* is adjusted value for engine start emission estimate. The value is adjusted to spatially identify relative emission estimates. The individual 'spikes' show estimated high concentrations of engine start emissions. Given the hour of the day (7-8 AM) the majority of the emissions from engine starts occur in residential areas. Thus, areas with large

populations (resulting in large numbers of engine starts) combined with higher emitting vehicles spatial distributions will have high emissions. In the figure, the highest spikes in the north east corner of the map are locations of dense multi-family development.

The total 24-hour engine start CO estimated for the area is 15,768,000 grams, the total estimated HC is 347,000 grams, and the total estimated NOx is 571,000 grams.

5.5.2. Minor Road Running Exhaust Emissions

Minor road running exhaust emissions are shown in figure 5.14. The same format as described in section 5.5.1 is used for the figure. As seen from this figure, spikes of intense emissions are not prevalent. This is due to the highly aggregate nature of estimating local road activity to large polygons. However, variability is evident among zones, indicating the impact of the road network configuration on travel time.

The total 24-hour minor road running exhaust CO estimated for the area is 1,195,000 grams , the total estimated HC is 50,000 grams, and the total estimated NOx is 60,000 grams.

5.5.3. Major Road Running Exhaust Emissions

Major road running exhaust emissions are shown in figure 5.15. As expected emission 'spikes' fall along the major road (shown as white lines). The highest spikes fall along the interstates. Arterials, especially the Peachtree Industrial Blvd Peachtree Road arterials, show significant emissions as well. The emissions estimates are not linear with volume, as appears, but are affected by the predicted modal behavior. When speed and acceleration profiles are used that have high variability (arterials, or low LOSs), there are higher emissions.

The 24-hour major road total running exhaust CO estimated for the area is 18,375,000 grams, the total estimated HC is 734,000 grams, and the total estimated NOx is 811,000 grams.

5.5.4. SCF Running Exhaust Emissions

Figure 5.16 represents speed correction factor (SCF) running exhaust emissions. Again, the emission spikes fall along the major roads. However, the emissions along arterials do not appear as significant as the previous figure. Mostly, this is due to the use of average speed as the measure of vehicle activity, excluding the effects of variable acceleration and deceleration. Thus, arterials and poor LOS road segments may have poorly represented emission estimates. While the level of spatial

108

aggregation used for reporting (4-5 km grid cells) may negate this impact, it is clear that facility-level and smaller grid cell aggregations of emission estimates will be affected.

The total 24-hour SCF running exhaust CO estimated for the area is 17,045,000 grams, the total estimated HC is 1,236,000 grams, and the total estimated NOx is 4,432,000 grams.

5.5.5. Total Emissions

The total emissions estimates were developed by adding the 100 meter aggregations of engine starts, minor road running exhaust, and major road running exhaust (aggregate modal). Total emissions for the study area are represented in figures 5.17 to 5.22. Figures 5.17 to 5.19 show emissions estimates for CO, HC, and NOx, between 7 and 8 AM. Figures 5.20 to 5.22 show the temporal variability found in the estimates occurring between 6 AM and 9 PM. The CO and HC estimates are characterized by the major road emissions and the spikes of engine start emissions. The NOx estimates are characterized by emissions on the major roads.

The figures showing temporal variability identify the impacts of the distributions seen in figures 5.6 through 5.12. It appears from the figures, that engine start emissions dominate the off-peak emissions, while running exhaust emissions dominate the peak hour emissions. This may be a function of the impact of congestion on the roads (higher variability in modal activity) during high traffic times. There also appears to be an unusual amount of engine start activity between 3 and 4 PM. Reviewing the temporal curves, all trip types except home-based-work are high between 3 and 4 PM.

5.6. Conclusion

Chapter 5 presented the results of model runs on a 100 sq. km area in northeast Atlanta, GA. The results of each module are presented along with detailed descriptions of the input data. Outputs are described in detailed as well.

High Emitter Engine Starts 7-8 AM
(1 dot = 1 start)

All Engine Starts, 7-8 AM
(1 dot = 20 starts)
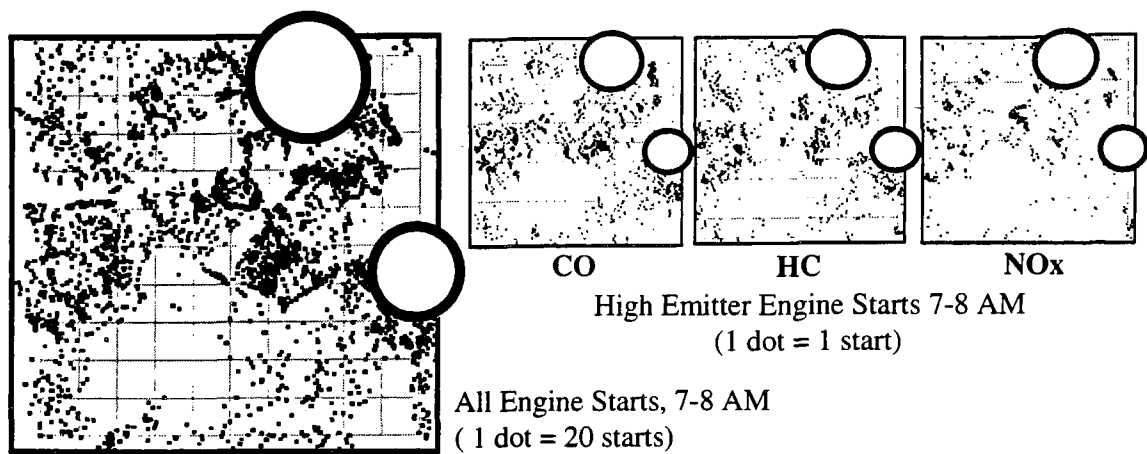
Figure 5.5 - High Emitter Engine Starts, 7-8 AM
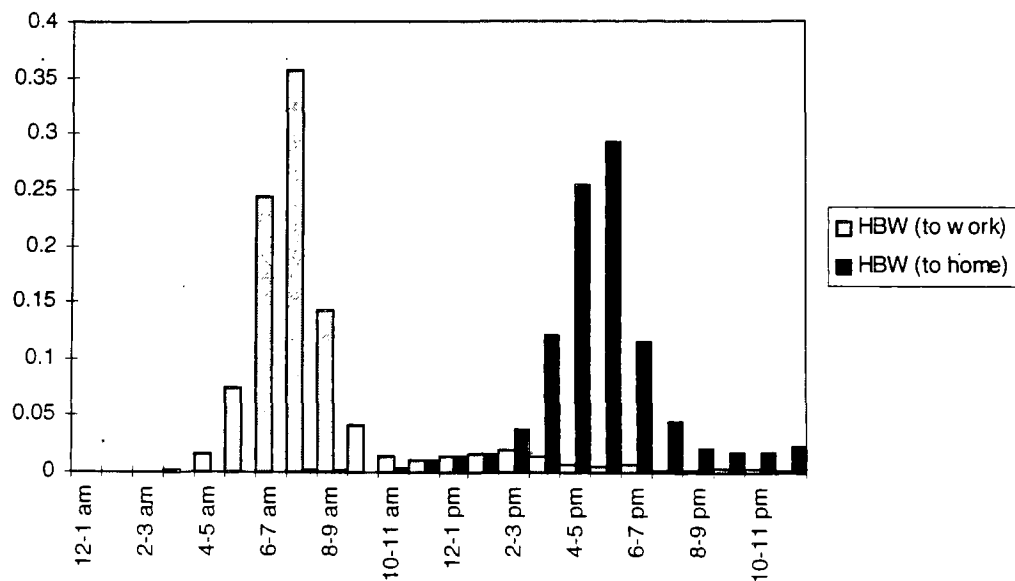
Figure 5.6 - Home-Based Work Trip Temporal Distribution
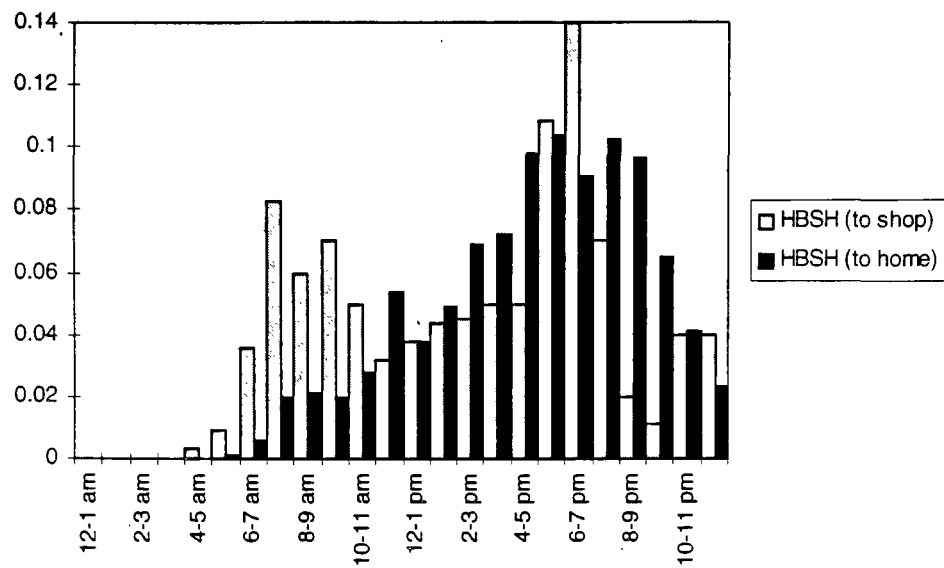
Figure 5.7- Home-Based Shopping Trip Temporal Distribution
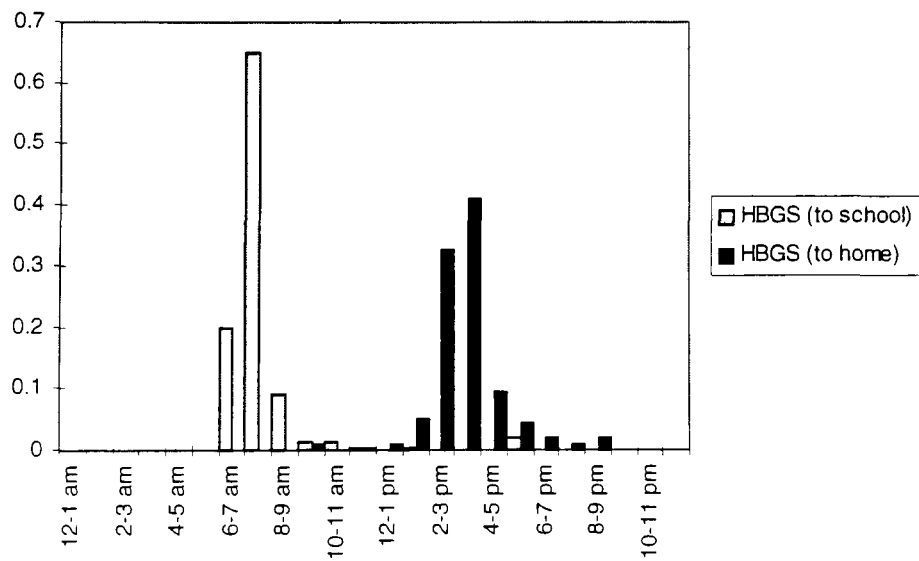
Figure 5.8 - Home-Based Grade School Trip Temporal Distribution
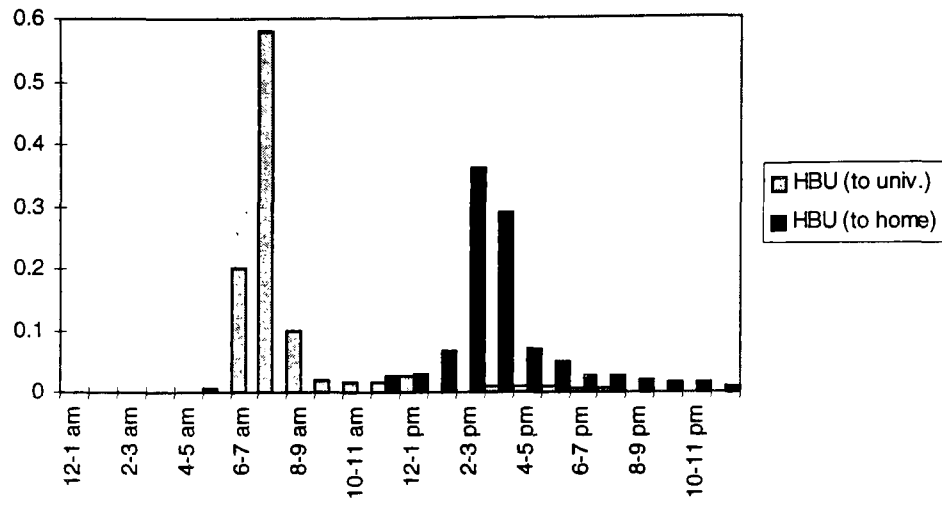
Figure 5.9 - Home-Based University Trip Temporal Distribution

NHB
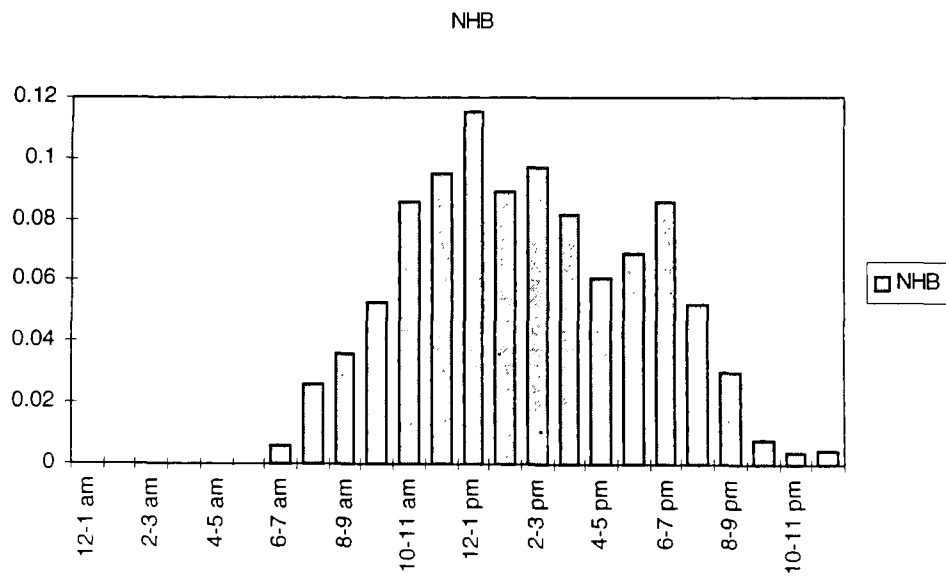
Figure 5.10 - Non-Home-Based Trip Temporal Distribution
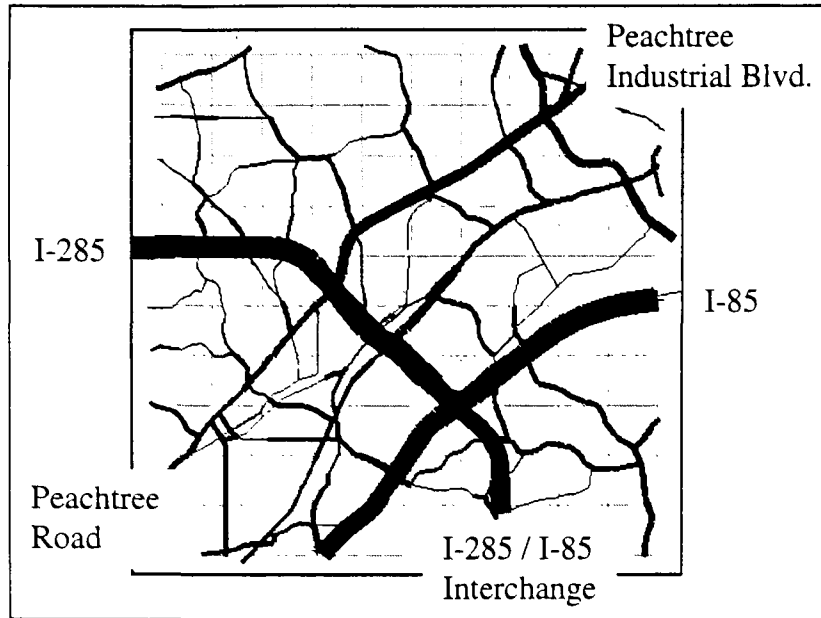
Figure 5.11 - Road Volume Density
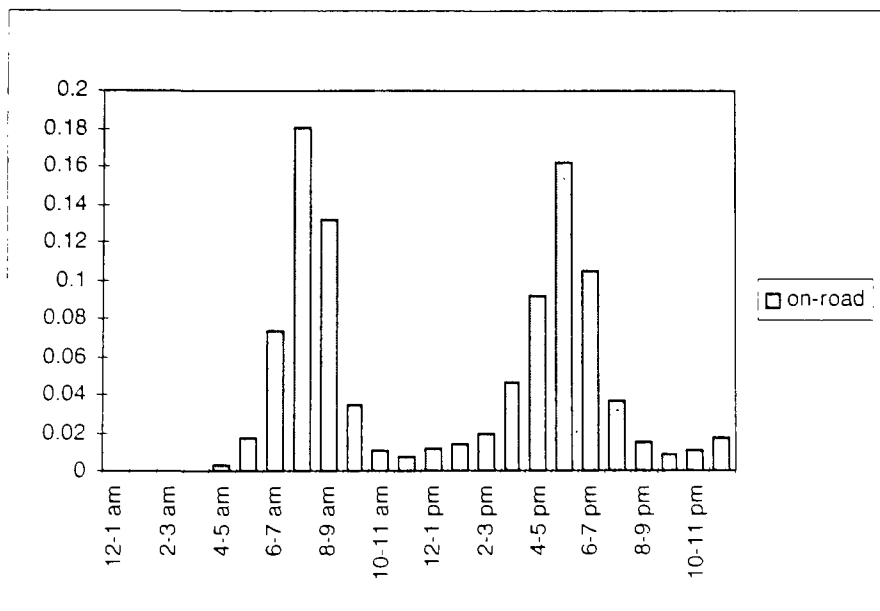
Figure 5.12 - On-Road Activity Temporal Distribution

Peachtree
Industrial Blvd.

I - 85

I - 285

I - 285 / I 85
Interchange

Peachtree St.

N

10 km x 10 km

Grid Cell = 0.5 km
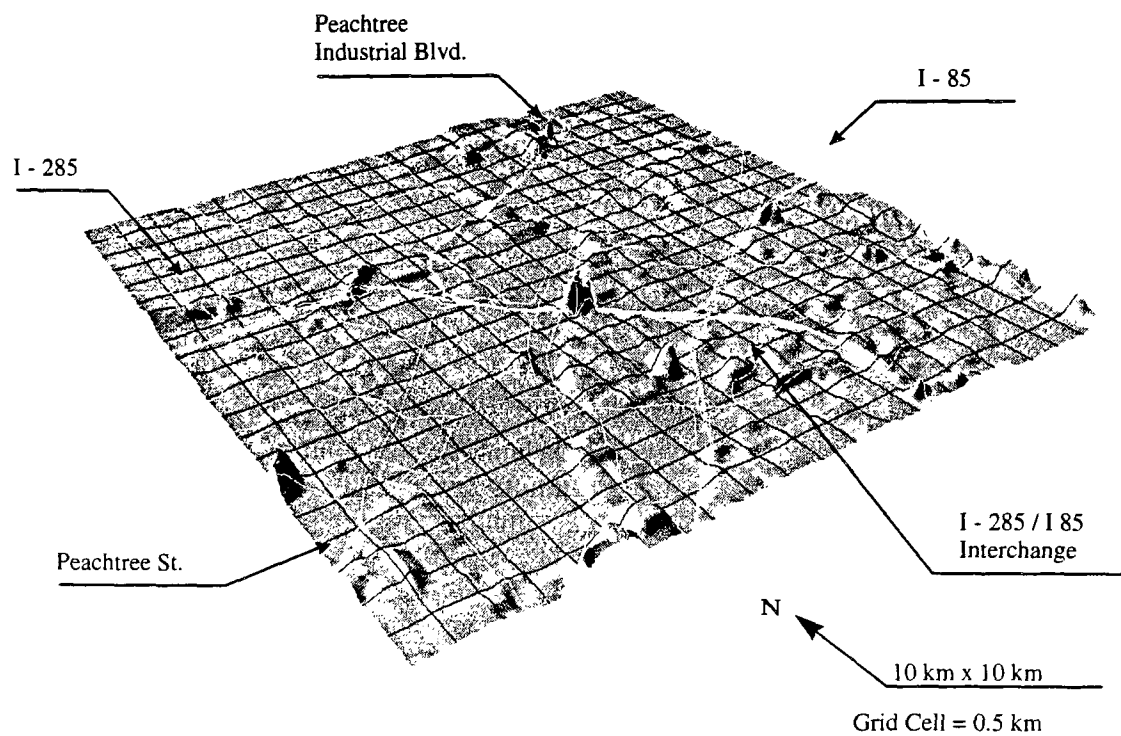
Figure 5.13 - Engine Start CO, 7-8 AM

Peachtree Industrial Blvd.

I - 85

I - 285

I - 285 / I 85 Interchange

Peachtree St.

N

10 km x 10 km
Grid Cell = 0.5 km
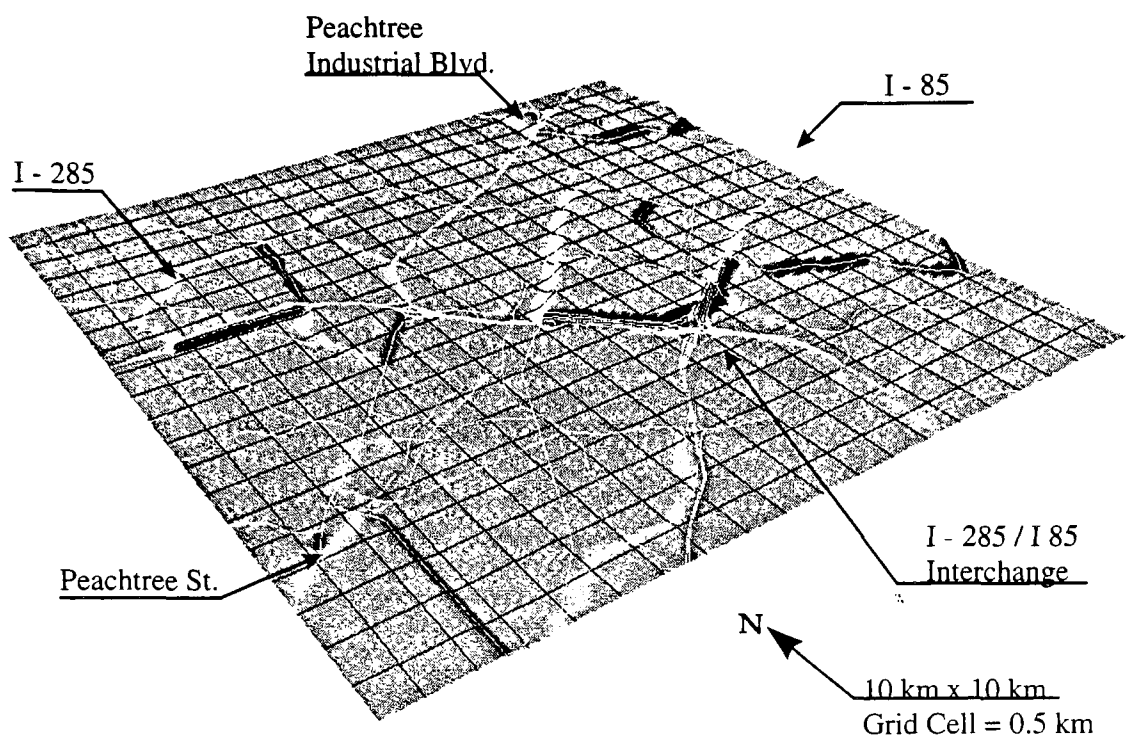
Figure 5.14 - Minor Road Running Exhaust CO, 7-8 AM
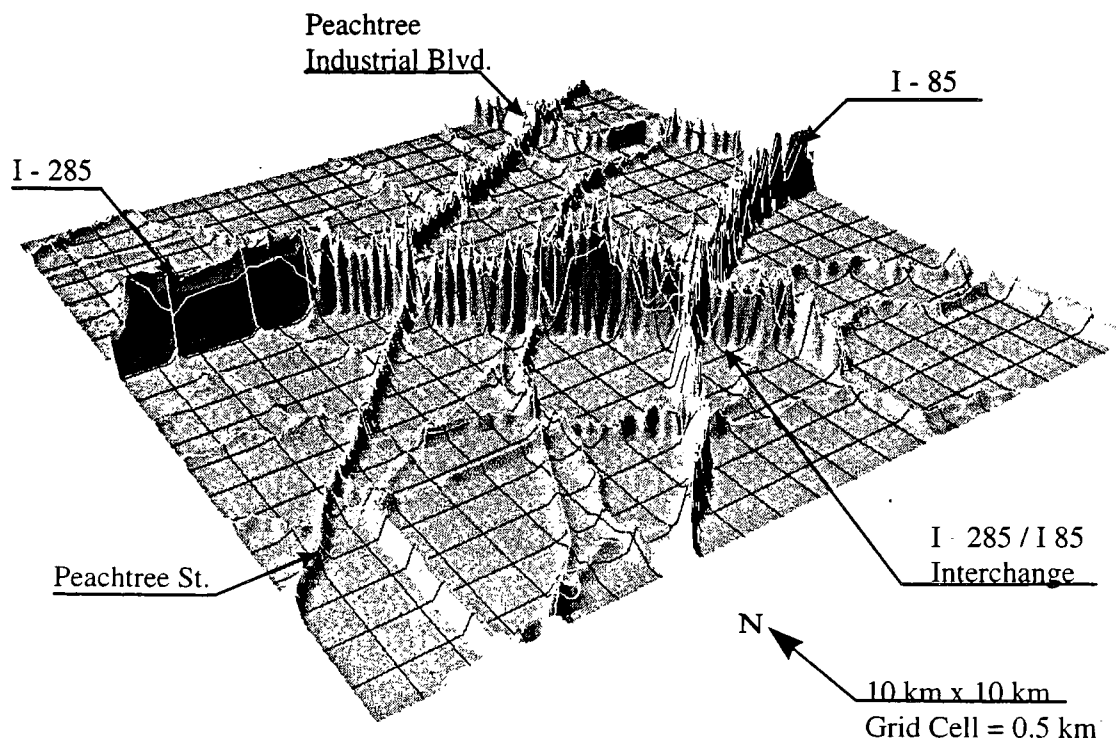
Figure 5.15 - Major Road Running Exhaust CO, 7-8 AM
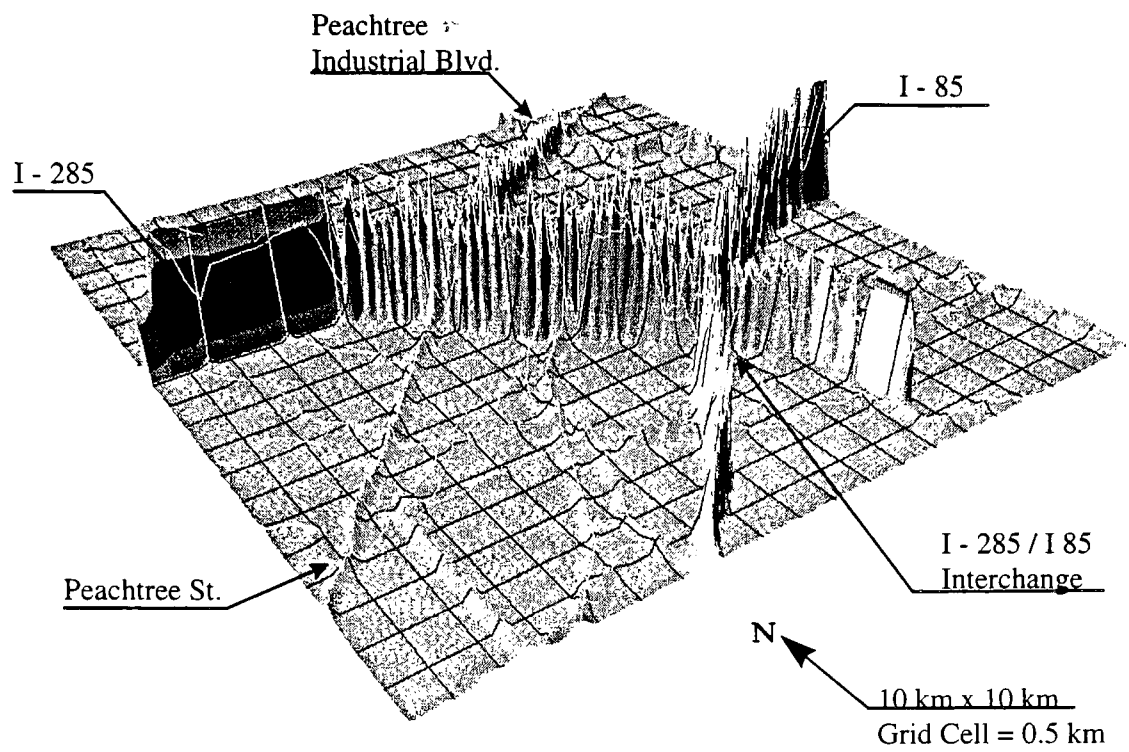
Figure 5.16 - SCF Running Exhaust CO, 7-8 AM

Figure 5.17 - Total CO, 7-8 AM

Peachtree
Industrial Blvd.

I - 85

I   285

I   285 / I 85
Interchange

Peachtree St.

N
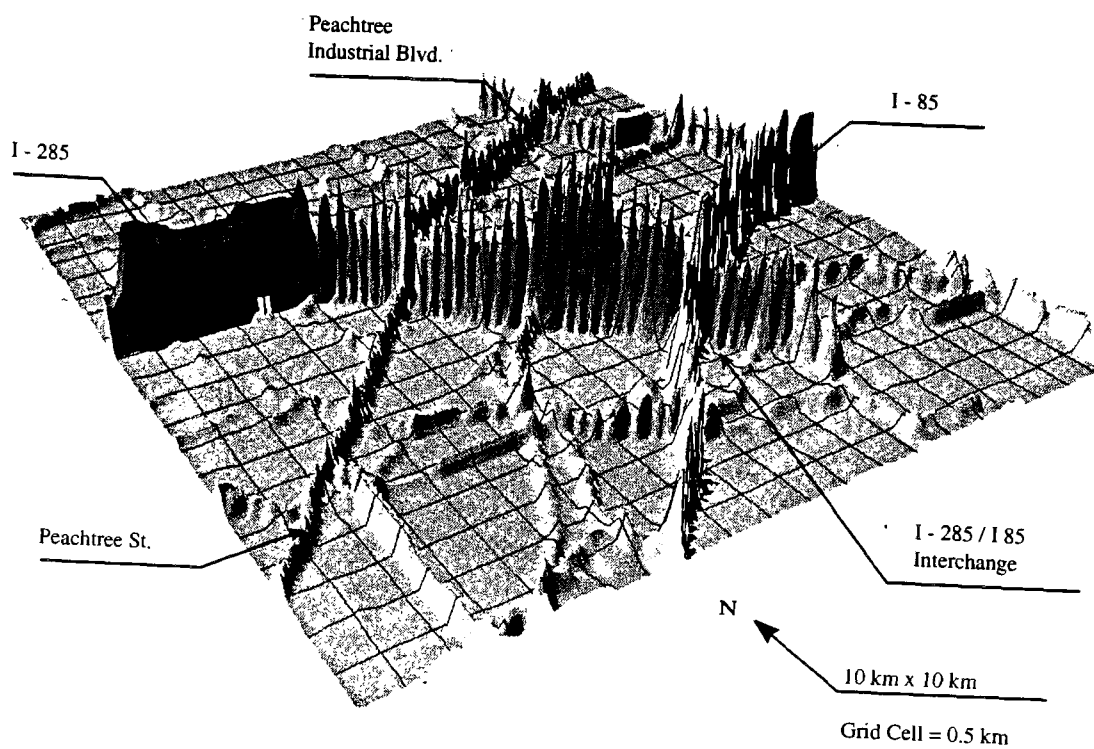
10 km x 10 km
Grid Cell = 0.5 km

Figure 5.18 - Total HC, 7-8 AM

119

Peachtree
Industrial Blvd.

I - 85

I - 285

I - 285 / I 85
Interchange

Peachtree St.

N
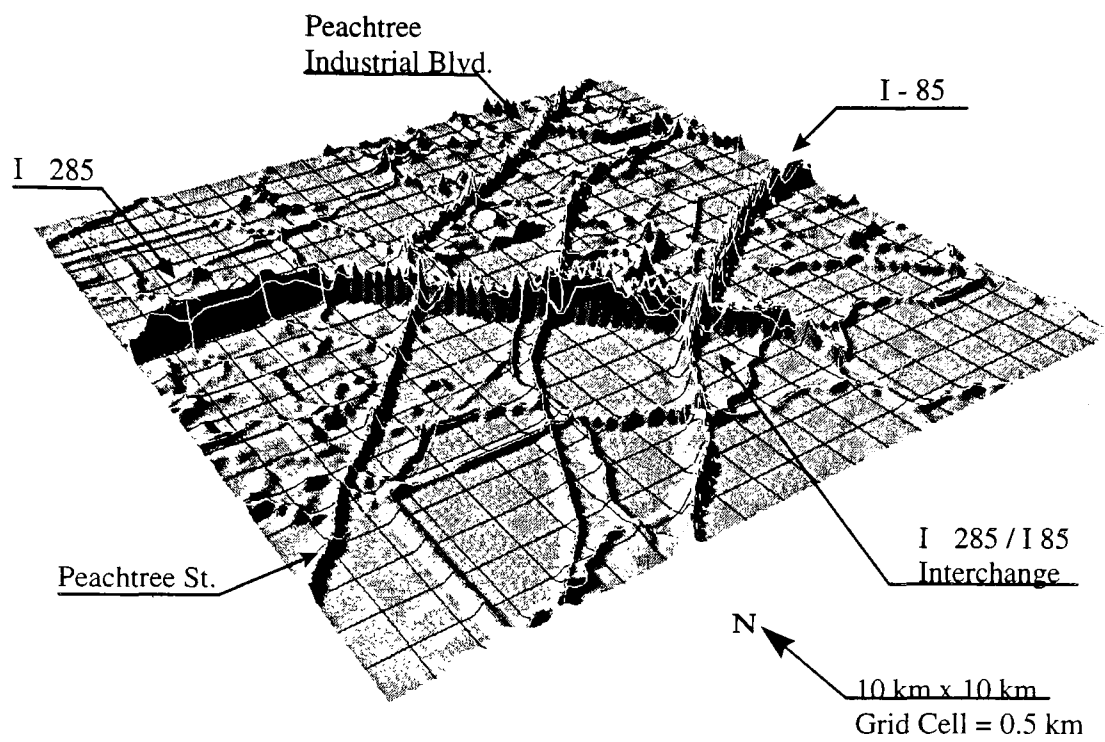
10 km x 10 km
Grid Cell = 0.5 km
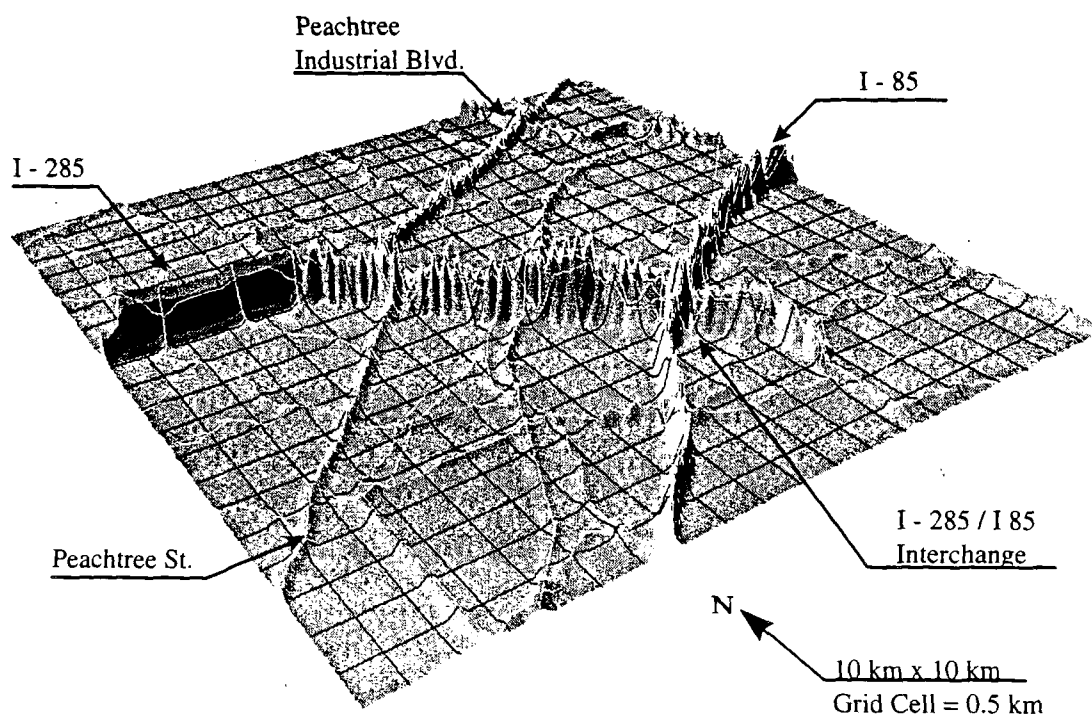
Figure 5.19 - Total NO$_x$, 7-8 AM

6-9 AM

10 AM - 1 PM

2 - 5 PM

6 - 9 PM
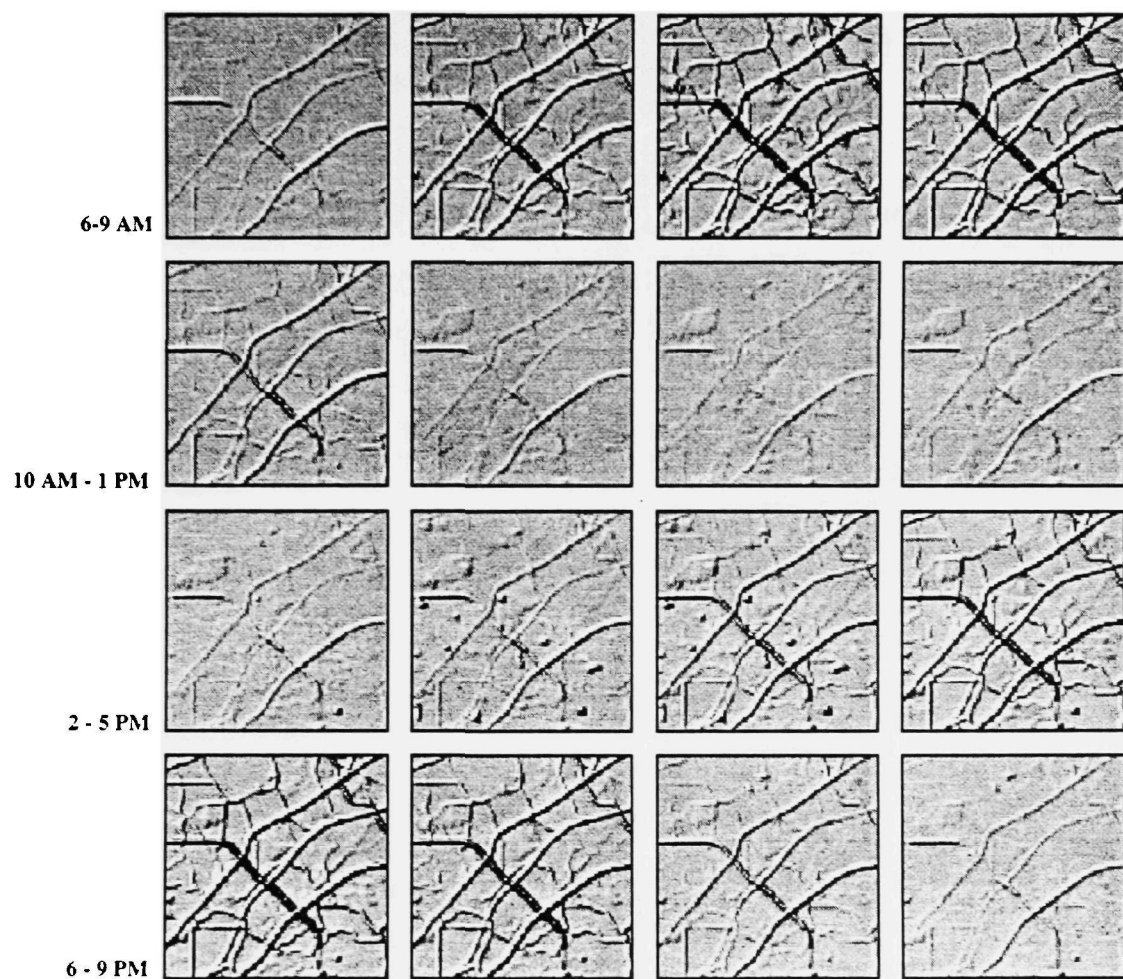
Figure 5.20 - Total CO, 6 AM - 9 PM

6-9 AM

10 AM - 1 PM

2 - 5 PM

6 - 9 PM
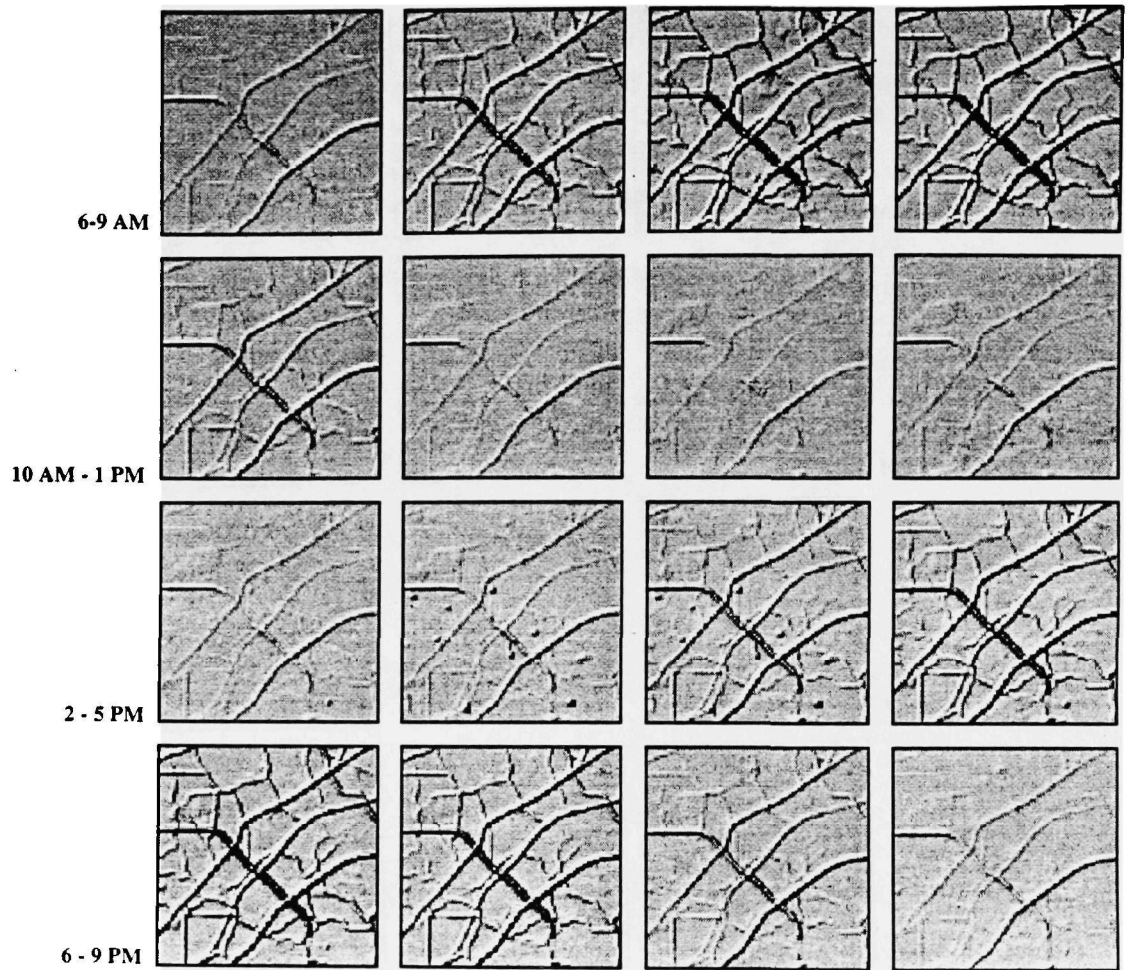
Figure 5.21 - Total HC, 6 AM to 9 PM

6-9 AM

10 AM - 1 PM

2 - 5 PM

6 - 9 PM
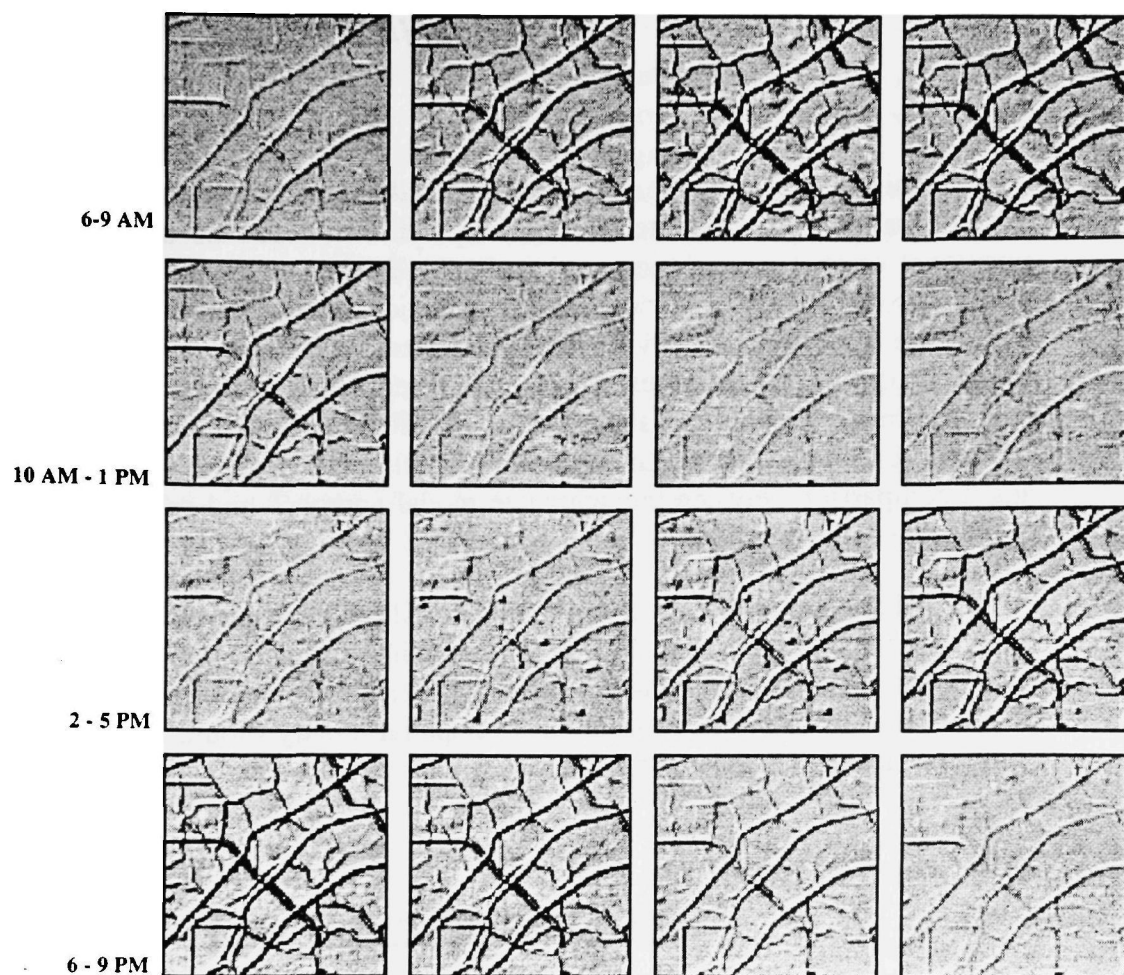
Figure 5.22 - Total NO$_x$, 6 AM to 9 PM

6.   MODEL EVALUATION

This chapter evaluates the model by analyzing and discussing potential sources of error, paying particular attention to the spatial data issues discussed in section 2.4.3; positional accuracy, resolution, and content. Studying the sources of model spatial error provides insight into developing validation studies and future research needs. One of the model objectives identified in chapter 3 was for the model to be statistically sound. Exploring the sources of error, and their propagation in the model, will also help determine appropriate strategies for developing confidence bounds around the spatially resolved estimates, an important model design feature. A sensitivity analysis and a comparison between the aggregate modal approach and speed-correction-factor approach are also provided.

A large amount of error in the model will be associated with the quality of the input data. While input data error is not explicitly discussed, it should be evident that any limitations associated with the input data impact the model results. It should also be noted that the input data's measures of spatial quality should focus on the relative positional accuracies among the datasets, not just the absolute accuracy.

6.1.   Spatial Environment

The spatial environment modules create the spatial entities sz, mr, and mz. Each entity was created by spatially manipulating input polygon and line data. During the spatial manipulation, potential positional errors arise that would impact the locational accuracy of the estimates. The following two sections describe the potential issues.

6.1.1. SZ

SZ was created using polygon-on-polygon overlay techniques on the four ARC/INFO coverages census, taz, ZIPcode and landuse (see figure 4.2). The technique merges two or more polygon networks into a single polygon network. The datasets share many common boundaries. However, the data were developed from different sources and resolutions resulting in different representations of common boundaries. In the sample area, the US Census blocks and the ARC's TAZs were generated from the original TIGER data, and, therefore, match very well. The ARC's

land use and ZIP code data were developed from different sources. The impact of this problem is that there is potential for misrepresentation of the *landuse / TAZ / census* combinations.

Figure 6.1 demonstrates the polygon-on-polygon overlay problem. The figure shows a portion of the sample area's polygon structure. The left side shows *census* polygons in black lines and *landuse* polygons in gray lines. The right side shows *census* and *landuse* polygons in gray, and the resulting *SZ* polygons in black. Point A shows a shared boundary between *landuse* and *census* that is represented differently. Point B shows polygon boundaries that may, or may not, represent the same features; it is too difficult to tell conclusively.

The polygon overlay process includes a 'fuzzy tolerance' that allows the user to define a threshold that is allowable for matching edges. In the study model example, a tolerance of 30 meters was allowed. The resulting polygon structure (*SZ*) is shown on the right side of the figure. Point C shows the same area as point A, but the potential 'sliver' polygon was removed because the lines were within the tolerance level. Point D shows borders that may have represented the same feature, but as the distance between the two edges deviated more than 30 meters, they were represented as separate entities.

Census and *Landuse*   *Census, Landuse,* and *SZ*
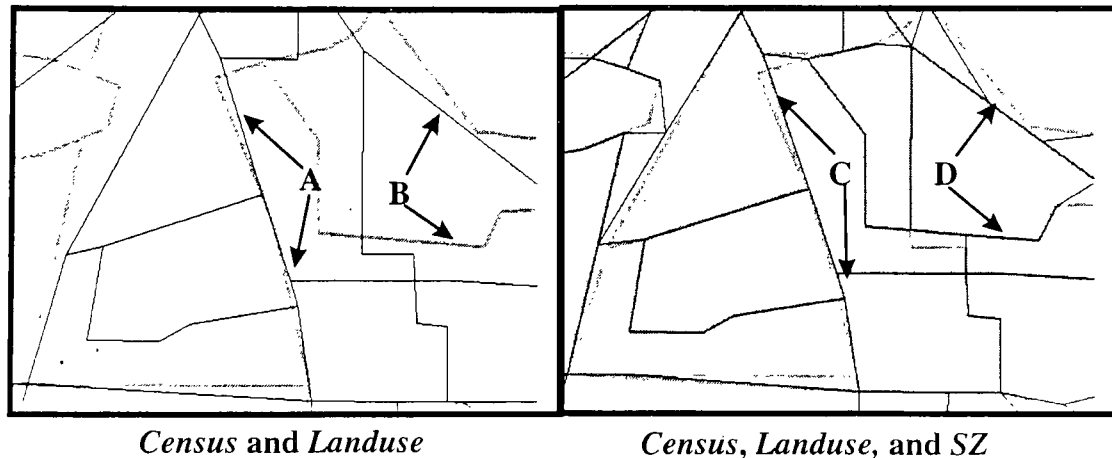
Figure 6.1 - Polygon Overlay Errors

As a result of the polygon overlay process, there are potential errors that exist in the spatial representation of the joined polygons. There are three potential impacts on the resulting data; data are lost, data are spatially misrepresented, and data combinations are incorrect. The biggest risk to emission output quality is the loss of data. By adjusting the boundaries of the spatial entities, any polygons less than 30 meters across would be removed. In the sample study area used in Chapter 5, 5 of 925 US Census blocks were lost during this process. Most of the blocks were road medians between divided highways and held no bearing on the emission estimates.

However, one was a Census block that contained two households. In this case, two households have little impact, but other datasets may have more, suggesting that the modelsuser would have to select a smaller tolerance level. The trade off to lower tolerance levels is increased spatial misrepresentation.

Spatial misrepresentation and incorrect data combinations are difficult to identify because the true values and positions must be known. Prior to model operation, detailed quality control steps in data development will prevent further error propagation. The model assumes that errors in input data exist, and therefore, a 'fuzzy tolerance' is used that can be adjusted to minimize data loss and maximize accurate representation.

6.1.2. MR and MZ

MR is created by selecting roads that are modeled in the travel demand forecasting model. Spatial errors associated with *MR* occur in the preprocessing steps, not in the formal model. The process of conflation, described in section 5.1.2, involves a great deal of user input, adding an aspect of human error.

The biggest concerns resulting from conflation errors are missing roads and miscoded roads. Matching some travel model links with the actual roads they represent can be difficult because the travel demand model network consists of abstract representations of roads. This is compounded by a lack of agreement or reporting about road classification in multiple datasets. Further, commercially available, accurate road datasets (similar to NAVTECH used in the study area) use significantly more detail than the travel models. One travel model link usually represents many road segments. During the conflation process, it is easy to miss one of the small segments, resulting in a 'gap' in the new network.

MZ polygons are created by defining the *MR* lines as polygon boundaries. There are no problems with the spatial accuracy of the polygons, other than those mentioned for the *MR* road network. However, there is one issue that is worth mentioning. The polygons are supposed to represent aggregations of local roads. The process of creating the *MZ* polygons does not actually consider the locations of these roads, but are defined as any polygon bounded by major roads (or the outside boundary). Therefore, medians from divided highways become defined as *MZ* polygons. While this poses no impact on the emission estimates, it can impact the amount of time required for model operation.

6.2. Vehicle Characteristics

The spatial errors associated with vehicle characteristics are significant and worth detailed study. There are several broad assumptions made in developing the spatially-resolved fleet distribution estimates. First, it is assumed that a vehicle's registered address is its 'home' location. This has not been proven or studied in previous research. Second, it is assumed that all the registered vehicles have the same probability of being operated at any given time. This not the case, but little evidence exists that justifies adjusting the fleet distribution to more accurately characterize operating vehicles. Third, it is assumed that any road segment's operating fleet distribution is composed of two groups of vehicles, a 'local' fleet and a 'regional' fleet. While some evidence exists to back up the idea [Tomeh, 1996], many questions remain about the specific definitions of the two groups of vehicles.

The potential negative impact of the above assumptions is reduced by predicting aggregate distributions rather than individual vehicles. The actual number of vehicles predicted at each entity, or whether the right vehicle is predicted at each entity, is less important than the predicted distribution. The only vehicle information used by the model is the fraction of each technology group at the zonal or road segment level, not the frequency. The model is more concerned with accurately characterizing the fleet, than accurately identifying the fleet.

Measures of the model's ability to predict the fleet distribution must come from future validation efforts. However, data do exist that indicate some biases found in the decoding process. Figures 6.2 and 6.3 show the degradation of the quality of the fleet estimate as vehicle characteristics are determined in the model. The basis for comparison (model year) comes from the raw vehicle registration dataset. As vehicle identification numbers (VINs) are decoded, model year information is predicted. Comparing the predicted and original model year distributions for the various datasets shows where bias has been introduced into the system.

Figure 6.2 shows the drop in the frequency of each model year for three steps in the decoding process; the VIN decoder, the removal of non-autos, and the assignment of vehicle weight. The VIN decoder operation results in a 7.7 % loss of data because the VIN couldn't be decoded. Most of those vehicles are older than 1980, when VINs were not standardized among manufacturers. Two odd 'humps' occur (frequency is overpredicted) in model years 1973 and 1978. After removing non-autos, only the 1973 hump remains. Further study revealed that the VIN decoder software was incorrectly assigning pre-1972 BMWs and Volvos as 1973 vehicles.

Adding the test weight to the vehicles (removing those without matches) resulted in a substantial data loss (39%). It also appears that the data loss is biased by model year, with 1988 vehicles underrepresented and 1995 and newer vehicle

127

unrepresented. Figure 6.3 shows the resulting distributions. The final distribution shows the impact of the data loss on the fleet distribution. . Pre-1972, post-1994, and 1988 vehicles are under-represented. Mid-1980s and early-1990s vehicles are over-represented.

6.2.1. Zonal Fleet

There are two concerns regarding the spatial allocation of the vehicles to zones; incorrect assignment, and non-residential trip distributions. Vehicles can be incorrectly assigned to zones because of address-matching problems. Non-residential trips use the fleet distribution of the current zone, disregarding the fact that the actual fleet distribution consists of vehicles originating from other locations.

The address-matching process can result in a small percentage of vehicles that were incorrectly assigned to zones because of errors in the address, an issue that has been well-documented in the literature. This problem is minimized in the model by having stringent matching guidelines; the vehicles' ZIP codes must match the candidate addresses' ZIP codes, the road types must match perfectly, and there can only be one error in spelling or incorrect address prefix (north, east, etc.). Further, the road dataset being used for address-matching could be missing new subdivisions or developments. If the registration dataset is newer than the last road dataset update, there could be vehicles that fail to match. However, the 'failed' vehicles are not discarded, but assigned a location based on their ZIP code.

The zonal fleet is developed from two sets of files, an address-matched file, and a ZIP code file (address match failures). To bring these two groups of vehicles together, the relationship between the SZ polygons and ZIPcode polygons must be identified. Each SZ is apportioned part of the ZIP code vehicles based on a comparison of the areas of the two polygons. A zone could have 78 vehicle address-matched within its boundaries, and an additional 10.3 vehicles assigned to it from the ZIP code. Since the concern of the model is the distribution, there are no problems that arise from non-integer frequencies.

Zones that do not have any address-matched vehicles are assigned the fleet distribution of the ZIP code. While some problems remain, zones that have new subdivisions will be assigned a fleet distribution that partially represents the vehicles registered at that location.

The issue regarding the fleet distribution of non-residential trips is not handled well ir ·· model. Since vehicles are not tracked during estimates of activity, there is no me. .nism for tying the origin fleet distribution to the destination. Unless the destination lies in a zone or ZIP code with a fleet distribution that is similar that of the origins, an incorrect distribution will be assigned. Given the dynamics of land use development, there is strong indication that strong bias will exist. For example, the

128

fleet distribution of vehicles leaving a commercial land use zone is assigned the fleet profile of registered vehicles in that zone and ZIP code, not the trip origin zones of the operating vehicles.
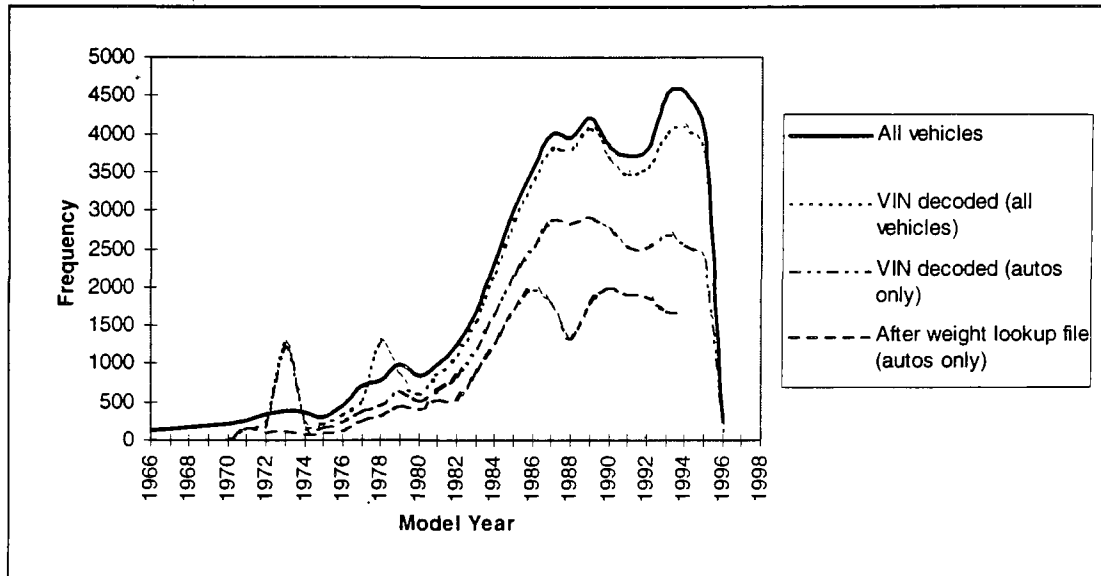
Figure 6.2 - Model Year Frequencies
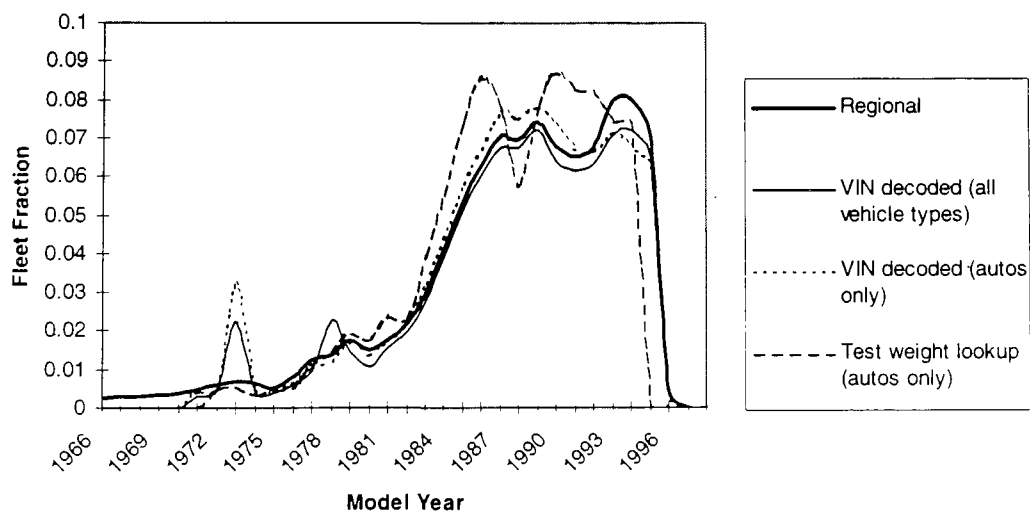
Figure 6.3 - Model Year Fraction

6.2.2. On-road Fleet

Problems associated with the on-road fleet distribution estimation stem from the unvalidated assumption that the on-road fleet can be summarized by combining a local fleet (defined as all vehicles within 3 km) and a regional fleet (all vehicles in the region). The instability of this approach is demonstrated by analyzing two sets of observed on-road vehicle datasets. Figure 6.4 shows registered vehicle locations for vehicles that passed through data collection sites in the study area. These data were collected and provided by the School of Earth and Atmospheric Sciences at Georgia Tech. License tags were captured on passing vehicles and matched to a registered vehicle's VIN and address. At site A, 674 vehicles passed through the data collection site. The figure indicates that spatial variation exists in the estimated origins of the observed vehicles. However, the size and shape of the spatial variability are unclear. Similarly, site B, with 13,481 vehicles, indicates spatial variability. The model currently uses a 3 km radius to define a 'local' fleet (10% of the observed vehicles fell in that range). It may be more accurate to select an alternative geometric (wedge, oval, network distance, etc.) search pattern involving road types, time of day, and network structure.

Site A: Arterial, 641 Vehicles   Site B : Ramp, 13481 Vehicles
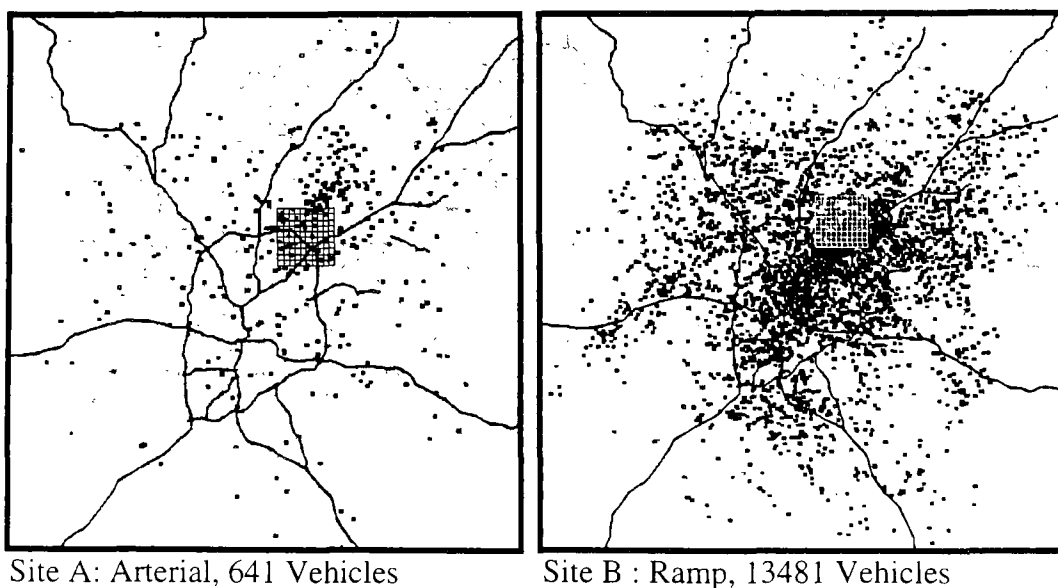
Figure 6.4 - Observed On-Road Vehicle Origins

6.3. Vehicle Activity

Spatial errors associated with the estimates of vehicle activity can be tied to previously mentioned problems with the spatial environment and the travel demand

model limitations. The model shares the problems associated with the use of the travel demand forecasting models in predicting emission-specific vehicle activity; inaccurate speeds, no feedback into the distribution phase, etc. (see section 2.3.1). These known travel demand model result limitations will not be discussed. However, trip disaggregation, the use of regional temporal distributions, and speed and acceleration matrices create some errors in vehicle activity estimates that are worth mentioning.

The disaggregation of trips by purpose to different landuse makes the broad assumption that the landuse data are discrete. All home-based trips are assigned origin engine starts to residential area. If homes can be found in other land uses, their engine start activity estimates will not be assigned to the correct location. The landuse data must be discrete to prevent this from occurring.

The use of *regional* temporal factors to distribute *zonal* and *road segment* activity results in errors. A series of spatial queries using 1990 Census data in Atlanta indicated that the fraction of people traveling to work between 6:30 and 7:00 AM was approximately 9% for people living within 2 kilometers of the central business district (CBD), and about 15% for people living between 8 and 10 kilometers from the CBD. By using regional temporal factors in the model, all zones and road segments are assigned the average, not allowing spatial variability. Thus, the peak hour 30 miles from the CBD will be the same as the peak hour 5 miles from the CBD.

The speed and acceleration profiles were used as a post-processor to the travel demand model's output. They are used to predict the modal distributions of the vehicles operating on the road. The model only includes matrices for interstates and ramps, forcing all lower classifications to rely on a single profile of mid-block estimates. As soon as new data are collected, validated, and available, the model structure can incorporate the new findings. The impacts of modal activity around signalized intersections could have a tremendous impact on the spatial variability of the estimates. As is, running exhaust emission estimates are highly correlated to volume. The potential variability found in future matrices could show that the highest emissions occur around major intersections, not high volume, low modal variability interstates. Further, the characterizing of dynamic modal activity into discrete bins and levels of speeds and accelerations could result in a certain level of error. Current research efforts are attempting to validate the approach.

6.4. Facility and Gridded Emissions

The spatial errors associated with the emission estimates come from aggregation. No spatial manipulation procedures are used to generate the facility-level emissions estimates. The facility-level estimates are, however, impacted by non-

131

spatial errors. The gridded emission estimates are generated by aggregating facility-estimates to vector grid cells of a user-defined size. During this process, spatial errors are incurred.

6.4.1. Facility Emission Estimates

New errors introduced to the facility-level emission estimates are generated by the process for determining emission rates. The emission modes (engine start, running exhaust) have gram per start or gram per second rates predicted by the hierarchical tree-based regression. The resulting emission rate values are discrete, with known confidence bounds. The accuracy of the emission rates is affected by the size and representativeness of the emission test dataset. The emission rates used in the model were developed from a dataset of approximately 3000 vehicle tests using about 700 individual vehicles. Currently, improvements and additions are expanding the dataset to over 10,000 tests.

6.4.2. Gridded Emissions

The errors associated with the aggregation of facility emissions to user-defined grid cells are spatial in nature. Vector grid cell polygons are overlaid with the *SZ*, *MR*, and *MZ* entities using the techniques described in sections 6.1.1 and 6.1.2. The 'fuzzy tolerance' used in this process is set as low as processing time will allow. There won't be any shared boundaries in this overlay technique and high 'fuzzy tolerance' will only degrade the spatial quality. Once the polygons and lines are split by the grid cells, the emission estimates are weighted (the proportion of the new entities' area or length and the original spatial entities' area or length) and summed by grid cell.

The size of grid cell selected by the user impacts the cell's accuracy. Larger cells will have more accurate estimates because errors (unbiased error) at the facility levels can be offset by aggregation. Small grid cells will have fewer entities falling within its borders, reducing the number of values to draw from. Further, grid cell sizes falling below the spatial accuracy of the origin datasets could spatially misrepresent the locations of emission estimates. Larger cells have the advantage of absorbing errors related to absolute position. Figure 6.5 shows grid cell aggregations from the sample study area described in chapter 5. Four levels are shown: 100 meter, 250 meter, 500 meter, and 1000 meter. The figure is useful in looking at the total emissions from different levels of aggregation. While the 1000 meter (1 km) grid cell is expected to be used for future photochemical models, additional information for research can be gleaned from smaller cell sizes.

132

100 Meter   250 Meter

500 Meter   1000 Meter
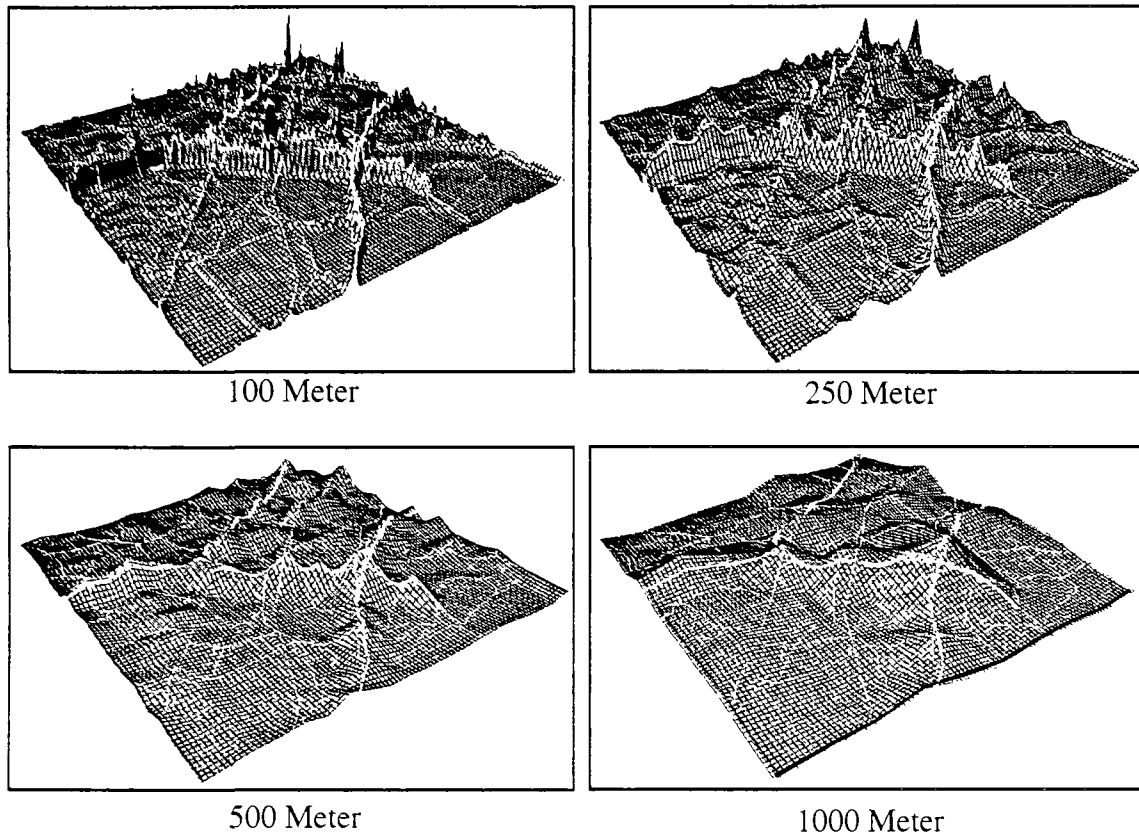
Figure 6.5 - Sample Grid Cell Aggregations

6.4.3. Sensitivity of Model

The model sensitivity can be measured in two ways: estimate accuracy and locational accuracy. The sensitivity of the estimate accuracy can be shown by running the model with a full range of input variables. The sensitivity of the locational accuracy depends on the spatial allocation of the estimate, given a full range of influential factors.

Figures 6.6 through 6.11 show how the emission rate varies for each technology group, level of service (LOS), and road grade. The graphs are for interstate activity only because speed and acceleration data for lower classifications do not yet exist. The percentage of the sample area's regional fleet in each technology group is provided in the graph as well. The very low percentage of some technology groups is the result of the problems mentioned earlier (section 6.2) regarding the determination of vehicle technologies.

All the technology groups have substantial estimated increases in emission rates for LOS F. The speed and acceleration profiles for interstate LOS F show substantially more variability in speeds and accelerations. Other LOS impacts are

133

fairly static, slightly increasing as traffic flow degrades from LOS A to LOS D. However, during LOS F activity (volume to capacities greater than 1.0 and average speeds less than 30 MPH), the model is predicting that emissions rates increase substantially.

The impact of road grade is seen for CO normal emitters, HC high emitters, and NOx high emitters. The graphs indicate that these technology groups have higher emission rates for steeper grades. As mentioned previously, the impacts of grade may be substantially under-predicted. Currently, the model adjusts the acceleration rates of vehicles based on the road grade. There are no mechanisms in the model that adjust emission rates based on engine load. It is expected that emission rates will vary significantly once these impacts are considered.

Unlike the emission estimates, the locational sensitivity of the model is not the result of a series of calculations. Estimates are allocated to zones or lines based on input data conditions. For example, the return trip of a home-based-shopping (HBSH) trip begins in a shopping area (commercial land use) and ends at home (residential land use). If the TAZ with a HBSH attraction has commercial land use within its boundaries, the emissions from the engine start are allocated evenly to all commercial areas. If no commercial land use is indicated by the data, the engine start emissions are all allocated evenly to the entire TAZ. All the sample area TAZs had residential and non-residential land uses, and all but two had commercial land uses.
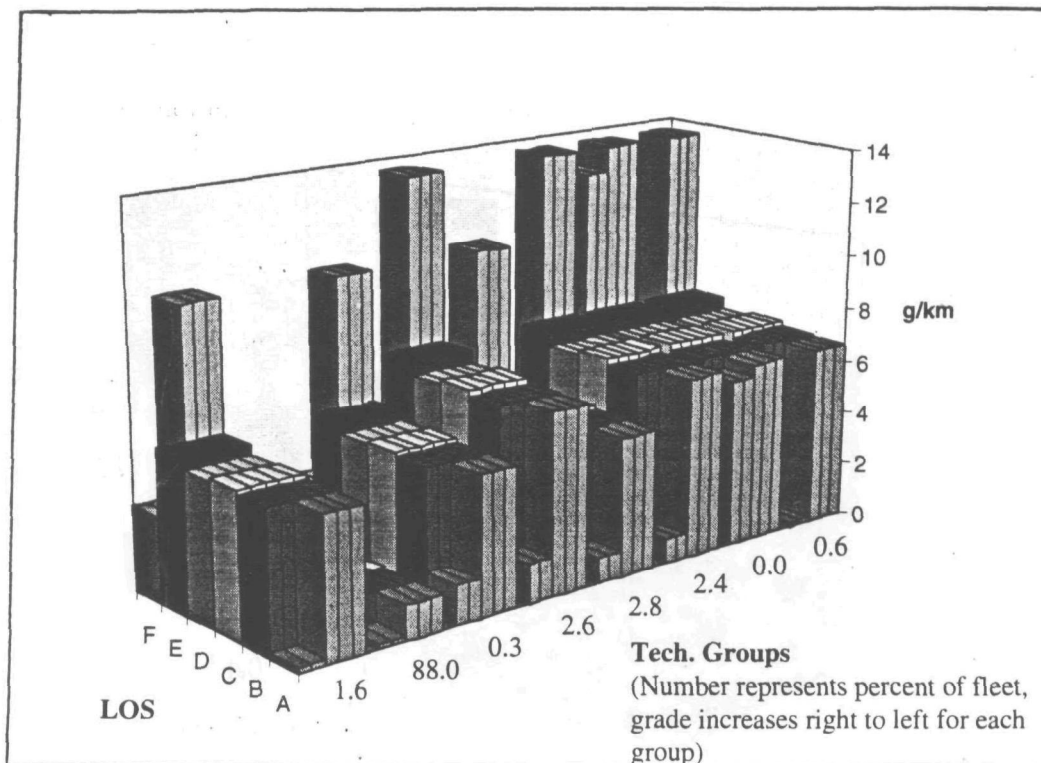
Figure 6.6 - CO normal emitter technology group emission rates by LOS and grade
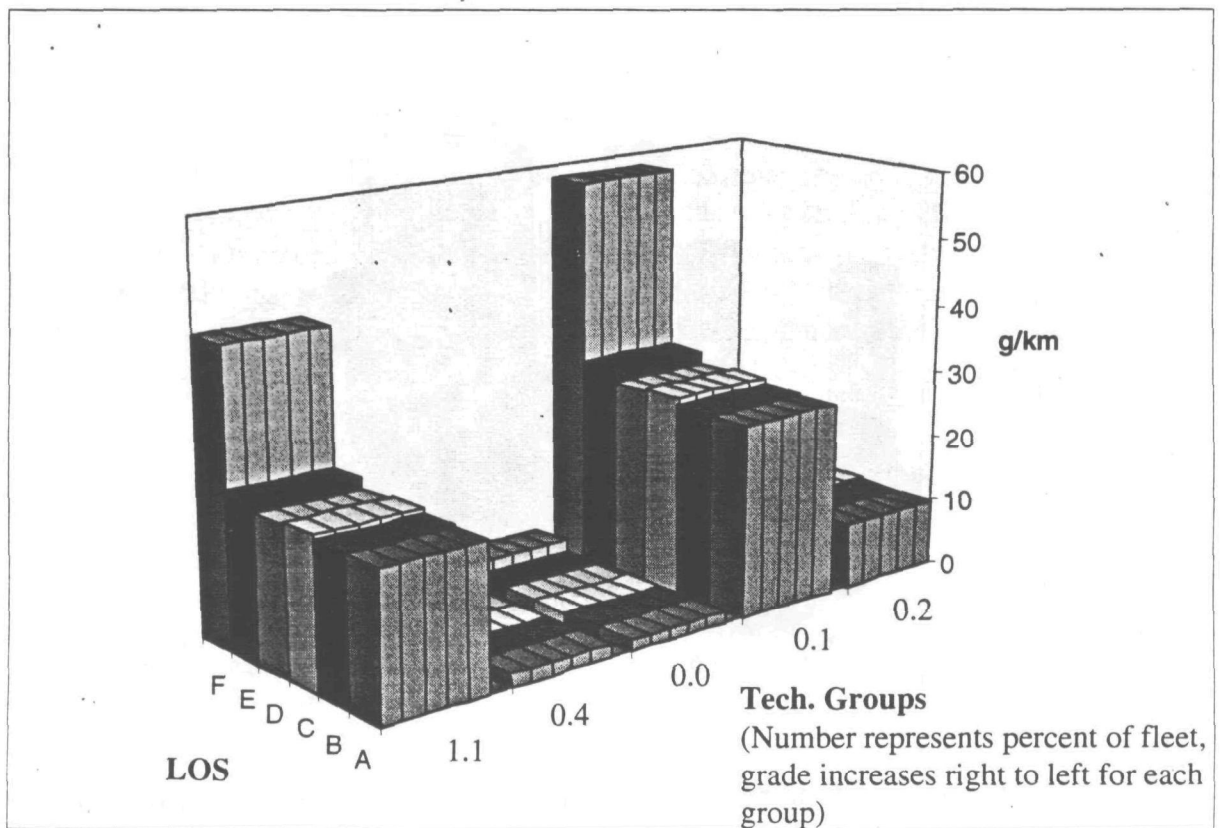
Figure 6.7 - CO high emitter technology group emission rates by LOS and grade
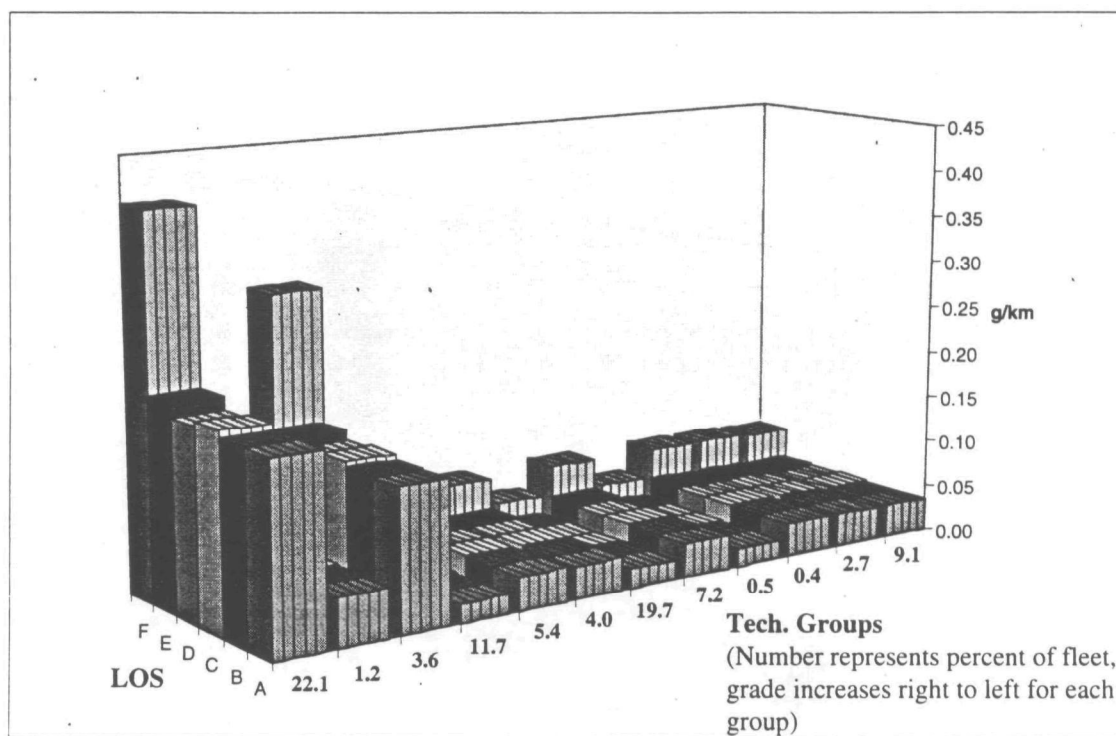
Figure 6.8 - HC normal emitter technology group emission rates by LOS and grade
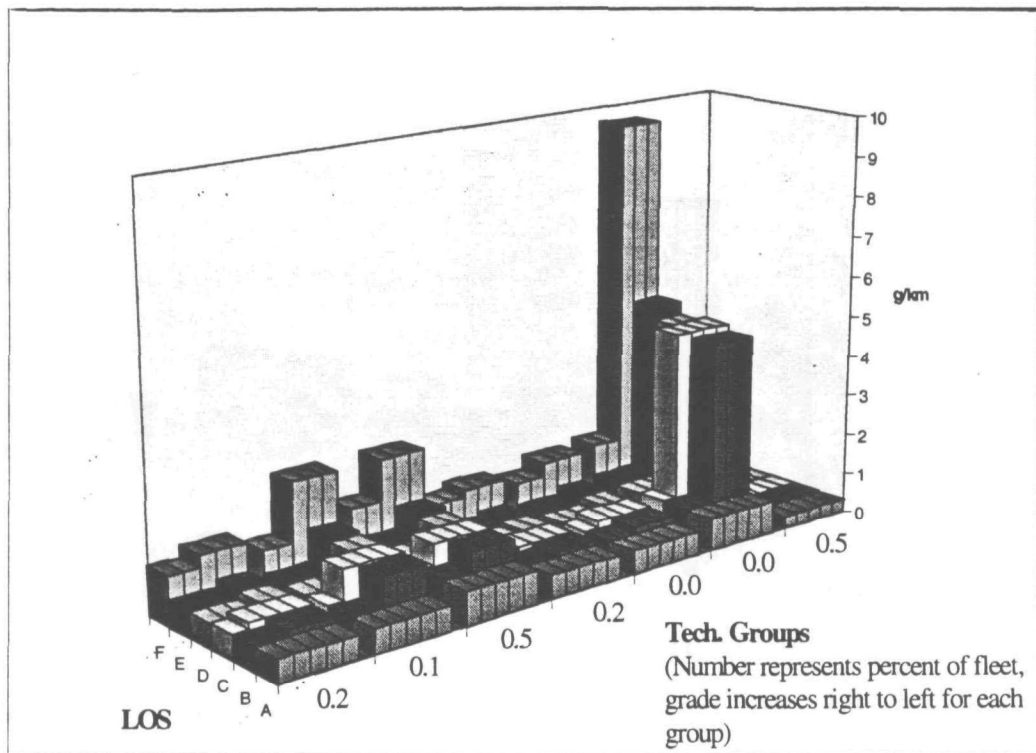
Figure 6.9 - HC high emitter technology group emission rates by LOS and grade
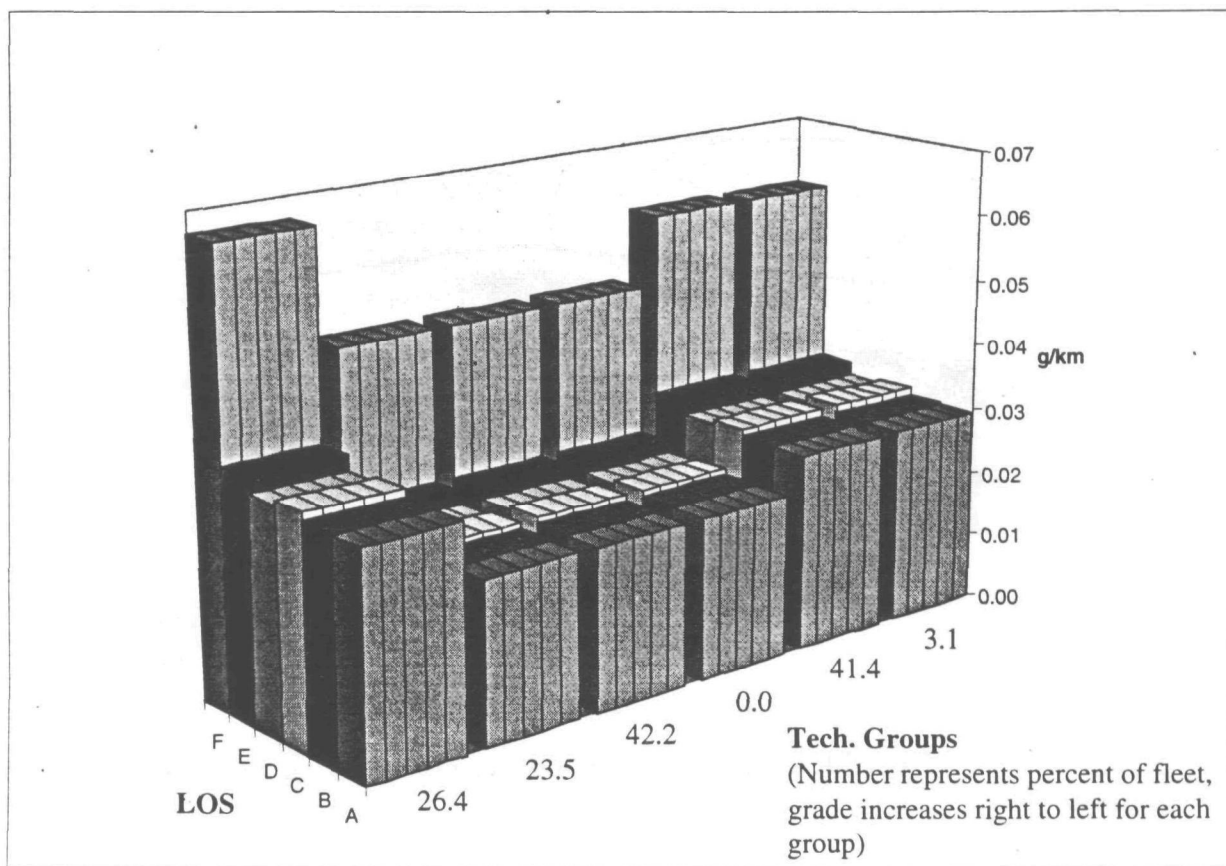
Figure 6.10 - NO$_x$ normal emitter technology group emission rates by LOS and grade

g/km

LOS

F E D C B A

Single Tech. Group
(1.4 percent of the fleet, grade
increases right to left)
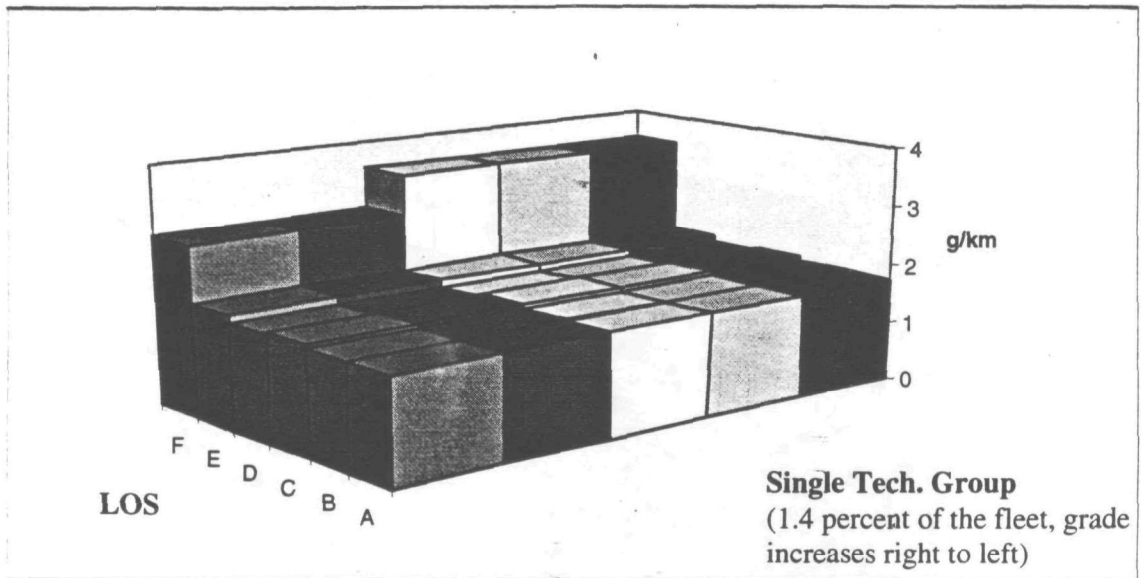
Figure 6.11 - NOx high emitter technology group emission rates by LOS and grade

6.4.4. MEASURE vs. MOBILE5a

To compare the USEPA's MOBILE5a and the new HTBR emission rates used in MEASURE, emission rates were determined for each speed and acceleration bin (0-80 mph, -10.0 to 10.0 mph/sec). Figures 6.12 to 6.17 show these profiles. Both emission rate models were used for each pollutant. The sample area results are also provided, showing a comparison between the hourly total grams of each pollutant. While much remains to be validated with MEASURE, this comparison provides some evidence for the future development of modal emission rate models.

Both data sets used in the analysis included the regional fleet distribution for the study area. The MOBILE5a rates were the running exhaust zero mile base emission rates (deterioration and start fractions effects removed). The HTBR rates used in MEASURE were similar; no start emissions or deterioration effects were included.

All the graphs show significant differences. The biggest impacts is the fact the MOBILE5a does not vary emissions by acceleration. A vehicle traveling at an average speed of 50 mph with minor variations in acceleration and deceleration is predicted to have the same emission rate as one with large variations. MEASURE indicates that these variations may have significant impacts on emission rates at certain thresholds of speed and acceleration activity.
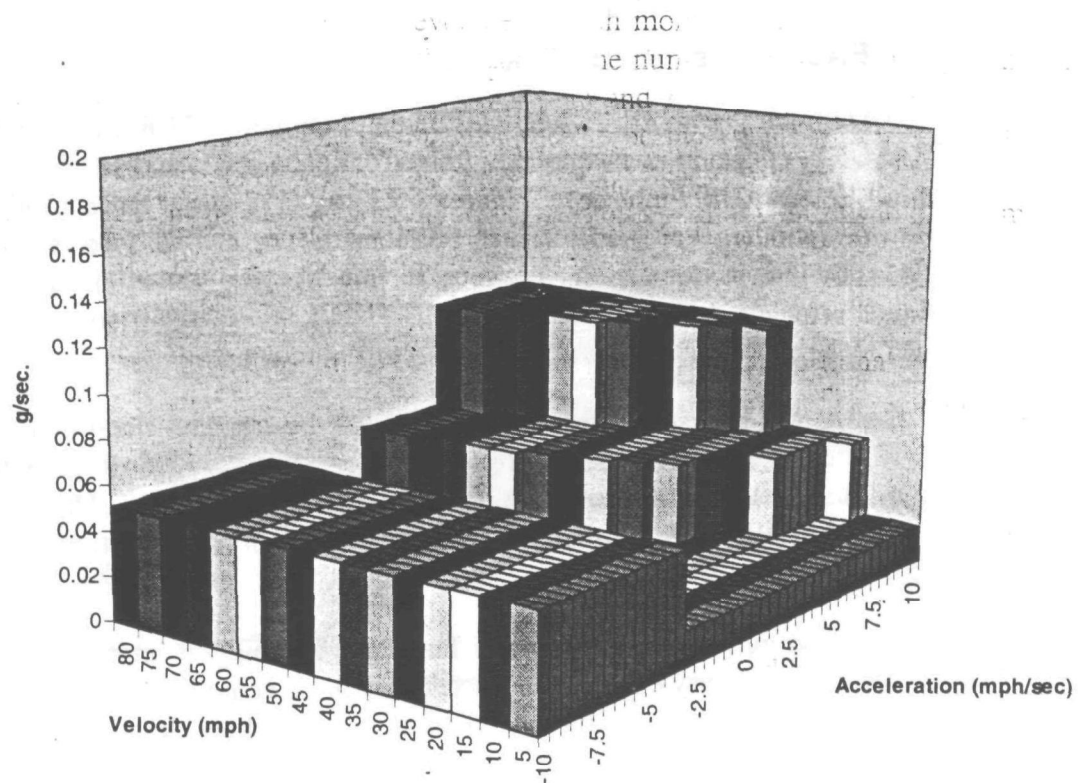
Figure 6.12 - MEASURE g/sec CO emission rates by velocity and acceleration for the study area's vehicle fleet
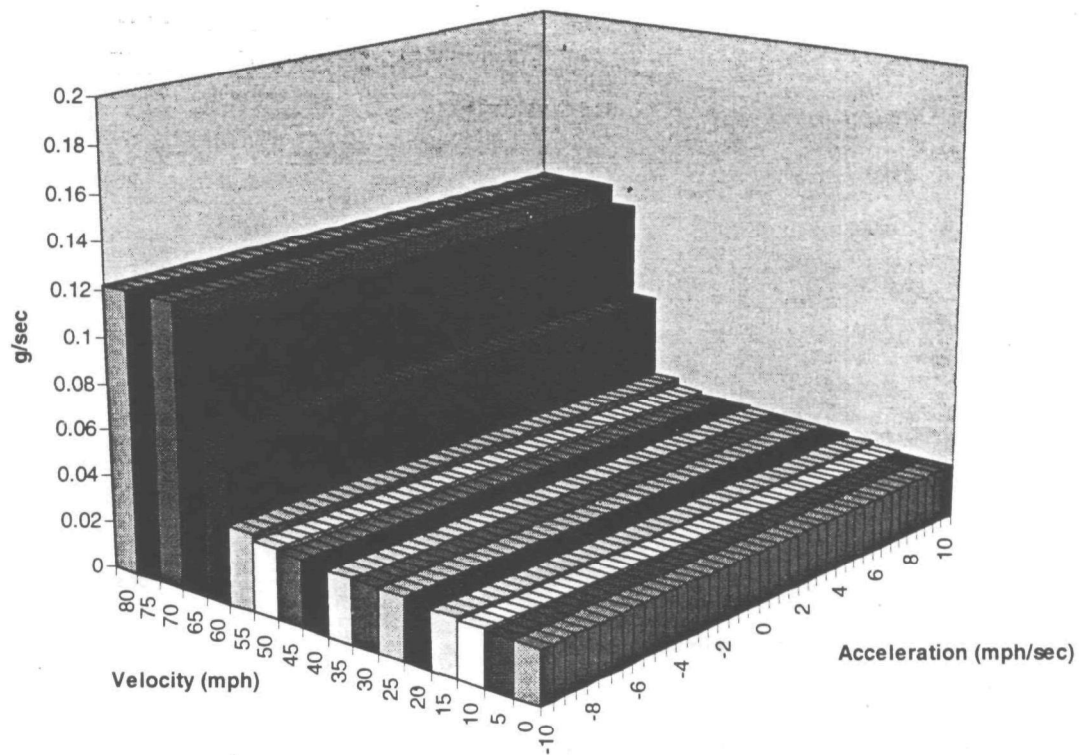
Figure 6.13 - MOBILE5a g/sec CO emission rates by velocity and acceleration for the study area's vehicle fleet
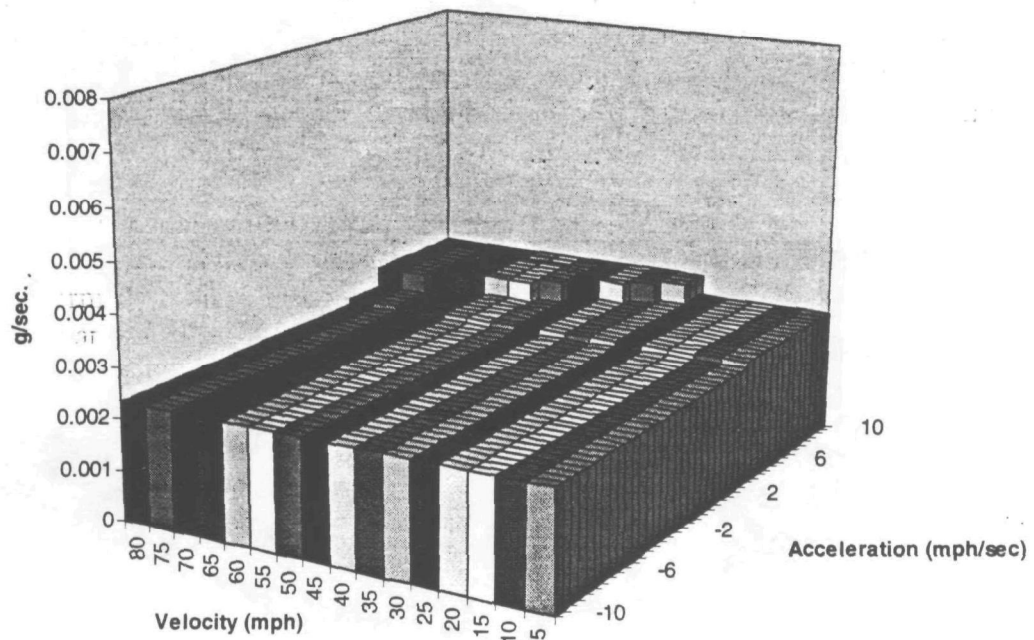
Figure 6.14 - MEASURE g/sec HC emission rates by velocity and acceleration for the study area's vehicle fleet
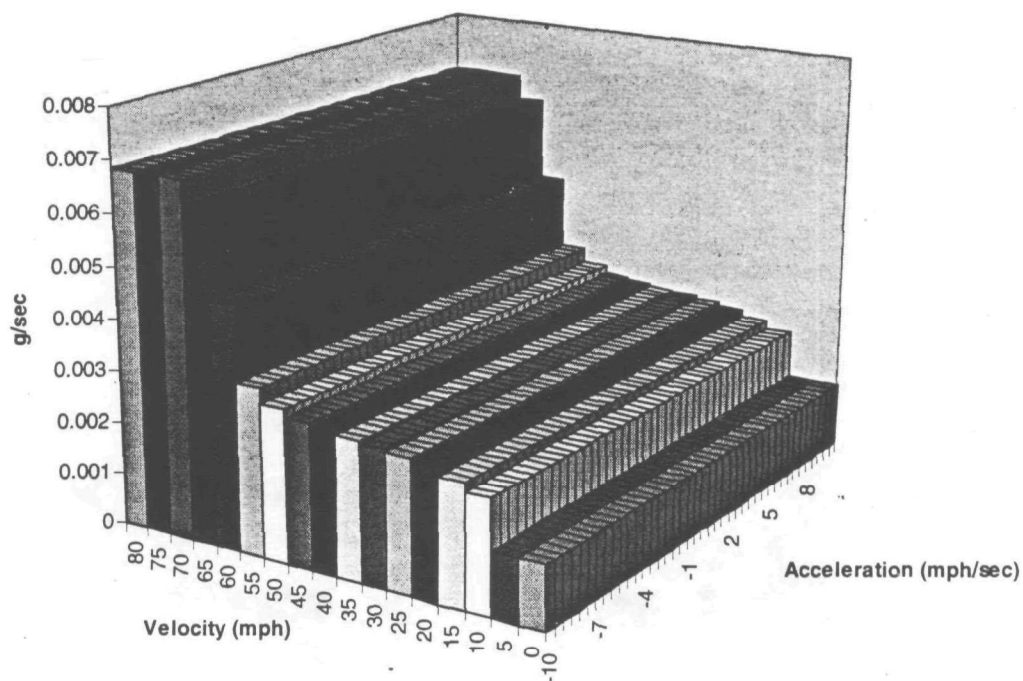
Figure 6.15 - MOBILE5a g/sec HC emission rates by velocity and acceleration for the sample area's vehicle fleet
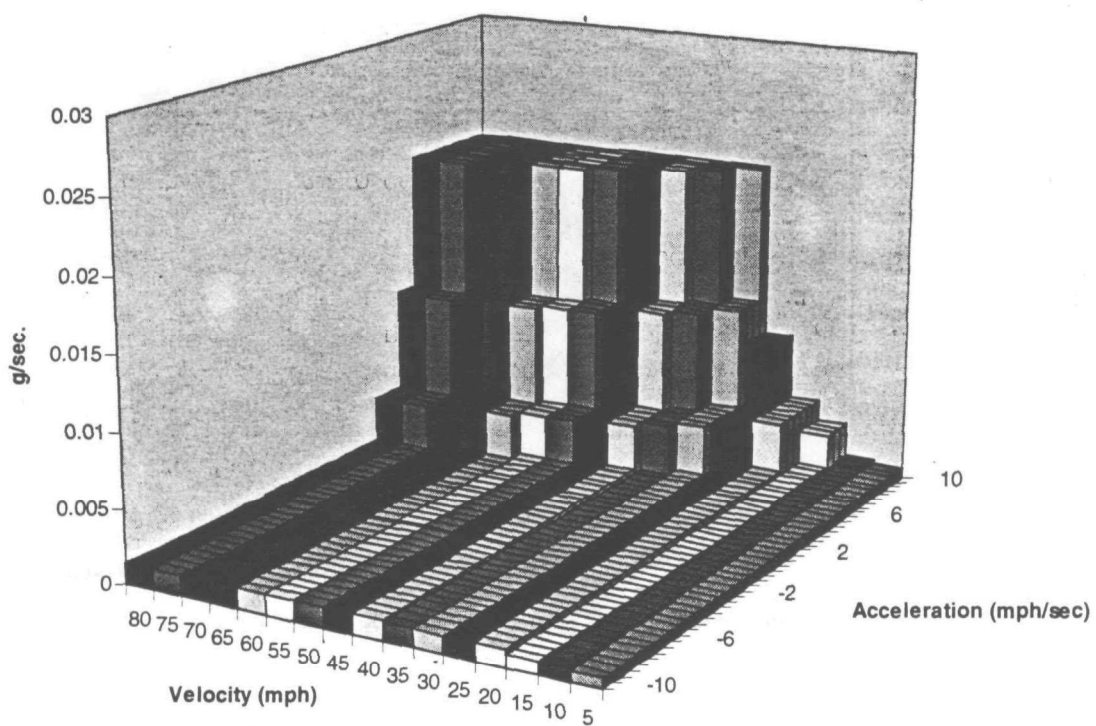
Figure 6.16 - MEASURE g/sec NOx emission rates by velocity and acceleration for the sample area's vehicle fleet
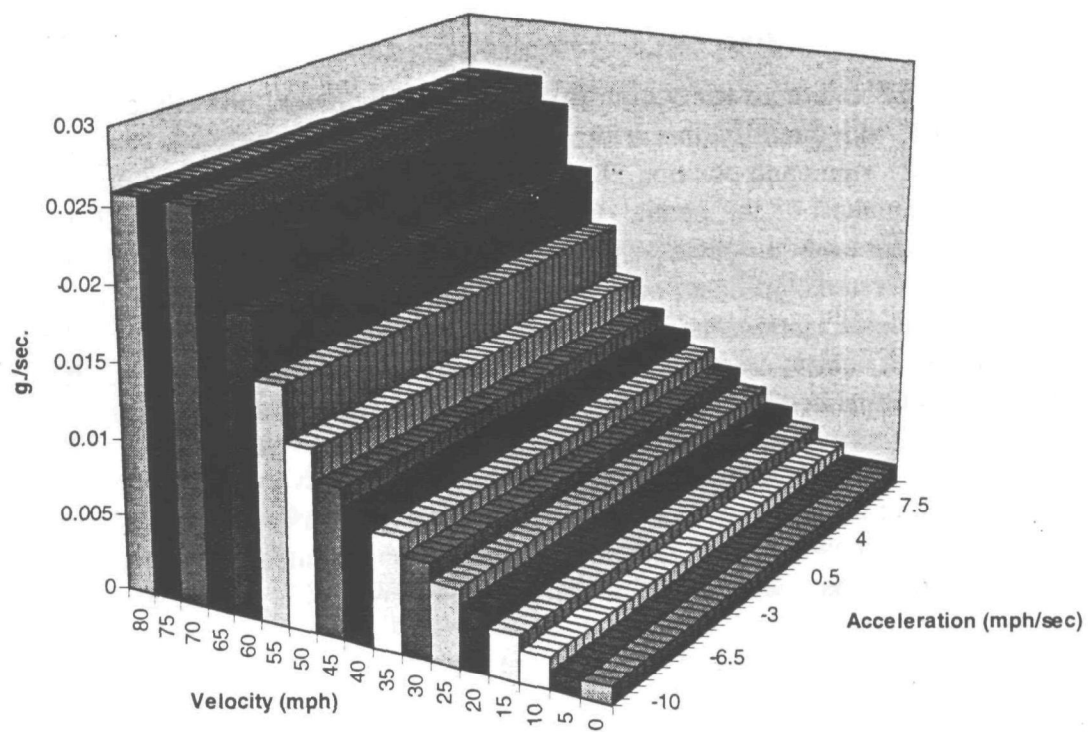
Figure 6.17 - MOBILE5a g/sec NOx emission rates by velocity and acceleration for the sample area's vehicle fleet

6.4.5. Conclusion

There are a variety of errors generated by model procedures. Future validation efforts will quantify the errors so that confidence bands can be predicted for the estimate value and position. During the process, particular attention should be paid to the estimates of the spatial variability of the operating fleet. Clearly, this model component has the greatest potential for spatial error. The use of regional temporal factors create significant non-spatial errors, particularly in off-peak hours. As long as accurate information is fed to the model, errors resulting from modeling procedures are significantly reduced (polygon overlay error, trip disaggregation error, etc.). Minimum grid cell sizes should be assessed for each model scenario.

The new modal emission rates indicate that vehicle technologies and vehicle operating profiles (speed and acceleration) have significant impacts on emission rates. While the new emission rate models need to be validated, there is strong evidence that MOBILE5a is insensitive to important emission-specific vehicle activity.

7. ANTICIPATED CONTRIBUTIONS TO MOTOR VEHICLE EMISSIONS ASSESSMENTS AND RECOMMENDED RESEARCH

This chapter will review the contribution to the fields of transportation and air quality planning presented in this report, and present a future research agenda. Individuals and groups have developed spatially-resolved emission estimates previously, but none have done so in a single, comprehensive unit. Groups have used geographic information systems (GIS) as a pre-processor, preparing data for outside modeling, and they have been used as a post-processor to help visualize results. The approach described in this report relies on the capabilities of GIS to manage the process from start to finish..

Research into transportation and air quality has made significant progress in the last several years, but results have not been formally incorporated into comprehensive and flexible modeling regimes. The proposed model framework successfully incorporates existing emission research into a single model. Further, the model is flexible to the addition of new research, an important design element needed due to the dynamic condition of transportation and air quality research. Finally, this report provides important insight into the current conditions of emission-specific spatial data, and provides a tool that can be used to define the limits of disaggregate modeling approaches. This exploration into the spatial modeling of exhaust emissions provides substantial progress towards the development of computer tools that can aid metropolitan transportation planners in their attempt to identify the impacts of transportation change on air quality.

7.1. Impacts

The potential impacts of this research are significant. It has been widely recognized that there are theoretical problems with the current modeling regime, especially with the speed correction factor emission rate models. If a spatially resolved modal emission model becomes accepted for use for conformity and inventory modeling, the types of mitigation strategies available to local and state governments change dramatically. Under the current modeling system, transportation planners and engineers have only two ways to reduce emissions: reduce vehicle miles of travel and/or optimize average speeds to points deemed significant by the models. Both options probably reduce mobility and accessibility enjoyed by the transportation

systems. If modal models are developed, much more diverse and creative strategies become available. Any strategy that reduces the number of high-emitting vehicles or reduces the occurrence of hard accelerations and decelerations will reduce mobile emissions. Reducing volume may be less important than improving traffic flow through ITS strategies, signal timing, or even lane additions. The new modal approaches may show that mobility and accessibility can increase as mobile emissions decrease.

Further, spatially-resolved estimates allow planners to prioritize certain locations for mitigation strategies because of their disproportional contribution to ozone formation due to topographic or climatic features. The spatially-resolved estimates at proper resolutions allow local transportation planners and traffic engineers to develop small-scale changes that reduce the net mobile emissions produced in their jurisdictions.

One other impact is that new car standards may be altered to reduce the occurrence of fuel enrichment resulting from high power demand. This may divert mobile emission reduction strategies away from operational conditions and more towards those strategies that impact engine starts.

7.2. Major Contributions

This section provides a discussion of how the objectives described in Chapter 1 were accomplished. The objectives were:

- *Develop an automobile exhaust emissions model that maximizes comprehensiveness, flexibility and user friendliness.*
- *Provide a research tool that allows for the testing of variable levels of motor vehicle emission model spatial aggregation.*
- *Demonstrate the benefits of using GIS for emissions modeling.*
- *Identify research and data needs for improved spatial and temporal emissions modeling.*

7.2.1. Model Design and Development

Chapter 3 lists specific model design parameters that were identified through background research. This section describes how the model is successful or unsuccessful in accomplishing those design goals.

The following model design parameters were successfully included:

- *All estimates (emissions, vehicle activity, etc.) must be capable of being validated.*
- *The model must be designed to easily incorporate new findings.*

150

- *The model must use available, or nearly available, data.*
- *The model must use as large a spatial scale as data will allow.*
- *Develop estimates of the production of automobile exhaust pollutants CO, HC, and NO_x in space and time.*
- *Separate and quantify high-emitting vehicle emissions.*
- *Include SCF emission rates.*
- *Include emission rates from the statistical approach.*
- *Include activity measures from the travel demand forecasting models.*
- *Prepare for inputs from future simulation models.*
- *Utilize geographic information systems.*
- *Appropriate documentation.*
- *Appropriate terminology.*
- *Modular system design.*
- *Open input and output data formats.*
- *Intuitive model process.*
- *Easy to understand and use.*
- *Model should reside in a GIS.*

The following model objectives were only partially accomplished:

- *The model must produce automobile exhaust emission estimates that are capable of being statistically verified.*

The resulting model is developed around data and procedures that have stochastic distributions. This factor makes the model statistically verifiable. However, actual verification of the current model results must wait until individual components are validated.

- *Anthropogenic NO_x estimate accuracy important in predicting ground-level ozone.*

NO_x is not treated in a manner distinguishing it from the other pollutants. However, the inclusion of modal parameters allows better predictions of NO_x, which varies with speed and acceleration.

- *Comprehensive representation of vehicle technologies.*

As mentioned in chapter 6, substantial differences exist between fleet averages and the ability of the MEASURE model to comprehensively represent the operating fleet in space and time. The major problems stem from the need to have a good routine for identifying vehicle characteristics given the vehicle identification number. When this is done, comprehensive representation will be accomplished. On-road vehicles distributions need more research.

- *Separate start, hot-stabilized, and enrichment emission quantities and locations.*

Start emissions are separate. Hot-stabilized and enrichment emissions are combined into the running exhaust estimates. This approach includes enrichment,

151

unlike other models, but not separately. Breaking the two modes (enrichment and running exhaust) into separate procedures requires more information unavailable for this stage of the research. However, the model framework can easily incorporate separate procedures.

7.2.2. Tool for the Exploration of Spatial Aggregation

The purpose of developing a tool that has flexible aggregation capabilities is that gridded emissions are required to predict ambient of levels of ozone and other pollutant concentrations. Future photochemical models will require a minimum one kilometer grid cell aggregate estimate of mobile source pollutant production by hour. As research into ambient air quality is conducted, the spatial scale of ozone formation and pollutant dispersion will continually be redefined, placing spatial parameters on input data. Further, emission rate models used for inventory purposes will have to have a level of spatial aggregation based on data availability and algorithm accuracy. To aid the local transportation planners in their efforts to reduce automobile pollution, models must develop accurate estimates for transportation facilities. The facility-estimates are aggregations of individual vehicles over time. Therefore, it is important that the amount of aggregation or disaggregation that can be accomplished with existing data and knowledge is identified. These issues are crucial in defining the scope of research being conducted around the country in emission modeling.

The current model has the capability to explore levels of spatial aggregation. It can use data from any size zonal aggregation, and it can re-allocate estimates to any size grid cell. Once validated, it can be used to help research the issues mentioned in the previous paragraph. The actual impacts of various levels of spatial aggregation on the accuracy of the emission estimates will vary with the spatial quality of the input data. As the spatial structure varies, so will the accuracy.

7.2.3. Value of Geographic Information Systems

Geographic information systems provide numerous advantages to the spatial modeling of exhaust emissions.

- *Spatial data organization*

Data in the model were organized based on their spatial character. Structuring the multiple layers of data in this manner provides data connectivity that would be difficult without GIS and topology.

- *Spatial data joining*

During the modeling process, datasets of different characteristics are merged together to form a single entity. Specifically, GIS allowed the travel demand forecasting model network to have improved spatial resolution by conflating to a

spatially accurate road database. This capability of GIS also allowed linkages to occur between the various area sources of information (TAZs, land use, Census, etc.).

- *Spatial query*

GIS provided the ability to search data by locational parameters. Specifically, the technique used to predict the on-road fleet distribution required for the identification of the fleet registered within a certain distance from the individual road segments.

- *Spatial aggregation*

GIS provided the ability to aggregate irregular polygon data and line data into regular user-defined grid cells. This capability makes GIS vital for efficiently developing mobile emission inventories, regardless of the modeling approach used.

- *Spatial data visualization*

The map-making and graphic display capabilities found in most GISs are extremely useful in communicating model results to individuals from various technical backgrounds. Given the importance of mobile emissions in determining transportation improvements, this feature has significant value.

7.3. Future Research

Future research is recommended in three major areas; model validation, model improvement, and model additions. Without model validation, the ultimate value of the model described in this report is lost. During model development and testing, it became evident that certain strategies could be improved with designed experiments. The following sections describe some specific actions that can be taken to expand the model for use by transportation and air quality planners.

7.3.1. Model Validation Strategies

The model validation strategies follow the modular nature of the software. Testing each module's results through designed experiments would answer the accuracy questions that would allow the model to proceed beyond the prototype stage. Prior to model validation, the input datasets should undergo significant quality assurance testing to identify errors and temporal conflicts with other datasets.

7.3.1.1.   Spatial Environment

The spatial environment module accuracy relies on the spatial accuracy of the input data, varying with new implementation sites. Procedures used in the model to

manipulate this information do not require validation. Each new modeled area should undergo a data verification and validation stage. At minimum, this should include identifying the last date the information was updated and the estimated absolute spatial accuracy of the data.

7.3.1.2. Fleet Characteristics

The process used to develop fleet characteristics needs substantial validation efforts. The intention of the modules is to develop an accurate profile of the operating fleet. The problems identified with the VIN decoder and 'lookup' routines can be solved through database and software development and are not mentioned in the validation phase. However, their effects on the accuracy of the estimate are included in the list of recommended validation efforts.

- *Zonal fleet distribution study:* A study is needed that can identify the distribution of vehicles that are registered within a zone. This could be accomplished through the ingress, egress study of a number of neighborhoods. The difficulty will be defining an appropriate sample size and definition (income, family size, etc.).
- *On-road fleet distribution study:* A study is needed that can identify the technology distributions of the on-road fleet. Currently, data capable of doing this have been collected (using video cameras) for over fifty sites. The difficulty is developing sample sets for different road classes at different times of the day. The current equipment can only be used across one lane of traffic, limiting the scope.

7.3.1.3. Vehicle Activity

Since vehicle activity estimates rely heavily on the travel demand modeling process, efforts to validate the travel model significantly improve the ability of this model to calculate errors. However, validating the engine start estimates can be accomplished through the same ingress / egress study mentioned previously.

- *Engine start activity study:* A study is needed that can identify the number of engine starts that occur by time of day within a zone. This could be accomplished through the ingress, egress study of some neighborhoods. However, this study would not need the video camera, but could rely on loop detectors. This would also allow data to be collected over a long period of time.
- *Road segment activity study:* A study is needed that can measure the volume and average speed of road segments modeled in the travel demand model by hour of the day. The Atlanta Advanced Traffic Management System (ATMS) could be used to validate interstate estimates. Other road classifications could be studied using other techniques. Again, an appropriate sample size will have to be determined.
- *Speed and acceleration profile study:* Studies that measure the speed and acceleration profiles accuracy are needed. This is currently being accomplished by researchers at Georgia Tech for interstate ramps.

7.3.1.4. Facility and Gridded Emissions

Facility emission estimates must be validated. Currently, remote sensing technology allows for road segment, hourly pollutant production estimates to be accomplished. The devices do not measure NOx, but it can be estimated using other pollutant concentrations. Measurements on multi-lane roads are difficult to do. An alternative for major roads is an upwind-downwind study where sensors are placed at regular intervals along both sides of a road. This type of study is expensive and unreliable in areas with large amounts of background pollution.

Outside of the remote sensing data collection, little can be done to directly measure emission production from operating vehicles. Many researchers are working on the issue, and technologies may develop that would make this possible.

7.3.2. Model Algorithm Improvement

There are some specific issues in the model design that could be studied to improve the accuracy of the estimates. While model validation is important in measuring current capabilities, these issues could improve the ability of the modules to accurately predict their phenomena.

- *Home location vs. registered address:* The registered dataset includes address fields that are supposed to represent the homes sites of the vehicles. Because registration tax rates differ among jurisdictions, and because people move, there is need to identify the proportion of the database that has incorrect data.
- *Fraction of total vehicle operation by vehicle type:* The registration dataset represents all vehicles that are licensed to operate on the road. The actual operating fleet may look quite different. It was evident in the two sites discussed in section 6.2.2 that the operating fleet may be much newer than the registered fleet.
- *On-road vehicle distribution search pattern:* In section 6.2.2, there was some evidence that indicated that the radial search pattern used in the model may be inappropriate for determining a local operating fleet. Research into the size and shape of the search pattern could significantly improve the capability of predicting the on-road fleet distribution.

7.3.3. Model Additions Research

The current model scope is limited to automobile exhaust emissions. Moving to a complete mobile emissions model involves adding much more information and data. Some of the major items are listed below:

- *On and off network grade distributions and impacts:* Comprehensively, the impacts of grade on engine load have not been identified in the research. Since road grade has spatial variability and could have significant impact on the load on

155

an engine, it should be included in the research design. This may mean moving to a more detailed emission rate model that has emission rates for engine load conditions.

- *More speed / acceleration matrices:* The model needs more speed and acceleration data to have a comprehensive view of modal activity on all road types. Currently, the model is limited to eleven different profiles.

- *Intersection activity:* Intersections are the one facility-type missing from the current model. Intersections will be significant to producing accurate emissions due to the extreme variability of modal activity.

- *Other motor vehicles types:* Currently, only automobiles are modeled because there are only a few vehicle emission tests for non-autos. A comprehensive mobile source model must include all vehicles types.

- *Load-based approach:* A load-base approach to predicting emissions will allow enrichment emissions to be separately identified, an original model design objective.

- *Non-exhaust mobile emissions:* Exhaust emissions only make up a portion of the overall mobile emission modes. Evaporative emissions need to be included in future models.

- *External / internal trips:* Currently, external / internal trips are excluded from the models predictions of start activity. The return trips of these vehicles are ignored, and they could represent a significant portion.

Overall, the model was successfully designed and developed according to research backed parameters. Substantial progress towards the development of a comprehensive mobile source inventory / impact model has been accomplished.

8. REFERENCES

[Awuah-Baffour 1997] Robert Awuah-Baffour, Wayne Sarasua, Karen Dixon, William Bachman, and Randall Guensler; "GPS with an Attitude"; Transportation Research Record; Transportation Research Board, Washington DC, 1998.

[Bachman 1998] Bachman, William, Jessica Granell, Randall Guensler, and John Leonard; "Research Needs In Determining Spatially-Resolved Sub-fleet Characteristics"; Transportation Research Record; Transportation Research Board, Washington, DC, 1998.

[Bachman 1996] Bachman, William, Wayne A. Sarasua, and Randall Guensler; "A GIS Framework for Mobile Source Emissions Modeling"; Transportation Research Record No. 1551. National Academy Press; January, 1996.

[Barth 1996] Barth, Matthew, Feng An, Joseph Norbeck, and Marc Ross; "Modal Emissions Modeling: A Physical Approach;" Transportation Research Record; National Academy Press; January, 1996.

[Benson 1989] Benson, Paul; "CALINE4 A Dispersion Model for Predicting Pollutant Concentrations Near Roadways" (FHWA/CA/TL-84/15); State of California Department of Transportation, Division of New Technology and Research; Sacramento, CA; November, 1984, Revised June, 1989.

[Bruckman 1991] Bruckman, Leonard, Ronald J. Dickson, and James G. Wilkinson; "The Use of GIS Software in the Development of Emissions Inventories and Emissions Modeling;" Proceedings of the Annual Meeting and Exhibition of the Air and Waste Management Association; June, 1992.

[Calspan 1973] Calspan Corporation; "Automobile Exhaust Emission Surveillance"; Environmental Protection Agency #APTD-1544; NTIS PB220-775; Office of Mobile Source Air Pollution Control; Ann Arbor, MI; May, 1973.

[CARB 1990] California Air Resources Board; Emission Inventory 1987; Technical Support Division; Sacramento, CA; March, 1990.

[Carlock 1993] Carlock, Mark; "An Analysis of High Emitting Vehicles in the On-Road Motor Vehicle Fleet"; The Emission Inventory: Perception and Reality, Proceedings of an International Specialty Conference; Air and Waste Management Association; Pittsburgh, PA; October, 1993; pp 207-228.

[Grant 1996] Grant, Chris, Randall Guensler, and Michael D. Meyer; "Variability of Heavy-Duty Vehicle Operating Mode Frequencies for Prediction of Mobile Emissions"; Proceedings of the 89th Annual Meeting; Air and Waste Management Association; Pittsburgh, PA; June, 1996.

[Groblicki 1990] Groblicki, Peter J.; Presentation at the California Air Resources Board Public Meeting on the Emission Inventory Process; General Motors Research Laboratories; Warren, MI; November 5, 1990.

[Guensler 1993a] Guensler, Randall; "Data Needs for Evolving Motor Vehicle Emission Modeling Approaches"; Transportation Planning and Air Quality II; Paul Benson, Ed.; American Society of Civil Engineers; New York, NY; 1993.

[Guensler 1993b] Guensler, Randall; "Vehicle Emission Rates and Average Vehicle Operating Speeds"; Dissertation; submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Civil/Transportation Engineering; Advisor: Daniel Sperling; Department of Civil and Environmental Engineering, University of California, Davis; Davis, CA; December, 1993.

[Guensler 1994] Guensler, Randall; "Vehicle Emission Rates and Average Vehicle Operating Speeds"; Institute of Transportation Studies, University of California, Davis, CA; 1994.

[Hallmark 1996] Hallmark, Shauna and Wende O'Neill; "Integrating GIS-T and Air Quality Models for Microscale Analysis"; Transportation Research Record No. 1551; National Academy Press; 1996.

[Harvey 1991] Harvey, G. and E. Deakin; "Toward Improved Regional Transportation Modeling Practice"; National Association of Regional Councils; Washington DC; December, 1991.

[Heywood 1988] Heywood, John B.; Internal Combustion Engine Fundamentals; McGraw-Hill Publishing Company; New York, NY; 1988.

[Jacobs 1990] Jacobs, Paul, Donald J. Churnich, and Mark A. Burnitzki; "Motor Vehicle Emissions and their Controls"; California Air Resources Board; Sacramento, CA; July, 1990.

[Johnson 1993] Johnson, Elmer; "Avoiding the Collision of Cities and Cars, Urban Transportation Policy for the Twenty-first Century"; Report by The Academy of Arts and Sciences and The Aspen Institute; Chicago, IL; September, 1993.

[Joy 1992] Joy, Richard W.; "Drive Thru Windows - A Case Study"; Proceedings of the Transportation Modeling Tips and Trip-Ups Conference; Air and Waste Management Association; Pittsburgh, PA; March 11-12, 1992.

158

[Kunselman 1974] Kunselman, P., H.T McAdams, C.J. Domke, and M.E. Williams; "Automobile Exhaust Emission Modal Analysis Model"; Calspan Corporation; Buffalo, NY; Environmental Protection Agency (EPA-460/3-74-005); Office of Mobile Source Air Pollution Control; Ann Arbor, MI; January, 1974.

[Meyer 1997] Meyer, Michael D.; "A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility"; Institute of Transportation Engineers for the Federal Highway Administration; Washington DC; July, 1997.

[Miller 1995] Miller, Terry L., Arun Chatterjee, and Cheng Ching; "Travel Related Inputs to Air Quality Models: An Analysis of Emissions Model Sensitivity and the Accuracy of Estimation Procedures"; Transportation Congress; Volume 1, pg 1149; 1995.

[NCHRP 1993] NCHRP Report 359; "Adaptation of Geographic Information Systems for Transportation"; Transportation Research Board; National Academy Press; Washington, DC; 1993.

[NRC 1991] National Research Council, Committee on Tropospheric Ozone Formation and Measurement; Rethinking the Ozone Problem in Urban and Regional Air Pollution; National Academy Press; Washington DC; 1991.

[Orthofer 1995] Orthofer Rudolph and Wolfgang Loibl; "GIS-Aided Spatial Disaggregation of Emission Inventories"; The Emission Inventory: Programs and Progress; Presentation at the Air and Waste Management Specialty Conference; RTP, NC; October, 1995.

[Outwater 1994] Outwater, Maren and William Loudon; "Travel Forecasting Guidelines for the Federal and California Clean Air Act"; Transportation Research Record; Transportation Research Board; January, 1994.

[Pollack 1992] Pollack, Alison, Jeremy G. Heiken, and Robert A. Gorse; "Comparison of Remote Sensing Data and Vehicle Emission Models: The High Proportion of Emissions from High Emitting Vehicles"; Proceedings of the 85th Annual Meeting of the Air and Waste Management Association; Pittsburgh, PA; June, 1992.

[Pozniak 1980] Pozniak, Donald J.; "The Exhaust Emission and Fuel Characteristics of an Engine During Warmup, A Vehicle Study (800396)"; Society of Automotive Engineers; Warrendale, PA; February, 1980.

[Reingruber 1994] Reingruber, Michael and William Gregory; The Data Modeling Handbook; John Wiley & Sons, Inc.; New York, NY; 1994.

[Sabate 1994] Sabate, Shelley and Archana Agrawal; "Proposed Methodology for Calculating and Redefining Cold and Hot Start Emissions"; California Air Resources Board; 1994.

[Siwek 1997] Siwek, Sarah J.; Summary of Proceedings, EPA-FHWA Modeling Workshop; Ann Arbor, MI; 1997.

[SOS 1994] Southern Oxidants Study; "The State of the Southern Oxidants Study: Policy-Relevant Findings in Ozone Pollution Research, 1988-1994"; W.L. Chameides and E.B. Cowling, eds.; North Carolina State University; Raleigh, NC; 1994.

[Souleyrette 1991] Souleyrette, Reginald, Shashi Sathisan, David James, and Soon-tin Lim; "GIS for Transportation and Air Quality Analysis"; Proceedings of the ASCE Urban Transportation Division National Specialty Conference on Transportation Planning and Air Quality; Santa Barbara, CA; July, 1991.

[Stopher 1993] Stopher, Peter; "Deficiencies of Travel-Forecasting Methods Relative to Mobile Emissions"; Journal of Transportation Engineering, ASCE; Vol. 119, No. 5; September/October, 1993.

[Tomeh 1996] Tomeh, Osama; "Spatial and Temporal Characterization of the Vehicle Fleet as a Function of Local and Regional Registration Mix: Methodological Development"; Dissertation, School of Civil and Environmental Engineering, Georgia Institute of Technology; Atlanta, GA; May, 1996.

[USEPA 1995] US Environmental Protection Agency; "National Air Pollutant Emission Trends, 1900-1994"; (EPA-454/R-95-011; NTIS PB 96-135678); Office of Air Quality Planning and Standards; Research Triangle Park, NC; October, 1995.

[USEPA 1997] US Environmental Protection Agency; Congressional Subcommittee Testimony by Carol Browner; CSPAN; Washington, DC; 1997.

[Venigalla 1995a] Venigalla, Mohan, Terry Miller, and Arun Chatterjee; "Alternative Operating Mode Fractions to Federal Test Procedure Mode Mix for Mobile Source Emissions Modeling"; Transportation Research Record No. 1472. p. 35-44; National Academy Press; Washington, DC; 1995.

[Venigalla 1995b] Venigalla, Mohan, Terry Miller, and Arun Chatterjee; "Start Modes of Trips for Mobile Source Emissions Modeling"; Transportation Research Record No. 1472. p. 26-34; National Academy Press; Washington, DC; 1995.

[Washington 1994] Washington, Simon; "Estimation of a Vehicular Carbon Monoxide Modal Emission Model and Assessment of an Intelligent Transportation

Technology"; Dissertation; Department of Civil Engineering, University of California, Davis; Davis, CA; 1994.

[Washington 1995]   Washington, Simon; "Modal Activity Correction Factors for Implementing Emission Rate Estimates"; <u>Transportation Research Record</u>; Transportation Research Board; January, 1995.

[Washington 1996]   Washington, Simon; "Considerations for the Development of New Mobile Source Emission Models"; Proceedings of the 76th Annual Transportation Research Board; Washington DC; January, 1996.

[Wolf 1998]   Wolf, Jean, Simon Washington, Randall Guensler, and William Bachman; "High Emitting Vehicle Characteristics Using Regression Tree Analysis"; <u>Transportation Research Record</u>; Transportation Research Board, Washington DC; 1998.

[Worboys 1995] Worboys, Michael F.; <u>GIS: A Computing Perspective</u>; Taylor and Francis; London, UK; 1995.

APPENDIX A

DATA DICTIONARY

The data dictionary will explain the design and function of each file and its underlying fields / attributes. It will divide the data into two components: input files and output files. Further, all files are ARC/INFO coverage files, Dbase 4 files, or ASCII text files. Chapter 4 provides a functional overview of all the input and output files. ARC/INFO coverages consist of several related files that are maintained by the software. For these databases, only the attribute file is described because it is usually the only one edited by the user. Fields in the attribute files maintained by ARC/INFO (area, perimeter, length, cover#, and cover-id) are not profiled.

The directory structure consists of 16 directories. All directory contents are discussed in detail later:

1. *aml* – Contains 'aml' scripts. These ASCII files are ARC/INFO macros that are called throughout the process.

2. *Code* – Contains 'C' programming code. The ASCII code files, and their binary object and executable files are stored in this directory.

3. *Em* – Contains emission estimate results. All final emission estimates are written to this directory.

4. *Grade* – Contains all files required to estimate the road grade for the study area.

5. *Grid* – Contains all vector grid coverages that are desired for the assessment (e.g., 1 km, 2km, 4km).

6. *Landmarks* – Contains all point data for specific landmarks (schools, universities, etc.)

7. *Lookup* – Contains all ASCII look-up files. These consist of the SCF emission rate file and Vehicle Test Weight file.

8. *Modalmats* – All the pre-processed modal matrices required for the aggregate-modal emission rates are stored here.

9. *Raster* – All the gridded, hourly, raster outputs are located here.

10. *Raw* – Contains readable copies of all the input files.

11. *Road* – Contains all data related to the road network.

12. *Sa* – Contains all the ASCII speed / acceleration matrices used to calculate average speed and modal conditions.

13. *Temp* – Initially contains nothing. But during processing, it is a working directory where temporary files are written and erased.

14. *Templates* – Contains empty data templates that speed up the writing of files with many fields. These are all INFO files.

15. *Tg* – Contains all vehicle-related information needed to determine emission-specific technology group distributions.

16. *Zone* – Contains all zonal based data (landuse, census data, TAZs, etc.).

INPUT Files:

ARC/INFO Coverages:

1. **GRID** – GRID is a vector polygon coverage that will be used as the basis for the output raster database. This database should consist of regular square sized polygons of any user-defined dimension. All attribute data are found in the polygon attribute table call 'grid.pat'. One way to get this database for the study area is to use the ARC/INFO command 'generate' with the 'fishnet' option. This database is stored in the 'grid' directory.

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	AREA	4	12	F	3
5	PERIMETER	4	12	F	3
9	GRID#	4	5	B	–
13	GRID-ID	4	5	B	–
17	GDID	4	5	B	–

The field GDID represents a unique ID field that is initially set to the internal id GRID# + 1. One is added because the first record is an external polygon representing undefined space. This undefined space is therefore assigned a 0 id. The GDID is used to track disaggregation that occurs during processing.

2. **LANDUSE** – The LANDUSE coverage is a vector polygon database consisting of residential, commercial, and other landuses. It is used to disaggregate TAZ trip information into smaller areas. For example, it is assumed that vehicle trips that begin at home and end up at work, begin in residential and end in non-residential land uses. This disaggregation allows the engine start emissions to be better spatially defined.

```
COLUMN   ITEM NAME      WIDTH OUTPUT  TYPE N.DEC
     1   AREA             4    12      F    3
     5   PERIMETER        4    12      F    3
     9   LANDUSE#         4     5      B    -
    13   LANDUSE-ID       4     5      B    -
    17   LU               3     3      C    -
```

The LU field consists of three characters describing the land use type. Acceptable values for this first version are 'RES', 'COM', or 'UNK'. 'UNK' polygons have landuses that are unknown or non-residential or non-commercial. For example, an institutional land use would be designated 'UNK'.

3. **CENSUS** – The CENSUS dataset is a vector polygon coverage of socioeconomic data. It is expected that these polygons would be Census Blocks defined in the 1990 US Census and stored in the 1994 TIGER files. Actually, they can represent other datasets (parcel level data, local economic zones, etc.). It is important that the fields be present as described below.

```
COLUMN   ITEM NAME      WIDTH OUTPUT  TYPE N.DEC
     1   AREA             4    12      F    3
     5   PERIMETER        4    12      F    3
     9   CENSUS#          4     5      B    -
    13   CENSUS-ID        4     5      B    -
    17   CBID             4     5      B    -
    21   HU90             8     8      I    -
    25   HU90/KM          8    16      F    5
    33   SOV              8    10      F    4
    41   CARPOOL          8    10      F    4
```

The CBID field is the identifier that is tracked throughout the modeling process. This field becomes a key field to link zonal activity and emission estimates. The HU90 field contains the number of housing units (1990) found within the polygon. This field is used to disaggregate home-based trips from the TAZ level to the Census Block level. The HU90/KM is the 1990 housing units per square kilometer. The SOV field is the fraction of 1990 workers that drove to work alone. The CARPOOL field is the fraction of 1990 workers that carpooled to work. The remaining trips were non-auto trips.

4. **TAZ** – The TAZ dataset is a vector polygon coverage of the local planning organization's traffic analysis zones. The zones are used in planning agencies as a spatial unit, summarizing trip origins and destinations that occur within each. The

trip estimates are used to predict the number of engine starts that occur within the zone.

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	AREA	4	12	F	3
5	PERIMETER	4	12	F	3
9	TAZ#	4	5	B	-
13	TAZ-ID	4	5	B	-
17	TZID	4	5	B	-
21	HBW_PRD	8	10	f	0
29	HBSH_PRD	8	10	f	0
37	HBGS_PRD	8	10	f	0
45	HBU_PRD	8	10	f	0
53	HBO_PRD	8	10	f	0
61	NHB_PRD	8	10	f	0
69	HBW_ATT	8	10	f	0
77	HBSH_ATT	8	10	f	0
85	HBGS_ATT	8	10	f	0
93	HBU_ATT	8	10	f	0
101	HBO_ATT	8	10	f	0
110	NHB_ATT	8	10	f	0

The TZID field represents the TAZ identifier that is tracked throughout the modeling procedures. It is a key field that is used often to link related data and subsequent estimates. The remaining fields identify trip types that are defined by the local travel demand forecasting models developed and used by local transportation planners. HBW_PRD are 24-hour home-based work productions (trips between home and work or work and home). HBSH_PRD are home-based shopping trip productions. HBGS_PRD are home-based grade-school productions. HBU_PRD are home-based university trip productions. HBO_PRD are home-based other productions (trips that begin or end at home and go to or return from someplace other than work, shopping areas, grade schools, or universities). NHB_PRD are trips that begin and end someplace besides home. All the remaining fields ending in ATT describe the attractions of each trip type.

5. **ZIP code** – The ZIPCODE dataset is a vector polygon database that represents 5 digit ZIP codes in the study area. The primary purpose of the ZIP code database is to identify vehicle type distribution locations for those vehicles whose address was unmatched. During the assessment of vehicle registration data (discussed later), individual vehicles are assigned coordinates based on their address. When the address location can not be successfully or confidently identified, the vehicle's location parameter becomes its registered ZIP code. This polygon database becomes the means of identifying location. It also provides another polygonal form for aggregating results for comparison between ZIP codes in a region.

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	AREA	4	12	F	3
5	PERIMETER	4	12	F	3
9	ZIPCODE#	4	5	B	-

```
13  ZIPCODE-ID        4    5    B    -
17  ZIPCODE           5    5    I    -
22  ZPID              4    5    B    -
```

The ZIPCODE field holds the 5-digit ZIP code number. The ZPID field is an ID that is tracked throughout the model.

6. **Allroads** – The ALLROADS dataset is a vector line database of all roads in the study area. The lines are used to identify the locations of emissions that occur as a vehicle moves through the road network. Emission estimates are estimated on a line-by-line basis. The fields are;

```
COLUMN    ITEM NAME        WIDTH OUTPUT  TYPE N.DEC
     1    FNODE#              4    5      B    -
     5    TNODE#              4    5      B    -
     9    LPOLY#              4    5      B    -
    13    RPOLY#              4    5      B    -
    17    LENGTH              8   18      F    5
    25    ALLROADS#           4    5      B    -
    29    ALLROADS-ID         4    5      B    -
    33    ARID                8    8      I    -
    41    TFID                4    5      B    -
```

The ARID field is a unique identifier for every line. This identifier is tracked throughout the modeling process and used a key field linking a number of related files. The TFID field is an identifier linking the travel demand forecasting network link. Every line in the local planning organization's network must be represented in the ALLROADS road dataset. TFID becomes a key field for linking travel demand forecasting model data and emission outputs.

7. **Landmarks** – The LANDMARKS dataset is a vector point database of schools and universities. These landmarks are special trip generators and attractors used in developing estimates travel behavior by local planners. Only grade school and university locations are used in this version.

```
COLUMN    ITEM NAME        WIDTH OUTPUT  TYPE N.DEC
     1    AREA               4   12      F    3
     5    PERIMETER          4   12      F    3
     9    LANDMARKS#         4    5      B    -
    13    LANDMARKS-ID       4    5      B    -
    17    CODE               4    4      C    -
```

The CODE field identifies whether the landmark is a grade school or a university. A code of 'G09' indicates a grade school. A code of 'D43' represents a university. The codes correspond to database definitions used by a landmark database available in many cities.

INFO Files:

1. **TDFN.DAT** – The TDFN.DAT datafile contains fields that are used in the travel demand forecasting software called TRANPLAN. This file holds the predicted road volumes, average speeds, and capacities that are output from the software. These items area used in MEASURE to identify the number of vehicles and levels of congestion for each modeled road segment.

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	TFID	4	5	B	–
5	FN	6	6	I	–
11	TN	6	6	I	–
17	ASSIGN_GR	1	1	I	–
18	DIST	4	4	I	–
22	AB_OPTION	3	3	C	–
25	ABSPEED1	4	4	I	–
29	ABSPEED2	4	4	I	–
33	ABDIR	2	2	I	–
35	ABLG1	2	2	I	–
37	ABLG2	2	2	I	–
39	ABLG3	2	2	I	–
41	ABCAP	6	6	I	–
47	ABVOL	6	6	I	–
53	BA_OPTION	3	3	C	–
56	BASPEED1	4	4	I	–
60	BASPEED2	4	4	I	–
64	BADIR	2	2	I	–
66	BALG1	2	2	I	–
68	BALG2	2	2	I	–
70	BALG3	2	2	I	–
72	BACAP	6	6	I	–
78	BAVOL	6	6	I	–

Only seven fields in this file are used by the model: TFID, ABSPEED1, ABCAP, ABVOL, BASPEED1, BACAP, and BAVOL. Fields with AB refer to the travel lanes moving in the 'from node' – 'to node' direction while BA refers to the reverse direction. The TFID field is a key field that is used to link to the ALLROADS attribute table. ABSPEED1 and BASPEED1 are the average modeled speeds in hundredths of a mile per hour. ABCAP and BACAP are the estimated capacities of the road by direction. ABVOL and BAVOL are the estimated 24-hour volumes of the road by direction. The remaining fields are not used by the model, but are maintained for future query capability. FN and TN are from and to node identifiers. ASSIGN_GR is the road classification. DIST is the actual distance of the TRANPLAN link. AB_OPTION and BA_OPTION are flags indicated road characteristics. ABSPEED2 and BASPEED2 are usually not used, but can represent average speed for a different set of conditions. ABDIR and BADIR identify the direction of each lane group. ABLG1-3 and BALG1-3 are fields that can identify factors for specific lanes or lane groups.

2. **TEMPORAL.FACTORS** – The TEMPOR AL.FACTORS (in the 'templates' directory) database contains multipliers for ourly travel activity. The file contains one record for each hour of a day, and fields for each trip purpose and one for on-road activity. For example, record 8 (representing 7-8 am) holds a value of 0.18080, meaning that 18.08% of daily on-road travel occurs during this period. These data were developed for Atlanta based on regional reports on travel behavior.

ASCII Files:

1. **SCF.CSV** – The SCF.CSV comma-delimited file is a lookup table for the running exhaust gram/second emission rates from MOBILE5a listed by 10-MPH increments and model year. It was created by running MOBILE5a for the given condition (100% fleet with certain model year, average speed of certain increment) with cold start percentages set to zero. Cold starts are calculated separately. There are five columns of data;
   - The first column is average speed in miles per hour.
   - The second column is the model year (1970-1994).
   - The third column is the CO emission rate.
     The fourth is the HC emission rate.
     The fifth is the NOx emission rate.
   This file replaces the need to run the MOBILE5a model repeatedly during MEASURE run time.

2. **ZONE.TWT and ZIP.TWT** – The ZONE.TWT and ZIP.TWT comma-delimited files identify vehicle characteristics by location. These files are the products of a series of programs that exist outside of MEASURE. The programs process department of motor vehicle registration data by address-geocoding and VIN decoding all of the records. During the address-geocoding process, individual vehicles are either successfully or unsuccessfully geocoded. Successful records are assigned a zone identifier that is the equivalent of CBID discussed previously and written to ZONE.TWT file. Unsuccessful records default to the ZIP code and are written to the ZIP.TWT file. The files are identical in structure; 10 columns of data:

   The first column contains the zonal or ZIP code identifier.
   The second column contains the vehicle identification number, or VIN.
   The third column contains the model year.
   The fourth column contains the er ission control equipment type (4 = Oxidation and Catalyst, 3 = Catalyst only, 2 - 'xidation only, 1 = none).
   The fifth column contains the fu: ivery type flag (4 = DS, 3 = Throttle Body. _ = Carburetor, 1 = Anythin:).
   The sixth column contains the cu: ich displacement of the engine.
   The seventh column contains the ' iicle test weight in pounds.
   The eighth column contains a flag or being a CO high emitter or not.

168

- The ninth column contains a flag for being a HC high emitter or not.
- The tenth column contains a flag for being a NOx high emitter or not.

3. **SZZP.ASC** – The SZZP.ASC comma-delimited file holds factors for joining technology7 group fractions from the ZIP code level to the zonal level (census blocks). Area and address-matching rates are used to factor the ZIP code fractions by multiplying both distributions by their appropriate match or failure rate. The ZIP code distribution is further factored by the area of the zone divided by the area of the ZIP code. The distributions are then combined to give the best estimate of the zonal technology group distribution. The file has five columns of data;
   - The first column contains the zone identifier.
   The second column contains the zone area in square feet.
   - The third column contains the ZIP code identifier.
   The fourth column contains the ZIP code area in square feet.
   The fifth column contains the address-geocoding success rate.

4. **GRADE.XY** – The GRADE.XY file is a comma-delimited file of points along roads that have grade data. The file is usually developed from the output of a dynamic attitude GPS device. The first column is the unique point identifier, the second column is the x coordinate (using whatever projection system is standard to the rest of the data) and the third column is the y coordinate. These locations are matched with other data and assigned to road segments based on their location.

5. **GRADE.GR** – The GRADE.GR file accompanies the previous file by providing the road grade reading at that location. The first column of this comma-delimited file is the unique identifier, and the second is the road grade reading.

OUTPUT FILES

The output files consist of a number of intermediate and final Dbase 4 files, ARC/INFO coverages, and raster GRID databases. Many of the intermediate outputs are saved for comparison, making all data available for analysis. Many files are created and processed in during individual module runs. These files are temporarily stored until the particular module is complete. The temporary files are not discussed in this data dictionary, only the outputs from each module.

Zonal Environment Module:

1. **Sz** – The SZ (start zones) database is an ARC/INFO polygon coverage that is the spatial intersection of the input polygons CENSUS, TAZ, LANDUSE, and ZIPCODE. Every SZ polygon has identifiers to the original polygon structures.
   · The SZ database is used as the spatial aggregation for estimating engine start

emissions. At the end of the model, each individual polygon has an estimate of the amount of CO, HC and NOx emissions that result from engine starts.

```
COLUMN   ITEM NAME      WIDTH OUTPUT  TYPE N.DEC
     1   AREA               4    12     F    3
     5   PERIMETER          4    12     F    3
     9   SZ#                4     5     B    -
    13   SZ-ID              4     5     B    -
    17   CBID               4     5     B    -
    21   HU90/KM            8    16     F    5
    29   TZID               4     5     B    -
    33   ZPID               4     5     B    -
    37   SZID               4     5     B    -
```

The CBID field identifies the CENSUS polygon. The TZID field identifies the TAZ polygon. The ZPID identifies the ZIP code polygon. The SZID field represents a unique identifier for each 'start zone'. The HU90/KM field is the 1990 Census households per square kilometer for that polygon.

2. **Sz.dat** – The SZ.DAT info file maintains important data for each 'start zone'. It is created by stripping attributes from the combined coverages that make up the 'start zones'.

```
COLUMN   ITEM NAME      WIDTH OUTPUT  TYPE N.DEC
     1   SZID               4     5     B    -
     5   AREA               4    12     F    3
     9   LU                 3     3     C    -
    12   HU90               4     5     B    -
    16   HU90/KM            8    16     F    5
    24   SOV                8    10     F    4
    32   CARPOOL            8    10     F    4
    40   HBW_PRD            8    10     F    0
    48   HBSH_PRD           8    10     f    0
    56   HBGS_PRD           8    10     f    0
    64   HBU_PRD            8    10     f    0
    72   HBO_PRD            8    10     f    0
    80   NHB_PRD            8    10     f    0
    88   HBW_ATT            8    10     f    0
    96   HBSH_ATT           8    10     f    0
   104   HBGS_ATT           8    10     f    0
   112   HBU_ATT            8    10     f    0
   120   HBO_ATT            8    10     f    0
   128   NHB_ATT            8    10     f    0
   136   ZIPCODE            5     5     I    -
   141   TZID               4     5     B    -
   145   CBID               4     5     B    -
   149   RES                8    10     F    0
   157   NONRES             8    10     F    0
   165   COM                8    10     F    0
```

The SZID field is the unique key field that links to the SZ polygons. AREA is the area of the SZ polygon in the current projection units. LU is the land use read from the LANDUSE polygons. HU90 is the estimated 1990 housing units for the SZ polygon. HU90/KM is the housing unit density in square kilometers. SOV and CARPOOL contain the fraction of travel to work by each type. The next 12 fields, starting with HBW_PRD, contain the estimates of the number of trips produced or attracted to the SZ polygon for each trip type. The ZIPCODE field holds the SZ polygons 5-digit ZIP code. TZID is the identifier to a TAZ polygon, and CBID is the identifier to the CENSUS polygon. RES contains the area of residential land use in square kilometers. NONRES contains the area of non-residential land use in square kilometers. COM contains the area of commercial land use in square kilometers.

Road Environment Module:

1. **Mr** – The MR database is a vector line dataset of major road centerlines in the study area. This database contains the entities that are used to aggregate emissions that occur from running exhaust on major roads. Major roads are defined as those modeled by the local planning organization's regional travel demand forecasting model. These roads are considered separately because of the prognostic data available through the forecast models. Individual lines represent road segments that start and end at crossings of other major roads.

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	FNODE#	4	5	B	
5	TNODE#	4	5	B	–
9	LPOLY#	4	5	B	–
13	RPOLY#	4	5	B	–
17	LENGTH	8	18	F	5
25	MR#	4	5	B	–
29	MR-ID	4	5	B	–
33	ARID	8	8	I	–
41	TFID	4	5	B	–

Outside normal system fields managed by ARC/INFO, each line has two identifier fields. The first is ARID. ARID is a key field link to the ALLROADS input database. The second, TFID is a key field link to the TDFN.DAT input file.

2. **Mz** – The MZ database is a polygon database of minor road aggregations. These polygons are bounded by lines from the MR database. The polygons are used to aggregate running exhaust emissions that occur of the major road network (those roads not modeled by regional travel demand forecasting models).

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	AREA	8	18	F	5
9	PERIMETER	8	18	F	5
17	MZ#	4	5	B	–
21	MZ-ID	4	5	B	–
25	MZID	4	5	B	–

The field MZID is a unique identifier that is used as a key field linking activity and emissions data estimated by the model.

Engine Start Activity Module:

1. **Sz-act.dbf** – The SZ-ACT.DBF file is a Dbase 4 file of engine start activity occurring within the SZ (start zone) polygons. The file is created by spatially disaggregating regional travel information and temporal factors from the input files.

```
COLUMN    ITEM NAME      WIDTH OUTPUT  TYPE N.DEC
     1  SZID               4     5      B    -
     5  ES1                4     5      B    -
        |
        ES24               4     5      B    -
```

The SZID field is an identifier to an SZ polygon. The ES1-ES24 fields contains the number of engine starts that occur within each ES polygon during that particular hour (ES1 = engine start from midnight to 1AM).

Running Exhaust Activity Module:

1. **Mr-act.dbf** – The MR-ACT.DBF file is a Dbase 4 file of running exhaust activity parameters that occur on each MR (major road) line.

```
COLUMN    ITEM NAME      WIDTH OUTPUT  TYPE N.DEC
     1  ARID               8     9      F    0
     9  TFID               8    11      F    0
    17  FREQUENCY          8    11      F    0
    25  SUM_SA1            8    25      F    6
    33  SUM_SA2            8    25      F    6
    41  SUM_SA3            8    25      F    6
    49  SUM_SA4            8    25   :  F    6
   .57  SUM_SA5            8    25      F    6
    65  LENGTH             8    24      F    5
    73  CLASS              8     2      F    0
    81  CAPACITY           8     7      F    0
    89  LANES              8     3      F    0
    97  SPD1-24            8    16      F    5
   289  V__1-24            8     7      F    0
   481  LO_1-24            1     1      C    -
   505  VCR1-24            8    12      F    1
```

The field ARID is a . to the ALLROADS database. The field TFID is a link to the TDFN AT file QUENCY is the number of grade data points for each record (m. road s nt). SUM_SA1 is the number of data points with grades less than -.045. SL 2 is the number of data points with grades between 0.045 and -0.015. S A3 is the number of data points with grades between 0.015 and +0.015. S _SA4 is the number of data points with grades between +0.015 ar +0.045. M_SA5 is the number of data points with grades greater

172

than +0.045. LENGTH is the length of the road segment in the current map units (meters). CLASS is the road classification. LANES is the number of one-way travel lanes for the road segment. SPD1-24 is the average travel speed for each hour in a day. VOL1-24 is the traffic volume for each hour. LOS1-24 is the level of service for each hour. VCR1-24 is the volume to capacity ratio.

2. **Mz-act.dbf** – The MZ-ACT.DBF file is a Dbase 4 file of the minor zone vehicle activity. The file stores activity information that is used in predicting emissions. The MZID field is an identifier linking the data to the MZID polygons. The MEAN_TRAVE field is the aggregate mean travel time from the centroid of each SZ polygon to the closest major road. The fields SUM_ES1-24 contain the aggregate number of engine starts (trips) that occur within the minor zone during the given hour.

Engine Start Technology Group Module:

1. **Estg.dbf** – The ESTG.DBF file contains all the emission-specific technology group distributions for each CENSUS polygon (1990 Census Blocks). The CBID field links to the CBID field in the CENSUS and SZ polygons. The FREQ field shows the total number of active vehicles estimated to reside in the polygon. The remaining fields are the distributions. Every individual vehicle is assigned a technology group for CO, HC, and NOx. Within each pollutant, vehicles are divided into high and normal emitters based on technology characteristics. Therefore, the fields are as follows:

   ESCON: Engine Start, CO, Normal
   ESCOH: Engine Start, CO, High
   ESHCN: Engine Start, HC, Normal
   ESHCH: Engine Start, HC, High
   ESNON: Engine Start, NOx, Normal
   ESNOH: Engine Start, NOx, High

2. **Esreg.dbf** – The ESREG.DBF file contains all the engine start emission-specific technology group distributions for the regional fleet. The file only contains the group fields and a frequency field.

Running Exhaust Technology Group Module:

1. **Mrtg.dbf** – This file contains aggregate modal running exhaust technology group distributions for each road segment in the MR line database. The ARID field links to the MR.AAT and the ALLROADS.AAT attribute tables. The FREQ field shows the total number of vehicles operating on the line in a 24-hour period. The technology group fields are as follows:

   HSCON: Running Exhaust, CO, Normal

HSCOH: Running Exhaust, CO, High
HSHCN: Running Exhaust, HC, Normal
HSHCH: Running Exhaust, HC, High
HSNON: Running Exhaust, NOx, Normal
HSNOH: Running Exhaust, NOx, High

2. **Scftg.dbf** – The SCFTG.DBF file is a Dbase 4 file of the Speed Correction Factor Technology Groups. These are the model year distributions used by MOBILE5a to determine appropriate running exhaust emission rates. The file links to the MR and ALLROADS line databases through the ARID identifier field. The FREQ field identifies the number of vehicles predicted to operate on that road segment during a 24-hour period. The remaining fields MY70-MY94 show the fraction of the operating vehicles in each model year.

3. **Rereg.dbf** - The REREG.DBF file contains all the running exhaust emission-specific technology group distributions for the regional fleet. The file only contains the group fields and a frequency field.

Engine Start Emissions Module:

1. **Es-em.dbf** – The ES-EM.DBF file is Dbase 4 file of the engine start emission estimates. The file contains the SZID identifier linking the estimate to the SZ polygon database. The remaining fields show the hourly estimates of each pollutant in grams. The fields are listed as CO1-24, HC1-24, and NOx1-24.

Running Exhaust Emissions Module:

1. **Mr-em.dbf** - The MR-EM.DBF file is Dbase 4 file of the aggregate modal running exhaust emission estimates. The file contains the ARID identifier linking the estimate to the MR and ALLROADS line database. The remaining fields show the hourly estimates of each pollutant in grams. The fields are listed as CO1-24, HC1-24, and NOx1-24.
2. **Mz-em.dbf** – The MZ-EM.DBF file is Dbase 4 file of the minor road emission estimates. The file contains the MZID identifier linking the estimate to the MZ polygon database. The remaining fields show the hourly estimates of each pollutant in grams. The fields are listed as CO1-24, HC1-24, and NOx1-24.
3. **Scf-em.dbf** - The SCF-EM.DBF file is Dbase 4 file of the speed correction factor running exhaust emission estimates. The file contains the ARID identifier linking the estimate to the MR and ALLROADS line database. The remaining fields show the hourly estimates of each pollutant in grams. The fields are listed as CO1-24, HC1-24, and NOx1-24.

Gridded, Hourly, Emissions Module:

1. **Grid-em.dbf** – The GRID-EM.DBF file is a Dbase 4 file of the final gridded, hourly emissions. The file contains an identifier GDID that links to the GRID

polygon database. The remaining fields show the hourly estimates of each pollutant in grams. The fields are listed as CO1-24, HC1-24, and NOx1-24.

2. **Raster Files** – The raster datasets created in this module are used for display purposes and only re-represent data previously estimated. Each file represents a pollutant-mode-hour specific estimate. Therefore the following raster files are created:

SZCO1-24
SZHC1-24
SZNO1-24
MRCO1-24
MRHC1-24
MRNO1-24
MZCO1-24
MZHC1-24
MZNO1-24
SCFCO1-24
SCFHC1-24
SCFNO1-24
TOTALCO1-24
TOTALHC1-24
TOTALNO1-24

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/R-98-097	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE A GIS-Based Modal Model of Automobile Exhaust Emissions		5. REPORT DATE August 1998
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) William Hendricks Bachman		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Georgia Institute of Technology School of Civil and Environmental Engineering Atlanta, Georgia 30332		10. PROGRAM ELEMENT NO.
		11. CONTRACT/GRANT NO. CR823020
12. SPONSORING AGENCY NAME AND ADDRESS EPA, Office of Research and Development Air Pollution Prevention and Control Division Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED Final report; 1/97 - 5/98
		14. SPONSORING AGENCY CODE EPA/600/13

15. SUPPLEMENTARY NOTES APPCD project officer is Carl T. Ripberger, Mail Drop 61, 919/541-2924.

16. ABSTRACT The report presents progress toward the development of a computer tool called MEASURE, the Mobile Emission Assessment System for Urban and Regional Evaluation. The tool works toward a goal of providing researchers and planners with a way to assess new mobile emission mitigation strategies. The model is based on a geographic information syatem (GIS) and uses modal emission rates, varying emissions according to vehicle technologies and modal operation (acceleration, deceleration, cruise, and idle). Estimates of spatially resolved fleet composition and activity are combined with situation-specific emission rates to predict engine start and running exhaust emissions. The estimates are provided at user-defined spatial scales. A demonstration of model operation is provided using a 100 sq km study area in Atlanta, Georgia. Future mobile emissions modeling research needs are developed from an analysis of the sources of model error.

17.	KEY WORDS AND DOCUMENT ANALYSIS	
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Pollution Exhaust Emissions Motor Vehicles Mathematical Models Information Systems Geography	Pollution Control Mobile Sources Geographic Information System (GIS)	13B 21B 13F 12A 09B, 05B 08F
18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 187
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)