

Centre SIM: Hour-by-hour travel demand forecasting for mobile source emission estimation

J. L. Kuhnau & K. G. Goulias
*The Pennsylvania Transportation Institute,
The Pennsylvania State University, USA*

Abstract

Urban Airshed Models for estimating air pollution concentrations in a region require travel demand models that can produce hour-by-hour mobile source emission estimates. Regional simulation models of this type are currently created and tested in Australia, Europe, Japan, and the United States using a variety of theories, decision-making formalisms, and operational implementation methods. On one hand, these relatively new conceptualizations and models of transport systems have improved in a substantial way the realism of computerized decision support tools and have the potential of improving quantification of environmental impacts and transport management/control strategies. On the other hand, however, these systems require a substantial amount of data and understanding about behavior that very often are not readily available and for this require additional research. In this paper first a brief comparative overview of conceptual designs, data requirements, and models used in computer simulation of regional transport systems is provided. Then, the basic ingredients of a model system called Longitudinal Integrated Forecasting Environment (LIFE) that contains a demographic simulator, a daily time allocation and travel scheduling system, and a Geographic Information System are also presented. One component of this model system emphasizing the spatial and temporal dimensions of travel demand is described in more detail together with its validation using observed traffic data. The paper concludes with a summary and a few directions for model improvements.

CITATION: Kuhnau, J. L., and K. G. Goulias. "Centre SIM: Hour-by-hour travel demand forecasting for mobile source emission estimation." In *Development and application of computer techniques to environmental studies*, pp. 257-266. WIT Press, 2002.

1 MOBILE SOURCE EMISSIONS MODELING

Transportation is responsible for a large portion of the air pollution emanating from fuel combustion causing morbidity, mortality, and ecosystem damage. In 1999, transportation emissions in the U.S. produced 78.6% of the total carbon monoxide (CO), of which 55.9% was contributed by highway vehicles. For the same year, transportation contributed 53.4% to the total nitrogen oxides (NO_x) produced, more than half of which (35.1% of total NO_x production) was due to highway vehicles. A little less than half (43.5%) of the volatile organic compound (VOC) emissions are from transportation with 29.6% of the total produced by highway vehicles. VOC with NO_x combine in the atmosphere and with the help of sunlight form ground-level ozone, which is the primary component of smog. Transportation continues to contribute to lead production with 12.8% of its total production in 1998, in spite of its spectacular decline since the introduction of unleaded gasoline worldwide (in 1970 transportation was contributing 82.3% of lead). Transportation's share of emissions is smaller for particulate matter, either solid or liquid, of 10 micrometers (2.1%) and 2.5 micrometers (7.6%), SO₂ (6.9%), and NH₃ (5.4%).

The legislative framework for safeguarding air quality and controlling emissions from transportation in the United States is supported by the Clean Air Act Amendments of 1970 (CAAA70), its additional amendments in 1990 (CAAA90), the Intermodal Surface Transportation Efficiency Act (ISTEA), the more recent Transportation Equity Act for the 21st Century (TEA-21) and Titles 23, 40, and 49 of the U.S. Code and Code of Federal Regulations. CAAA90 and ISTEA introduced a different "philosophy" in dealing with air pollution from transportation sources. They both required an integration of transportation plans and investments with the goal of improving air quality. For example, they mandated the application of transportation control measures (TCMs – one example are strategies and actions to decrease the number of vehicles driven alone by promoting carpooling) with clear implementation schedules and goals. In addition, they introduced the process of conformity in plan development – in essence a detailed check on compatibility among transportation plans and programs with air quality attainment of specific standards. A similar process appears to be evolving in the UK as described in Beattie et al. [1] and the European Framework Directives of 1996 such as the "Auto Oil I Programme" (for a discussion see Fenger [2]).

The CAAA90 focused on technological improvements by mandating progressively tighter vehicle emission standards, cleaner fuels, and vehicle inspection and maintenance programs. New emission standards have been required for automobiles from model-year 1994 and new bus standards were also introduced. In 1997, NO_x reduction standards were also introduced. Since air pollution is particularly acute in urban environments, it motivates many of the stipulations in CAAA90, which redefined the National Ambient Air Quality Standards (NAAQS) as maximum concentrations of pollutants that cannot be exceeded. In addition, it also gives authority to the U.S. Environmental Protection Agency to impose highway fund sanctions when non-compliance is found [3]. NAAQS specify maximum acceptable concentrations beyond which

unhealthy conditions exist and also defines the “criteria pollutants” to be regulated in order to meet the NAAQS on carbon monoxide (CO), nitrogen oxides (NO₂), sulfur dioxide (SO₂), particulate matter with aerodynamic size smaller than 2.5 μm (PM_{2.5}), particulate matter with aerodynamic size smaller than 10 μm (PM₁₀), ozone (O₃), and airborne lead (Pb). Details on the standards and their history and associated legislation can be found in Wark et al. [4].

Simulation models that link travel activity with automobile emissions are therefore needed in many applications assessing the effectiveness of policies and projects in controlling air pollution. These models are very often used to evaluate public investments in technologies and services to decrease the emission of compounds that either negatively influence air quality (e.g., oxides of nitrogen, carbon monoxide, hydrocarbons, toxic compounds) or are thought of contributing to global warming (e.g., carbon dioxide, in 1999 in the U.S. 32.8% of carbon dioxide emissions were from transportation fossil fuel consumption). These simulation models, however, need to operate at a variety of geographical and temporal scales. The Environmental Protection Agency (EPA) and the National Research Council (NRC) in the United States have identified three analytical scales of resolution that are needed within a new program called the new generation model (NGM) initiative. These three categories of model development are:

- 1) *National and regional scale of pollutant production (Macroscale)*. Applications in this area are better served by analyses at a high level of geographical aggregation (e.g., an entire county as a unit) correlating average travel activity to average emission productions in typical driving cycles. Travel activity is mainly represented by vehicle-kilometers traveled by each of a small number of vehicle classes (e.g., personal autos, buses, trucks). Factors of emission production (in grams per mile) are then multiplied by these miles “produced” to yield total amounts of pollutant emission estimates. These on-the-road automobile pollutant productions are then added to other estimates of mobile source pollutants and stationary source pollutants to develop large area estimates. In this category we find scenarios of new technology introduction (e.g., reformulated gasoline, new inspection and maintenance programs, the introduction of alternatively fueled vehicles) and estimates of their effect on total pollutant produced.
- 2) *Transportation policy and scenario evaluation (Mesoscale)*. This second area of applications targets the assessment of policies that may have smaller scale and localized effects (e.g., a city-based parking management program or other transportation control measures). Models developed in this area aim at linking vehicle activity on roadway segments and in other smaller geographic areas such as a traffic analysis zone (e.g., a city block) to emission production and policies to decrease it.
- 3) *Individual facility emission estimation (Microscale)*. This third area aims at assessing the impact of traffic management strategies for individual intersections (e.g., optimal signal timing) and corridors of groups of intersections. The U.S. EPA also includes in this scale highway-based corridors and their environmental impact assessments. In this category we

also find models that estimate individual vehicle emissions second-by-second.

Subdivision of model development in these three categories is somewhat artificial because models in one category can be used to assess projects that are at a higher level of geographical aggregation by developing interfacing procedures operating across different scales. In fact, the U.S. EPA developed a vision of model integration and model development to account for this [5]. Research is active in developing components of this NGM vision with models in two distinct specialty areas. The first, under the direct supervision and funding of the U.S. EPA, is in the emission factor estimation. The second, in the transportation activity modeling, is even more active at the microscale level leaving a gap in mesoscale model development. In this paper, we focus on transportation activity modeling targeting applications in the microscale and mesoscale categories. The model component presented here, called Centre SIM, is designed to provide in each hour of a day estimates of traffic volumes on each road of a geographic region. The degree of resolution for this model system can become extremely fine. In its current version, however, it covers a spectrum of resolution that starts from an individual intersection and its hourly traffic and moves up to the zonal and highway segment level by simulating hour-by-hour traffic flows (number of vehicles per hour per highway segment). Unlike other simulation software such as TRANSIMS and MEASURE [5], Centre SIM also simulates the hourly presence of persons in each zone. This offers a unique opportunity to integrate (on-road and off-road) mobile source emission models with stationary and area source emission models. In this way large scale telecommuting (stay at home to work) programs can be assessed in a holistic way accounting for decreases in fossil fuel consumption for travel and possible concomitant increases in fossil fuel consumption for electricity use at homes.

2 REGIONAL SIMULATION MODEL OVERVIEW

Regional transportation simulation models are typically based on computerized systems including the characteristics of the resident population, statistical models of travel behavior, and a representation of the transportation system. These model systems estimate the amount of traveling persons engage in, where the trips originate and end, what modes of transport are used, and the routes used for travel. These four groups of estimates also define the steps in one version of the models discussed here and named the four-step model. Inputs to these models include socioeconomic data about the population, the spatial distribution of activity locations such as residences, work places, and retail businesses, and the elements of the transport system such as roadways and transit lines and terminals. Further, the simulated region is divided into smaller areas known as traffic analysis zones (TAZs). The geographic information is most conveniently represented as an electronic map with associated databases embedded. The output of these models includes, but it is not limited to, traffic volumes of each roadway considered by the model, percent of trips by each mode (buses,

passenger cars, non-motorized), travel times from an origin to a destination, and average speed on each roadway segment considered.

Although there are many methods and approaches in travel simulation, they can be loosely categorized into the three levels of resolution scales described previously in this paper: macroscale, mesoscale, and microscale. The first transport simulation models were developed in the 1950's and 1960's to plan for major new infrastructure investments such as the U.S. interstate highway system. These models were very approximate since precise models were not necessary to determine the number of lanes required for a new facility. However, in the past decades significant shifts have occurred in the modeling paradigms. A brief general classification of the model categories, their data requirements, and their advantages and disadvantages are shown in Table 1.

Table 1. Travel Model Overview

Model Type	Data Required	Advantages	Disadvantages
Macroscale	Number of households, classification of households by income and size, average income, number of vehicles	Data are readily available, models are well understood and available	Imprecise predictions, cannot simulate policy changes, assumes equilibrium
Mesoscale	Classification of households by income and size, inventory of businesses by type and size, activity patterns of persons, spatial distribution of businesses and residences	Data are available, models are sometimes understood and availability varies conversely with level of detail	Significant amount of data needed for classifications and inventories, assumes equilibrium, interactions among persons and the environment are ignored
Microscale	Library of activity schedules, inventory of persons in the population, inventory of activity locations	Significant improvement over above models, time and change explicitly considered, can incorporate human adaptation and learning	Interactions among persons and the environment are ignored, data needed are rarely available, models require additional research

Centre SIM is one pragmatic application of a model that has its theoretical roots in the Longitudinal Integrated Forecasting Environment (LIFE) model system that is based on activity-theory (a philosophical framework for studying different forms of human praxis as multiple developmental processes involving

individuals, households, and their social groups). LIFE contains two parallel streams of modeling directions using two distinct sets of tools [6]. The model has been evolving from a first version that used data from past applications [7] to a less approximate and more complete version for regional simulation [8].

Trip-based macroscale and mesoscale models have been somewhat successful in their ability to predict total daily travel for a 24-hour period and average daily traffic (ADT) on major roadways. However, these coarse levels of prediction and forecasting precision are not adequate to address the spatially and temporally localized transportation problems currently faced by urban areas. The accurate distribution of total travel demand in time and space are essential to accurately estimate traffic volumes on specific roadways during peak travel hours (e.g., 7:00-8:00 am) and subsequently estimate on-road mobile source emissions. The Centre SIM model includes survey-based hourly activity patterns to account for the temporal distribution of travel and a business and residence inventory in a geographic information system (GIS) environment to replicate the spatial patterns of activity locations, trip origins, and trip destinations. The Centre SIM model is based on the traditional four-step model while incorporating explicit temporal and spatial characteristics of travel behavior.

3 CENTRE SIM APPLICATION

The Centre SIM model is a regional simulation of Centre County, Pennsylvania, a region of approximately 136,000 persons that includes The Pennsylvania State University with over 40,000 students and over 12,000 faculty and staff. The unique characteristics of the Penn State population in terms of time and space are accounted for in the simulation of the county population. The spatial foundation of the Centre SIM model is a complete inventory of businesses containing the address, business type classified by standard industry code (SIC), and number of employees of each business. The businesses are first matched to a digital map of the simulation area that includes the roadways and intersections of the transportation system, as well as the TAZs of the study area.

The business type and size were used to estimate the capacity, in terms of the number of persons that can be served, of each establishment for the activities of work, shopping, recreation, and other. The activity capacities served as a measure of attractiveness for each activity, and were aggregated to the TAZ level to obtain the attractiveness of each TAZ. Population data from the 2000 U.S. Census were then used to estimate the capacity of each TAZ for the home activity. The spatial distribution of the five activity types forms the basis of a 24-hour accounting system of the population (named the zone presence model), their activities, and their activity locations.

The spatial distribution of activity locations combined with the temporal activity patterns of six designated population segments (Penn State students, Penn State faculty, Penn State staff, unemployed, professionals, workers) are the basic elements in the development of this activity-based travel demand model. The activity patterns for each population segment were calculated from a 1996 two-day activity survey of Penn State students, faculty, and staff [9]. The activity distribution for each population segment represents the probability of

participation in each activity at each hour. Assumptions were made about the similarity of some of the population segments in order to use the Penn State activity survey for the entire Centre County population. For example, it was assumed that the population segment ‘Workers’, consisting of persons with fixed work schedules such as retail and manufacturing employees, would have similar activity patterns to those of Penn State staff that also have fixed work hours. Similarly, it was assumed that the category ‘Professionals’ would have similar activity patterns to Penn State faculty because both have flexible activity schedules. However, no local data were available about the activities of unemployed and retired persons and stay-at-home parents, so survey data from other sources were used [8].

The activity distributions multiplied by the number of persons in each population segment results in the number of persons engaged in each activity in each hour. This output, combined with the spatial attractiveness of each TAZ for each activity, results in a zone presence model consisting of the number of persons in each zone and their activities by time of day. An example of the student zone presence model output for one hour of the day is shown in Figure 1. Trip generation, the amount of travel in the simulation area, is then estimated by comparing the number of persons in each population segment, activity, and TAZ in consecutive hours. A greater number of persons in a TAZ and activity in one hour compared with the next hour indicates a trip must have begun from the TAZ of interest during the time period.

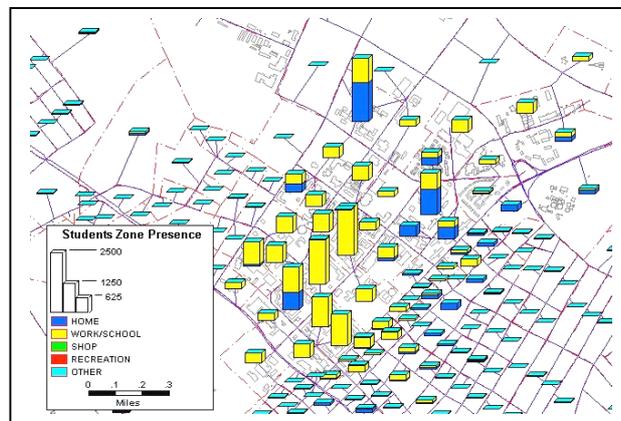


Figure 1: Zone presence output for time segment 9:00-10:00 am. Source: [8].

However, this method will underestimate the total number of trips because people within a population segment are indistinguishable and this method calculates only the minimum number of trips that must have occurred. For example, if there are 10 persons at work in a zone at one time point and 10 persons at work at the next time point, it is impossible to determine if the same people are present during the entire period (0 trips), whether some portion of the group left and were replaced by other persons (>0 trips), or if all persons left and

were replaced by other persons (10 trips). In addition, this method of trip generation misses short trips with origins and destination in the same zone and time segment. To solve both these problems, conditional travel probabilities were estimated from the survey data. The probabilities were calculated as the probability of making a trip during a time period given the population segment and the current activity. For example, a Penn State staff member at work at 9:00 am, the start of the workday, is unlikely to make a trip during the next hour and the corresponding travel probability is 0.08, reflecting this unlikelihood.

Hourly trip generation was estimated as the number of trips that have been made plus the conditional travel probabilities multiplied by the unknown persons (i.e., unknown in terms of whether they made a trip in the time period). Mode split factors, the percentages of travel by available transport modes, were then applied by population segment, resulting in the number of vehicle-trips per hour. The mode split factors were estimated by population segment and time of day based on the same Penn State survey data used for the activity patterns. The time-of-day and population considerations are key elements of the mode split step due to the almost-exclusive use of the transit system by students, a captive market because of limited parking facilities on the Penn State campus.

Both trip distribution, the spatial pattern of trip origins and destinations, and traffic assignment, the route chosen for each trip, are based on travel times on the roadway network. However, travel times on most links are very different at 4:00 am than at 4:00 pm. Three network conditions, with corresponding travel times, were identified to more accurately simulate destination and route choice in the Centre SIM model: the peak network, with maximum vehicle volumes and longest delays; the mid-day network, with a significant number of vehicles but less delay than during peak periods; and the off-peak network, with low vehicle volumes when vehicles can easily travel at or above the speed limit. The use of these three distinct networks more accurately simulates the destination and route choices of travelers.

Traffic volumes were next assigned to the appropriate network for each hour, resulting in hourly directional volumes on each roadway. These volumes were compared with observed data to validate the model output. For the Centre SIM model, mean average hourly directional traffic counts at 50 locations were used for validation. The geographic peaking and directional characteristics are important improvements of the Centre SIM model over traditional aggregate trip-based models that consider only total daily roadway volumes (ADT). Unacceptable differences between model output and observed values were defined as output volumes more than two standard deviations away from the mean observed volumes. Where unacceptable model outputs were identified, calibration adjustments were made. Calibration adjustments included changes in link travel times, turn penalties (representing the delay at intersections), mode split factors, and activity patterns.

The final Centre SIM model results demonstrate its improved accuracy over aggregate models. Traditional models typically produce as output roadway average daily traffic, which does not account for temporal or directional peaking. In addition, even recent validation studies report as target acceptable differences between observed traffic volumes and predicted traffic volumes by these models

between 7% and 25% (varying by the type of roadway). For individual roadway segments' daily volumes the desirable percent deviation between observed and predicted volume is 21% to 60% with the low end for high volume roadways and the high end for low volume roadways [10]. For the roadways that were checked, the Centre SIM model predicted ADT with 3.6% to 13.0% error. This result demonstrates that the model can predict total daily travel with acceptable precision. In addition, the model accurately simulates, for example, link volumes during the peak hour with an average of 20% error compared to previous aggregate models that showed 40% prediction error for the same links in the peak hour. An example of the model output for the evening peak hour is shown in Figure 2.

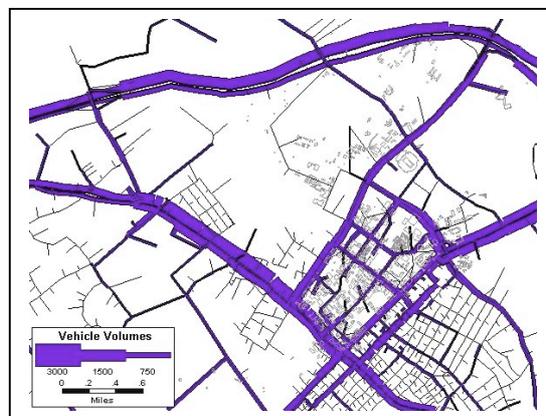


Figure 2: Travel demand model for time segment 5:00-6:00 pm. Source: [8].

This is the first model to incorporate spatial and temporal models within the traditional aggregate four-step model system, and it is clear that these are significant improvements to the travel demand modeling practice. In addition to the model's improved prediction capabilities and accuracy, the spatial and activity foundations of the model make it sensitive to land use policy changes and travel demand management scenarios. Finally, the Centre SIM model has a flexible structure that can accommodate future additions, changes, and improvements. For example, an explicit land use simulator could be used for long-range predictions and a mobile emissions model can be added to predict air quality during peak hours. The development of the NGM will be key for this capability. The explicit consideration of the spatial and temporal characteristics of travel in the Centre SIM model provides the precision necessary to quantify the effects of congestion in terms of its impacts on the environment. In addition, it is a considerable improvement over past applications designed for smaller urban areas.

The model presented here however is limited in a few critical ways. First, a good portion of the population was not surveyed and its behavior was estimated using data from elsewhere. Second, a more extensive validation is needed to

detect localized errors due to assumptions that have not been tested. Third, to improve travel prediction by time-of-day a finer degree of temporal resolution is required (e.g., 15 minutes instead of 1 hour) to capture all the trips made by the resident population and the visitors to the study area. Fourth, goods movement was completely neglected in this application and for some periods in a given day and specific highway links trucks may be the largest contributors to traffic flow. Fifth, for forecasting a battery of other models is needed to simulate population demographics and land use changes. All these are current topics of model development in Centre SIM.

References

- [1] Beattie, C.I., Longhurst, J.W.S., & Woodfield, N.K. Air quality management: evolution of policy and practice in the UK as exemplified by the experience of English local government. *Atmospheric Environment*, **35(8)**, pp. 1479-1490, 2001.
- [2] Fenger, J. Urban air quality. *Atmospheric Environment*, **33(29)**, pp. 4877-4900, 1999.
- [3] Savonis, M.J. Toward a strategic plan for transportation- air quality research, 2000-2010. *Transportation Research Record*, **1738**, pp. 68-73, 2000.
- [4] Wark K., Warner, C.F. & Davis, W.T. *Air Pollution. Its Origin and Control*, Addison-Wesley: Menlo Park, CA, 1998.
- [5] United States Environmental Protection Agency. EPA's new generation mobile source emissions model: initial proposal and issues. Report EPA420-R-01-007, 2001.
- [6] Goulias, K.G. A longitudinal integrated forecasting environment (LIFE) for activity and travel forecasting. *Ecosystems and Sustainable Development III*, ed. Y. Villacampa, C.A. Brebbia & J-L. Uso, WIT Press: Southampton, UK, pp. 811-820, 2001.
- [7] Goulias, K.G., & Kuhnau, J.L. Computer simulation for regional transport systems: overview and an example. Presented at the Transport in the 21st Century International Conference, Warsaw, Poland, September 15-19, 2001.
- [8] Kuhnau, J.L. "Activity-based travel demand modeling within the urban transportation planning system." Masters thesis, Department of Civil and Environmental Engineering, The Pennsylvania State University, USA, 2001.
- [9] Alam, S.B. "Dynamic emergency evacuation management system using GIS and spatio-temporal models of behavior." Masters thesis, Department of Civil and Environmental Engineering, The Pennsylvania State University, USA, 1998.
- [10] Barton-Aschman Associates, Inc. & Cambridge Systematics, Inc. *Model Validation and Reasonableness Checking Manual*. Report FHWA-EP-01-023. Travel Model Improvement Program, Federal Highway Administration: Washington, D.C., 2001.