DEVELOPMENT OF SOFTWARE FOR SPATIALLY AND TEMPORALLY RESOLVING MOTOR VEHICLE ACTIVITY DATA

FINAL REPORT
CONTRACT NO. 98-322

PREPARED FOR:

CALIFORNIA AIR RESOURCES BOARD
RESEARCH DIVISION
1001 I STREET
SACRAMENTO, CA 95814

PREPARED BY:

YUE-HONG CHOU, PRINCIPAL INVESTIGATOR
DEPARTMENT OF EARTH SCIENCES
UNIVERSITY OF CALIFORNIA, RIVERSIDE

JULY 31, 2000

For more information about the ARB's, Research Division's research and activities, please visit our Website:

http://www.arb.ca.gov/research/research.htm

DISCLAIMER

The statements and conclusions in this Report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

ACKNOWLEGDMENTS

This report was submitted in fulfillment of ARB contract number 98-322: Development of Software for Spatially and Temporally Resolving Motor Vehicle Activity Data by University of California, Riverside, under the sponsorship of the California Air Resources Board. Work was completed as of August 15, 2000.

Researchers at California Air Resource Board that made significant contributions to this project include Mark A. Carlock, Michael T. Benjamin, Pablo Cicero, Jeffrey R. Long, and Hector Maldonado. Rick Everett, a graduate student at UCR, and Darby McMiller who graduated from California State University, Fullerton, helped in data processing and programming of the application. Mike Vuong of GeoSpatial Technologies provided voluntary services to revise the StreetLinker software for this project.

TABLE OF CONTENTS

ABSTRACT

The objective of this project is to develop a customized Geographic Information System (GIS) software application to spatially and temporally allocate Global Positioning System (GPS) instrumented vehicle activity data and provide summary statistics at a range of spatial and temporal resolutions. Major research activities were made in three areas: (1) correct interpretation of GPS-instrumented vehicle data in terms of street networks and accurate translation of vehicle location data into meaningful routing information necessary for construction of trip patterns, (2) effective organization of both digital street maps and vehicle data in a comprehensive database required for compilation of critical statistics for analysis of mobile source emissions at desirable geographic configurations, and (3) incorporation of mapping and database management capabilities into a user-friendly, automated procedure suitable for analysis of the spatially and temporally resolved motor vehicle activity based on the huge volumes of GPS-instrumented vehicle data. A analytical application is built on ESRI's ArcView GIS for ARB researchers to carry out a wide range of motor vehicle travel data analyses.

EXECUTIVE SUMMARY

Background

Accurate emission inventory models are critical to the development of air pollution control strategies and regulations in California. This was formally recognized in September, 1996 with passage of SB2174 (Polanco Bill), requiring the Air Resources Board (ARB) to approve and validate emission inventories for mobile, stationary, areawide, and biogenic sources every three years. The primary undertaking of this project is to develop a customized Geographic Information System (GIS) software application to spatially and temporally allocate Global Positioning System (GPS) instrumented vehicle activity data and provide summary statistics at a range of spatial and temporal resolutions. This software application is designed to operate seamlessly with existing GIS software owned by the Air Resources Board (ARB), represents a critical step in facilitating spatial allocation of mobile source emissions in California at a sub-county level. Availability of such a powerful and user-friendly software application enables the in-house collection and analysis of the large volumes of second-by-second GPS data necessary for development of a statistically robust motor vehicle activity data set. As such, the result of this project provides an important new tool for improving the spatial and temporal resolution of the motor vehicle emissions inventory, as well as assist in conformity analyses.

METHODS

The first step of this project was to determine if there existed a software package suitable for ARB to use for analysis of GPS vehicle travel data. A software application developed by TransCore was carefully evaluated in terms of software suitability, street matching accuracy, and system requirements. At the time of evaluation, the TransCore software was not a finished product and required major modifications. Relatively minor modifications needed include (1) streamlining the entire operation so that the user does not need to run functions of p-line connecting, path smoothing, street matching, and summary generating separately for each single trip file, (2) providing a batch process so that one can process multiple trip files at once, (3) converting the obsolete "prn" DOS output format into a common Windows format such as "dbf" or "mdb". Major modifications of the software required for this project include (1) developing an editing mechanism so that the TransCore software's matching errors can be interactively corrected by the user, (2) revising the street matching algorithm to recognize one-way road segments, (3) enhancing the street matching algorithm to process attribute data, instead of just the graphic objects only, and (4) developing a direct ArcView interface rather than just generating ArcView shapefile output.

The most critical concern about the TransCore software is that it would not be ready for this project because its DOS version never worked while its Windows conversion was still at a beta testing stage in 1999, even though its development started in 1995-96. It would require a minimum of six months, and more likely one year or longer, for the TransCore software to become useful for this project. The Windows version evaluated contained many hard-coded common variables and parameters, such as file names and path directories, thus a ready-to-use product would take TransCore quite some time to accomplish.

The only alternative solution for this project was to use StreetLinker of GeoSpatial Technologies. StreeLinker was a finished product with all the required functions for this project. The only additional work needed is that, because StreeLinker was a MapInfo application, conversion of the interface into an ArcView application would be required, . The MapInfo to ArcView conversion of the StreetLinker interface could be done within two to three months and thus the use of StreeLinker fits the project's schedule well. The research team also determined that ESRI's ArcView should be used as the GIS tool for system development and Etak digital street maps should be adopted for this purpose.

The second step of this project involved development of the ArcView-based GIS application for activity data allocation. The design of the system takes into consideration the accuracy of final vehicle data output, data processing speed, hardware requirements, future modification and update, and friendliness for ARB users. Configuration of the system is based on the requirements that the developed system will not require additional software, efficiency of spatial data organization, and organization of trip data and analytical results. The system was developed using the Avenue programming language and tested intensively using the empirical vehicle travel data provided by ARB researchers.

The final step of this project involved installation of the developed software system at ARB's El Monte facility, documentation of the application for both regular users and ARB's in-house Avenue programmers, training for ARB researchers to learn to operate the system, and ongoing technical support.

RESULTS

Spatial data obtained from multiple sources, including state and county boundaries, air basins, districts, ARB's 5KM grid, and Etak's street maps, were organized into a comprehensive GIS database. Digital street maps of all the fifty eight counties of California were processed into a StreetLinker Multiple-Regions data structure, with all the road classifications re-defined to meet ARB's specific requirements. Other spatial data layers incorporated in this project include state boundaries, county boundaries, air basins, air districts, and ARB's 5KM grid. All the street segments, originally organized by county, were overlaid onto each of the four map layers of geographical divisions (i.e. County, Basin, District, and 5KM Grid), and re-constructed so that every segment bears unique county, basin, district, and 5KM grid code. The data structure also includes organization of raw GPS data, processed vehicle history files, trip tables, and output statistics. Spatial data processed in this project are illustrated in Fig. 1 below.

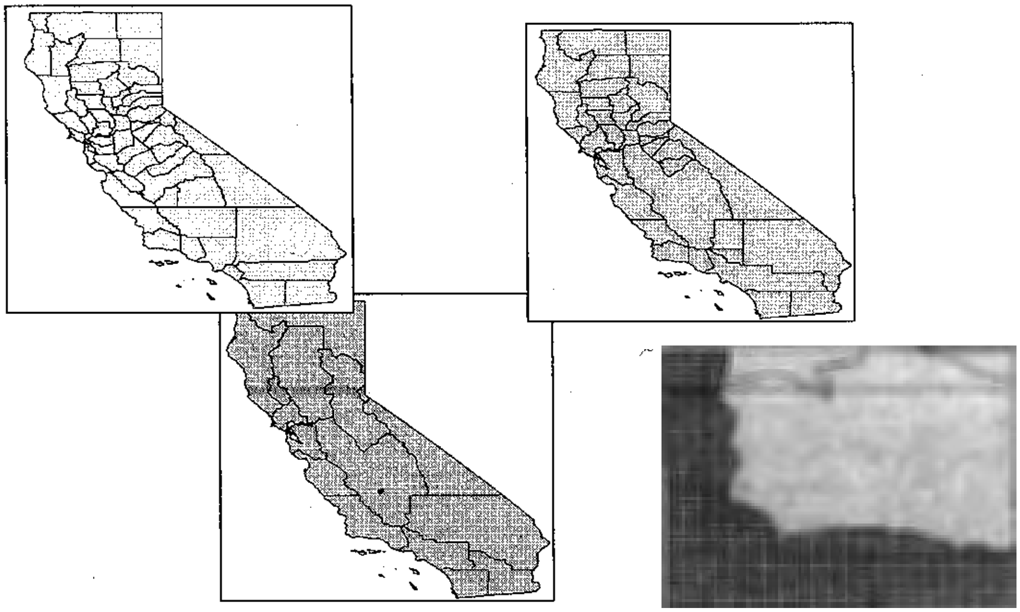


Fig. 1 County, basin, district, and 5KM Grid were the spatial data

Fig. 1 shows the processing of the spatial data, including the county boundaries, air basins, districts, and the statewide 5KM grid. All these data layers were overlaid to define spatial resolution of vehicle data analysis. In addition, the statewide Etak street maps, Fig. 2, were overlaid to the spatial data so that every segment contains unique information about county, basin, district, and 5KM grid.

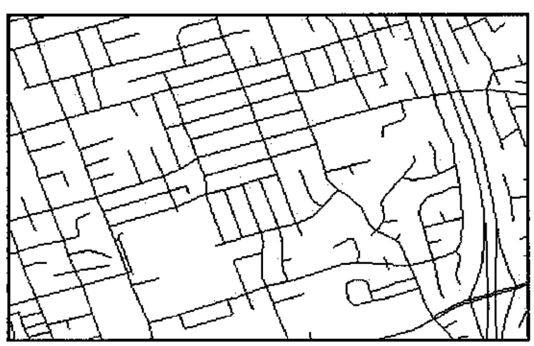


Fig. 2 Statewide Etak street maps were processed for the analysis

Each trip is a set of GPS data collected from one vehicle on a specific day. The GPS points of the trip are first evaluated for validity, filtered by a user-defined time or distance threshold, and then converted into a point coverage for storage and map presentation. Using point to line matching algorithms that the principal investigator has previously developed, each GPS point is then converted into a point location on a most-likely near street. The converted point (Linked Point) is also organized in the GIS for further analysis and map display. Each trip is then converted into a series of street segments that form a route pattern on the GIS display.

The GIS application provides two major analytical functions: trip analysis and statistical analysis. The trip analysis function allows for ARB researchers to select any number of vehicle trips and generate travel-related statistics of the selected trips. With this function, the researcher may cross-examine one specific trip or generally analyze any set of trips. Selection of trips can be based on either spatial or temporal resolution at the existing scale. The statistical analysis function allows the researcher to conduct statistical analyses of vehicle travel patterns, such as VMT by speed by facility type. The results of trip analysis can be displayed on the ArcView GIS, and the results of statistical analyses can be organized into files and tables for further study.

The developed system allows for ARB researchers to (1) convert vehicle GPS data into a database for trip and statistical analyses, (2) generate multiple map layers including original GPS point locations, point locations interpreted by street network, and street segments that form each trip, (3) conduct trip analysis to show basic statistics associated with each trip and any set of selected trips, and (4) conduct statistical analysis to obtain statistics of vehicle travel data both spatially and temporarily.

Vehicle Activity Stats (County) By Speed By Facility

Stats	Count	Sum_Vmt	Sum_Num_starts	Sum_Soak_time	Sum_Idle_time
San Diego - Collector 55 - 60	1	0.0461	0.0000	0.0000	0.0000
San Diego - Collector 60 - 65	1	0.1373	0.0000	0.0000	0.0000
San Diego - Collector 65 - 70	4	0.2035	0.0000	0.0000	0.0000
San Diego - Freeway 0 - 5	2	0.1400	0.0000	0.0000	6.4500
San Diego - Freeway 10 - 15	4	0.2965	0.0000	0.0000	4.9000
San Diego - Freeway 20 - 25	3	0.5348	0.0000	0.0000	0.0000
San Diego - Freeway 25 - 30	3	0.8787	0.0000	0.0000	0.0000
San Diego - Freeway 30 - 35	4	0.7724	0.0000	0.0000	0.0000
San Diego - Freeway 35 - 40	4	1.0270	0.0000	0.0000	0.0000
San Diego - Freeway 40 - 45	5	0.6770	0.0000	0.0000	0.0000

Table1. Sample StreetLinker table of VMT by county by speed by facility type

INTRODUCTION

Accurate emission inventory models are critical to the development of air pollution control strategies and regulations in California. This was formally recognized in September, 1996 with passage of SB2174 (Polanco Bill), requiring the Air Resources Board (ARB) to approve and validate emission inventories for mobile, stationary, areawide, and biogenic sources every three years.

The two major inputs to the ARB's motor vehicle emission inventory model (MVEI) are emission factor and activity data. Currently, activity data inputs to MVEI, such as vehicle miles traveled (VMT) and number of vehicle trips and starts, rely heavily upon travel demand model output. However, instrumented vehicle studies suggest activity diary and telephone surveys, which provide the data inputs to many transportation models, may not accurately reflect the actual driving behavior of participants due to their reliance on participant recall. Another shortcoming of conventional activity diary and telephone survey methods is their inability to record the exact location of trip starts and ends, as well as the specific roadway links traveled and the speeds on those links. Development of an activity data collection tool capable of accurately recording not only how people drive, but also where and when they drive, would improve the accuracy of transportation models and consequently the spatial and temporal allocation of motor vehicle emissions.

The ARB has recently completed the latest update to the state's mobile source emission inventory model. This latest update is a significant improvement over the previous version of the model in that many parameters have been resolved spatially to a county level and temporally to an hourly basis. For example, based on analysis of a large instrumented vehicle data set, 24 hour activity matrices have been developed for the emission inventory model to better characterize the temporal distribution of such activity data parameters as trip starts, trip ends, time-on, time-off, resting time, and vehicle miles traveled (VMT). Although this represents a significant improvement in the temporal resolution of motor vehicle activity in California, it is limited spatially to a statewide resolution due to the inability to track the locations of the instrumented vehicles used.

A number of studies, including ARB in-house studies, have shown the potential of temporally and spatially resolving motor vehicle activity data by overlaying GPS-instrumented vehicle data on digital street maps using a Geographic Information System (GIS). A critical step in this process is matching the vehicle position, as measured by the GPS, to the digital street representing the route driven. Due to positional errors in both the GPS and the digital street map, accurate map matching requires either human operator intervention or sophisticated software algorithms. As thousands of hours of second-by-second GPS-instrumented vehicle data are necessary to adequately reflect motor vehicle activity in California, analysis of the data using human operator intervention is impractical.

The first step of this project is to thoroughly evaluate a software application that is currently available for automated analysis of GPS-instrumented vehicle data. The application was developed by TransCore (formerly JHK & Associates). A careful review of the software, jointly conducted by the principal investigator and ARB researchers, suggested that the TransCore software is not suitable for this project for the following reasons.

At the time of evaluation, the TransCore software was not a finished product and required major modifications. Relatively minor modifications needed include (1) streamlining the entire operation so that the user does not need to run functions of p-line connecting, path smoothing, street matching, and summary generating separately for each single trip file, (2) providing a batch process so that one can process multiple trip files at once, (3) converting the obsolete "prn" DOS output format into a common Windows format such as "dbf" or "mdb". Major modifications of the software required for this project include (1) developing an editing mechanism so that the TransCore software's matching errors can be interactively corrected by the user, (2) revising the street matching algorithm to recognize one-way road segments, (3) enhancing the street matching algorithm to process attribute data, instead of just the graphic objects only, and (4) developing a direct ArcView interface rather than just generating ArcView shapefile output.

The most critical concern about the TransCore software is that it would not be ready for this project because its DOS version never worked while its Windows conversion was still at a beta testing stage in 1999, even though its development started in 1995-96. It would require a minimum of six months, and more likely one year or longer, for the TransCore software to become useful for this project. The Windows version evaluated contained many hard-coded common variables and parameters, such as file names and path directories, thus a ready-to-use product would take TransCore quite some time to accomplish. A more detailed explanation of the evaluation is provided in a later section.

Alternatively, another software package designed for GPS data analysis, StreetLinker developed by GeoSpatial Technologies (GST), was adopted for this project. StreetLinker was previous developed as a GIS application based on the MapInfo GIS. The main advantage of using StreetLinker is that is already contained all the functions needed for thie project, except that it was a MapInfo application. During this project period, programmers of GST converted the MapInfo version StreetLinker into ArcView version to support this project.

A related issue is to determine which GIS system should be used for this project. Comparisons between ArcInfo and ArcView, both are ESRI's GIS products, suggested that ArcView is more suitable for ARB researchers and should be adopted for building the GIS application for this project.

The GPS data analysis software outputs such activity parameters as average speed and length of idle on a link- or trip-specific basis. However, in order to develop mobile source emission inventories at other spatial and temporal resolutions, another software application is required which will aggregate the vehicle activity data output and generate statistics of greater value to mobile source emissions modeling.

The primary undertaking of this project was to develop a customized Geographic Information System (GIS) software application to spatially and temporally allocate Global Positioning System (GPS) instrumented vehicle activity data and provide summary statistics at a range of spatial and temporal resolutions. This software application is designed to operate seamlessly with existing GIS software owned by the Air Resources Board (ARB), represents a critical step in facilitating spatial allocation of mobile source emissions in California at a sub-county level. Availability of such a powerful and user-friendly software application enables the in-house collection and analysis of the large volumes of second-by-second GPS data necessary for development of a statistically robust motor vehicle activity data set. As such, the result of this project provides an important new tool for improving the spatial and temporal resolution of the motor vehicle emissions inventory, as well as assist in conformity analyses.

The developed GIS application was installed at the El Monte facility on August 16, 2000. The principal investigator demonstrated the functionality and operation of the developed package. An intensive training took place at the same facility on the same day.

MATERIALS AND METHODS

This project is divided into three tasks described in details below.

Task 1 – Evaluation of TransCore Activity Data Analysis Software

Task 1 of this project was to evaluate the activity data analysis capabilities of the GPS Travel Time Data Collection System software developed by TransCore. Three major issues were addressed in evaluating the software: (1) Software Suitability, (2) Street Matching Accuracy, and (3) System Requirements. The significance of these issues is discussed below.

Software Suitability

First, it is important to assess whether the TransCore software is suitable for ARB's analysis of motor vehicle activity. The software generates summary reports that show the route of a trip, the streets that are used, and the independent and cumulative travel time and average speed on each street segment (link). While such statistics provide useful and critical information for spatially and temporally allocating GPS instrumented vehicle activity data, to what extent the output meets ARB's requirements for deriving summary statistics at a range of spatial and temporal resolutions had to be examined.

Street Matching Accuracy

Second, the usefulness of the TransCore software depends on how accurate the GPS vehicle locations are matched to the street network. In principle, TransCore mathematically matches GPS vehicle locations to a street network through a series of algorithms. In general, such algorithms generalize a group of GPS-translated point locations into a line feature in order to identify the most likely street segment through which the vehicle had traveled. While there exist several possible sources of errors in the street segment approximation procedure, the software provides no direct mechanism for error checking. To determine if any identified street segment is accurate, one has to plot all the identified links in order for any abnormal link to show up. For instance, if the plot shows an isolated segment which is completely separated from other segments, it may indicate an error in the matching procedure. How accurate the matching algorithms are for the state of California had to be empirically tested using the actual GPS data. There are two ways to test the accuracy of the TransCore software in street matching. First, if both GPS data and reliable activity diary data are available, then the GPS data can be processed through the TransCore software and the result can be compared with the actual route according to the diary. Second, if the above data are not readily available, an alternative procedure is to use another existing street matching software package to generate a route and compare the results. In this task, the accuracy of the TransCore software was carefully examined in order to ensure that the derived summary statistics of vehicle activity data are accurate enough.

System Requirements

Third, the TransCore software will be applied to the street map of all fifty eight counties in California. Typically, digital street maps purchased from data providers are organized by county while vehicle trips are not confined to county boundaries. As such, restructuring street maps into more appropriate geographical organizations is necessary for effective analysis of vehicle travel data. Once street maps are adequately reorganized, researchers at ARB will be able to apply the software to generate meaningful summary reports and translate such reports into statistics suitable for analysis of mobile source emissions. How ready and easy it is to use the software must be carefully assessed. What other steps must be taken, and how much effort is required for each step, were all evaluated in this task.

In summary, in Task 1 efforts were made to identify the activity data parameters output, verify the map matching accuracy, and assess the suitability of the TransCore software in providing the input data required for fulfilling the remaining tasks.

TASK 2 – DEVELOPMENT OF ACTIVITY DATA ALLOCATION SOFTWARE

Task 2 of this project was to develop a customized GIS software application which shall provide gridded motor vehicle activity data within California at a user selectable range of temporal and spatial scales using activity data output by the GPS analysis software. This task involveed three major steps: (1) System Design, (2) System Configuration, (3) Programming and Testing.

System Design

The first consideration in designing the GIS application is to determine the programming language most appropriate for ARB. The application was to be developed using a programming language which shall allow seamless operation within either ArcInfo or ArcView GIS software packages. Criteria used in selection of the programming language include: (1) accuracy of the final activity data output; (2) data processing speed; (3) hardware requirements for the size of the database for analysis; (4) ease of future program modification and update; and (5) features and friendliness of the user interface. Based upon these criteria, a recommendation was made to use the ArcView GIS for the intended application and naturally the Avenue script language is the programming language

A second issue concerning system design is the capability of manipulating spatial and temporal resolutions in the application. The software provides motor vehicle activity data at five user-selectable spatial resolutions (5 km grid cell, county, air basin, district, statewide), as well as at temporal resolutions ranging from an average hour or average day to user specified hours or days. The configuration of the 5 km grid was determined with ARB's person in charge of the organization of spatial data in order that output of this software can be easily used in other models. Using a commonly accepted nomenclature, grid cells within the coverage were assigned unique identifiers to allow tabular reporting of results at the 5 km grid cell resolution. To carry out this task, ArcInfo compatible GIS coverages of California county and air basin boundaries were provided by ARB.

The software was designed to report the following activity statistics in both tabular and graphical (map) formats at the temporal and spatial resolutions specified above:

- Vehicle miles traveled (VMT) by speed by facility (roadway) type
- Number of starts
- Distribution of soak times (i.e. vehicle resting time)
- Distribution of idle times
- Distribution of cross-county and cross-air basin VMT
- Distribution of out-of-state VMT

System Configuration

In configuring the GIS application software, several issues were taken into consideration. First, the system must be complete and require no additional software package for ARB to conduct analyses of GPS-based vehicle travel data. In other word, as the required data are available, and the GPS data analysis software provides the desired output, ARB researchers should not be required to use other third party software for the analysis of the GPS data.

A second important concern is the efficiency in street map organization. As mentioned earlier, analysis of GPS-instrumented vehicle data requires GPS point locations be translated into street network information and thus digital street maps were purchased for this purpose. However, it is not only inefficient but also impractical to append the street maps of all the 58 counties together for this type of street-level travel analysis, because the file size will be extremely large and any spatial search will take a long time to complete. The problem is that, while digital street maps are typically organized by county, travelers are not confined by county borders, especially in metropolitan areas. How can the street data be most efficiently organized in order for ARB to easily conduct GPS data analysis was examined in depth. Utility tools providing functions for automation of street map processing are included in the application software. Such tools are designed for ARB researchers to effectively pre-process and reorganize digital street maps for analysis of vehicle data.

Third, street segments are not confined to 5 km grid cells, meaning that many street segments organized in the street map will cross grid cells. In some cases a segment will even cross more than two cells. Since street-matching algorithms are restricted to street segments (links), it is necessary for the street data to be pre-processed in such way that all the multiple-cell segments are dis-aggregated into shorter segments, each within a pre-defined 5 km grid cell. To do this, street maps were overlaid onto the 5 km grid and every segment crossing grid cells was divided into multiple segments with their topological (spatial) relationships reconstructed. Detailed requirements of this process were assessed in the system configuration. The required steps of operational procedures were incorporated in configuring technical specifications of the system.

Fourth, the organization of the system output and the presentation layouts are included in the system. How the output statistics are organized for easy use in the future, and how each table is to be presented in a most efficient way, were considered in the system configuration.

TASK 3 – INSTALLATION, DOCUMENTATION, TRAINING, AND SUPPORT

The primary deliverable for this project is a software application meeting the specifications described in Task 2. The software program was written in ArcView Avenue, and was provided to the ARB on CD-ROM's and designed to operate in a Microsoft Windows environment.

Documentation of this program includes a description of the software development, documented source code, and a users guide. In addition, a one-day training program was carried out at ARB facilities in El Monte on use of the software on August 15, 2000. Technical support will be provided to ARB at no additional cost for a period of six months. These deliverables will be considered the sole property of the ARB and shall be delivered to ARB staff upon completion of the project.

Programming and Testing

The GIS application was programmed in Avenue, the script language of the ArcView GIS. During the project period, the application program was extensively tested using empirical vehicle GPS data provided by ARB.

RESULTS

The results of this project are described below according to the three tasks specified in the preceding section.

Task 1 – Evaluation of TransCore Activity Data Analysis Software

Evaluation of Software Suitability

The first criterion in evaluating the TransCore software is whether or not the software is suitable for ARB's analysis of motor vehicle activity. The software generates summary reports that show the route of a trip, the streets that are used during the trip, and the independent and cumulative travel time and average speed on each street segment (link). While such statistics provide useful and critical information for spatially and temporally allocating GPS instrumented vehicle activity data, to what extent the output meets ARB's requirements for deriving summary statistics at a range of spatial and temporal resolutions was carefully examined.

In terms of the suitability of the software for ARB's intended analysis, there are several positive observations.

First, one major advantage of using the TransCore software is that it produces summary reports that can be translated into a format suitable for the intended analysis of vehicle data. The main utility of the software is to convert vehicle GPS point data into information associated with nodes and links in street networks. The software processes GPS point locations into such traffic-related statistics as travel distance, travel time, average speed, speed standard deviation, maximum speed, minimum speed, number of stops, and percent stopped. All the traffic-related statistics are organized by street segments (links) that can be related to the attributes associated with the network. For this project, the information available in the summary report can be aggregated spatially or temporally according to ARB's desired resolution.

Second, the program is free of charge to ARB, according to Mr. Laird of FHWA. The program consists of two main components, a vehicle-based data collection unit and a host-based data conversion unit. The data collection unit must be installed in every vehicle for GPS travel data collection while the host unit is needed at the ARB's site for analysis. Since there could be a large number of vehicle installations, plus multiple seats of ARB's systems for analysis, it would be a significant budget concern if ARB needs to pay for the license fees.

Third, installation of the data collection units is straightforward, according to the provided documentation. Although the software was not actually installed in vehicles, the operational procedure of such installation is identical to that of installing a typical commercial product of GPS data recording and should not be any complicated task.

On the other hand, there exist several potential problems concerning the use of the TransCore software for this project.

First, in terms of readiness of the software, at the time of evaluation the program was still under development and was not a matured product. The DOS version was not quite easy to install and test. It was developed using programming tools that are no longer compatible with the current mainstream programming tools such as C++ or other Windows-based programming language. The Windows version is basically the original DOS program plus a Windows interface, instead of a true Windows-based program. The implication for the current project is that it will be difficult to port modules of the program for major enhancement and it may be impossible to interface ArcView with the software directly, unless major modifications of the software could be made within a very short time period. In other words, significant changes in the programming structure would have to be required if we were to develop a seamless interface between the TransCore software and a GIS application using ArcView.

Second, related to the above issue was that one could not be sure about how well the TransCore software development team might accommodate ARB's needs regarding the required modifications. This concern is due to the fact that the program was supposed to have been well developed quite a while ago and yet it was still undergoing beta testing. The DOS version was difficult to evaluate as it was developed in 1995-1996 and that version apparently was not fully completed. To make the DOS version program work, one had to find an older 486 PC to install the software. The DOS version came with only one trip file without the corresponding street map, and thus evaluation of that version was not possible. Since the beta version of the Windows-based software became available in mid-June of 1999, the principal investigator had installed the program and tested it using all the included sample GPS trip data. The functionality and performance of the Windows version was essentially identical to those of the DOS version. Running the program was quite cumbersome at this stage and the interface still requireed major improvement before it could become a relatively matured product suitable for general use. Also, the current operation of the host system is not user-friendly for ARB's intended analysis. According to the current status of the software, it would require TransCore another 4-6 months to make the program ready for general application, even before any interface program with ArcView could be developed.

Third, in consideration of time requirement for the current project, it was necessary to take into account the additional time needed for installation and training of the mobile unit software. In other words, additional time would be needed to install the data collection component of the software in vehicles participating in the travel data analysis. Apparently, field data collection must be delayed for several months, or longer if unexpected problems do arise. If ARB planned to develop the GIS application within the next few months, there would have been not enough time to intensively test the application with a sufficiently large amount of vehicle trip data.

Fourth, some modifications of the TransCore software are necessary in order to incorporate the software into the GIS application. Minor modifications required for this purpose include (1) introducing an optional utility for one to run through the independent procedures of p-line connecting, path smoothing, street matching, and summary generating without the need to stop for operator's interaction at each step, (2) providing a batch job utility to process multiple trip data sets at once without the need of operator's manipulation for each set of trip data, (3) converting the output format of the current summary report "prn" file, the obsolete DOS format, into a common database format such as "dbf" or "mdb". Even though these modifications may not require major programming efforts, how soon this could have been done was questionable given the fact that the program was still at a beta testing version.

According to Mr. Laird of FHWA, minor modifications of the TransCore software could be requested through FHWA without additional charge. Nevertheless, given the current status and the progress made through the development of the program, the required modifications did not seem to be immediately feasible.

Fifth, in programming terms, the program should be re-compiled into either a COM object or a dynamically linked library (DLL) and thus can be executed within the environment of GIS function calls. To do so, the programmer will have to convert the current program into a more suitable programming language such as C++ code. This will require major modifications of the program and whether or not FHWA can arrange with TransCore to make such a major change is questionable.

In addition, a few major modifications of the TransCore software were required if ARB would consider to use the application. First, an editing mechanism must be developed for interactive correction of the TransCore software's street-matching errors. It is clear that no one single algorithm will always generate perfect street matching. TransCore software does not offer any means of correcting erroneously matched data points. As such, one would have to take whatever that is produced by the software and no mechanism can be used to adjust the matching result. Second, the TransCore software matches GPS points to street segments based on the characteristics of the line features. However, if a road segment is of a one-way street, then the matching could be totally wrong as the vehicle may be placed on the wrong side of the freeway. Third, the current implemented street-matching algorithms in TransCore are based on the graphical objects without considering attributes. As such, it lacks a self-correcting mechanism to relocate errors. For instance, if a GPS point is placed as one isolated segment totally disconnected from the rest of the network, there is no way to know where the matching went wrong. To enhance the software with the above required functions will require major modifications and does not seem to be feasible within the project period. Fourth, in order for this project to use the TransCore software, a direct ArcView interface is needed. Currently the TransCore software just generates ArcView shapefile output. What is needed is a direct interface so that one may process data and conduct analyses within the GIS itself.

The most critical concern about the TransCore software is that it would not be ready for this project because its DOS version never worked while its Windows conversion was still at a beta testing stage in 1999, even though its development started in 1995-96. It would require a minimum of six months, and more likely one year or longer, for the

TransCore software to become useful for this project. The Windows version evaluated contained many hard-coded common variables and parameters, such as file names and path directories, thus a ready-to-use product would take TransCore quite some time to accomplish.

Evaluation of Street Matching Accuracy

The usefulness of the TransCore software depends on how accurate GPS vehicle locations can be matched to the street network. TransCore matches GPS vehicle locations to a street network through a series of logical checks. In general, their algorithms generalize a group of GPS-translated point locations into a line feature in order to identify the most likely street segment through which the vehicle had traveled. While there exist several possible sources of errors in the street segment approximation procedure, the software provides no direct mechanism for error checking. To determine if any identified street segment is accurate, one has to plot all the identified links in order for any abnormal link to show up. For instance, if the plot shows an isolated segment that is completely separated from other segments, it may indicate an error in the matching procedure. How accurate the matching algorithms are for the state of California needs to be empirically tested using the actual GPS data.

There are two ways to vigorously test the accuracy of the TransCore software in street matching. First, if both GPS data and reliable activity diary data are available, then the GPS data can be processed through the TransCore software and the result can be compared with the actual route according to the diary. Second, if the above data are not readily available, an alternative procedure is to use another existing street matching software package to generate a route and compare the results, provided that both a TransCore data collection unit and the corresponding street map are available. Since neither of the two ways is possible, in this task, the accuracy of the TransCore software was examined through the analysis of the sample data. The key concern of the evaluation was to ensure that summary statistics of vehicle activity data are appropriately derived.

The accuracy of street matching depends on three main factors: (1) the accuracy of GPS data, (2) the accuracy of street data, and (3) the accuracy of the street-matching algorithm. Although in this project the key element for evaluation is the street-matching algorithm, the first two factors still need to be discussed as they are closely related to the evaluation of the street-matching algorithm.

Accuracy of GPS data

The accuracy of GPS data is affected by several factors, among them the introduced selective availability (SA) being the most severe. The theoretical position error for a high-quality GPS receiver is 35 m with SA and 20 m without SA, although in worst cases the spatial error could be as much as 100 m with SA. However, without SA, for commercially available and relatively low cost GPS receivers, the position accuracy is about 25 m with velocity accuracy about 0.1 m/s. Real-time differential GPS and post processing of GPS data with additional information obtained from base stations may significantly improve the accuracy at a higher cost. In general, those higher-quality, more expensive Differential GPS receivers may reduce the error to within 1 – 5 m and with a velocity error of 0.05 m/s. In reality, the position accuracy of GPS data tends not to be as good as what is specified by commercial products. As a rule of thumb, most people in the field of GPS application consider 30 m as the common average position error of less costly commercial GPS receivers.

In translating GPS position data into a street network, the obvious problem is that GPS points always jump back and forth around the actual position in a rather random manner. For surveyors to determine the precise location of a point feature, typically the GPS needs to be set at that location for a sufficiently long period of time in order to collect a sufficient number of data points and derive their geometric mean to represent the position. This is not applicable to moving vehicles since they are not supposed to stop at the same location for a long period of time. How to compensate the GPS data errors in moving vehicles has been a major area of research in the past few years. For our current project, the problem of GPS data accuracy cannot be completely separated from that of street map accuracy.

On May 1, 2000, the US government announced the removal the intentionally introduced SA for commercial GPS applications. The tremendous cost that was required to correct the GPS data due to SA is completely eliminated. It implies a major improvement for our project in terms of GPS data accuracy, such that, the costly differential correction is no longer needed and the much more affordable, commercial GPS receiver can now provide an average 10 m spatial accuracy suitable for this project. As a result of the release of SA, the vehicle travel data to be recorded by ARB become reliable.

Accuracy of Digital Street Maps

The accuracy of digital street map varies from product to product. In evaluating the use of a specific set of digital street map in terms of our current project, we considered both attribute accuracy and position accuracy. In addition, as there is always a budget constraint, one must also be aware of how complete a data set is and if its price is within the acceptable range.

The quality of digital street maps can be generally classified into three levels in terms of commonly available products: TIGER, ETAK, and the Thomas Brothers' Map (TBM). Other vendors are available and their quality is generally somewhere between these three classes.

TIGER is the least cost option and, as such, is also of the lowest level of accuracy. The most important advantage of TIGER over other products is its national coverage and extremely affordable cost. These two advantages, however, are not of significant concerns for the current project. First, the project is in California and thus there is no need to have a national coverage. Second, the extremely low cost is a plus yet the accuracy is a more important concern.

In terms of position accuracy, TIGER street lines are at best as good as the National Map Accuracy standards for rural counties derived from the map of 1:100,000 scale. In this case its position accuracy is about 50 m (167 ft). The accuracy level could be a misleading indicator because the actual accuracy level could be much worse due to the fact that there are quite a number of missing segments. If a segment does exist in the street map, then it is believed to be within 50 m from its actual position. If the segment is missing, then we will have a bigger problem relating the vehicle position to streets simply because the street segment does not exist in the map. A third area of main concern is the lack of appropriate attribute information, especially regarding freeways. In the TIGER file, freeways are coded as a single line feature representing the center line of the freeway. The single line feature cannot carry any one-way information and thus tracing the movement of a vehicle on freeways using GPS data could be difficult.

It is expected that the new revision of TIGER for the next census, TIGER 2000, will be much improved from its 1997 version, although how good those maps will be remains to be seen. A closely related product is GDT street map. GDT data are generally derived from TIGER and may be classified as TIGER-enhanced street map. In major urban counties such as Los Angeles and Orange, GDT data have been enhanced with significant improvement over TIGER. In less urbanized counties such as Riverside and San Bernardino, vast areas of roads are not much different from those in the TIGER file. The price of TIGER, processed for GIS applications by some data vendors, is about $2,000 for the state. The GDT data is $10,000 for the state.

The next level of digital street map is that of ETAK. ETAK has statewide coverage for California and thus is quite suitable for our project. The position accuracy of ETAK is significantly better than that of TIGER for urban counties. ETAK follows the National Map Accuracy Standards. The accuracy is about 12 m in urban areas where street segments are derived from the USGS's topographic series at 1:24,000 scale. For rural areas, however, the accuracy drops to the same level as TIGER, which is about 50 m derived from the 1:100,000 scale maps. In addition to the higher position accuracy in urban areas, ETAK maps provide the following important advantages. First, its coverage is significantly better than TIGER, i.e., in most counties in California, ETAK has significantly more street segments than TIGER-based maps. This property is important in providing much better street matching of vehicle GPS data. Second, ETAK maps have double line segments with one-way information for freeway segments. As such, translating GPS point data into street segments can be much more accurate because the direction can be traced, assuming that vehicles always follow one-way directions. Third, while TIGER has a well-established road classification system, in most areas streets in TIGER are roughly classified without using most of the classification items. ETAK's classification of roads is different from, and is generally better than, that of TIGER's. This property is important for our analysis of vehicle travel data in terms of different types of roads. These three properties make ETAK a better candidate for our project in linking GPS data to street maps. The standard price is $2,000 per high-density county such as Los Angeles and $1,000 per regular county.

The top-quality street maps are those of TBM. TBM has limited coverage in the nation, mostly only available for the west coast region including California. In urban areas with high street density, the position accuracy is much enhanced over the years through major efforts and TBM's contracts with local agencies. In areas where TBM did substantially enhancement project sponsored by local agencies, the position accuracy could be as good as within 5 m. Nevertheless, such high level of accuracy is not available everywhere in California. In rural counties the accuracy of TBM, although is generally superior to that of TIGER and ETAK, is still much less that the highly improved urban counties. The completeness of road coverage of TBM is far better than that of TIGER or ETAK, meaning that their street segments are more complete than any other vendor's data. The road classification is very well defined and implemented. In general, urban counties can enjoy a very high quality street maps from TBM if they can afford. Rural counties, however, are not as lucky as their urban counterpart because there are still numerous missing segments. The main problem for this project to use TBM is its prohibitively high cost and annual license fees. The pricing depends on the number of street segments. Los Angeles county is far more expensive than, say, Santa Barbara. The price can be anywhere between $10,000 and $20,000 for a county, and the user needs to pay the license fee annually.

The consideration of position accuracy needs to include the layout of streets. In urban areas, a typical city block is between 60 m and 150 m. In high density areas, adjacent street segments are about 60 m apart. In rural areas, road segments tend to be separated by 200 m or farther. In most cases, high density street segments are of the same type. For our project, if a GPS position is mismatched to an adjacent street, as long as they are of the same road type, the problem is not severe. The main problem is for areas where segments of different types are close together. For instance, if a residential street is about 20 m from a freeway segment, then the mismatching of a vehicle location will create a big difference. Nevertheless, if a position is mismatched, as long as it can be corrected using the next set of GPS data, then the effect of mismatching can be eliminated. Such self-correcting mechanism is not possible if street classification is not appropriate or if one-way information is not available. For this reason, TIGER-based street maps are not suitable for the project, unless they have been significantly enhanced.

Street matching should be easier in rural areas as long as the correct segments exist. For instance, if two adjacent segments are 200 m apart, and the position accuracy of the street map is 50 m plus the GPS error of 30 m, then the GPS point can still be correctly linked to the right segment. This does not apply to areas where roads of different classes intersect. For the self-correcting mechanism to take place, it is essential to have direction information and appropriate road classification.

Based on the above considerations, TIGER-based maps are not suitable for the current project while TBM maps are too expensive and of unnecessarily high quality. ETAK maps, or other digital street maps of the same level of quality and price, should be adopted for the current project.

Accuracy of Street Matching Algorithm

The majority of efforts made in this first task were to evaluate the accuracy of the street-matching algorithm of the TransCore software. The following discussion is based on the examination of eight sets of trip data provided with the beta version of the software. Among the included nine data sets of trips, Trip 1 cannot be evaluated due to the fact that it was used in the previous DOS version and its parameters were hard-coded. As a result, the evaluation is based on running the software for eight sets of trip data. The data come with ArcView shape files of street networks that cover part of their test area. Since all the sample data were provided by TransCore, we can assume that the results should represent those of the best possible quality. All the trip data and the routes derived from the program were plotted in ArcView for examination.

The following discussion is based on Trip 5 of the eight sample files provided by TransCore. The trip was made on July 13, 1999.

General Street Matching Methodology

Matching GPS point data to street networks can be achieved in one of two major approaches, matching points to nodes or matching points to links (segments). The TransCore software is built to match vehicle point locations to street segments. The methodology is described in their earlier documentation. According to the documentation, there appears to be no mathematical flaws in the matching algorithm. Nevertheless, the matching algorithm *per se* is not a mathematical problem, i.e., the iterative procedure is not a linear programming question where one my determine if the method is right or wrong mathematically in order to reach an optimal solution from the given set of objective function and constraints. The procedure, indeed, is defined by a series of logical checks where criteria were set arbitrarily. It is a matter of whether the procedure produces a result acceptable for our project.

The biggest challenge of the street-matching algorithm is how the different sources of data errors are handled in the procedure. As described above, data errors come from both GPS position data and street maps. How to compensate for the errors and obtain a suitable matching result can only be evaluated from experiments. The eight trips provide enough data for our evaluation of the street-matching algorithm.

Street matching based on the characteristics of line features will generate a good matching result if two critical conditions are satisfied. First, the GPS data must maintain the characteristics of the line features to be matched. Second, the street segments are accurately coded with enough between-node vertices. In other words, if the data are very clean, the line-feature matching algorithm will produce a good matching result. The reality, however, is far from ideal. Alternatively, street-matching can be based on nodes and vertices, where nodes are either street intersections, street ends, or intermediate location on long street segments, while vertices represent point locations that are used to describe the shape of the line segments. In general, matching points to nodes or vertices is a much faster process than the line-feature option. In general, if the GPS data are matched to nodes or vertices, and then processed through the attributes of street segments, then the data errors can be better handled and the processing will be more efficient.

In terms of methodology, which approach is more appropriate depends on both the quality of the GPS data and the accuracy of the street map. Even mathematical errors can be easily identified once the matching results are plotted. In this regard, we evaluated the street matching of the software based on the results of the eight trips examined.

GPS Points and P-lines

The sample GPS data of the eight sample trips were surprisingly good in terms of how well they delineate line features. GPS point data at this level of accuracy must be obtained either from a differential GPS or the data must have been post-processed using the data of a base station. Fig. 3 illustrates that the data do not jump and keep a constant direction, which indicates a relatively high accuracy level of GPS data, either from a differential GPS or a post-processed file at the time of recording. In either case, as the GPS points form very clearly defined line features, these data are suitable for the link-matching approach. According to the documentation, two GPS units used in the FHWA demonstration project were configured for differential correction, employing DCI's RDS 3000 FM receiver. These units are capable of producing the high-accuracy data points as shown in the sample data.

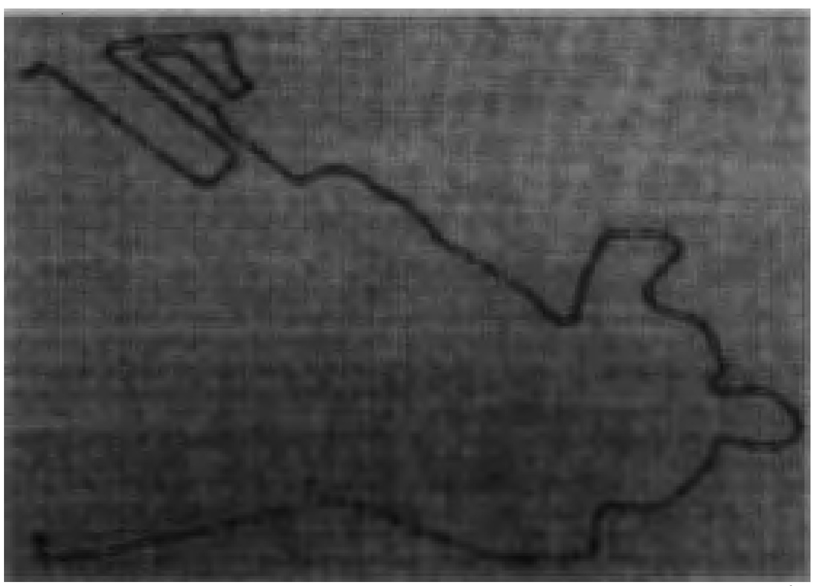


Fig. 3 GPS points of Trip 5, TransCore sample data.

The TransCore software allows one to generate P-lines from the GPS data by sequentially connecting adjacent data points to form line features with multiple vertices. The P-lines generated from the data are of very high quality since they represent the linear connections of the point locations. The thick line feature in Fig. 4 illustrates such a P-line.

At this stage, the P-lines have no relation to street segments. Since the GPS point data are of very high quality, the P-lines can be used to evaluate the quality of the street segments. Apparently the street map is not as accurate as the GPS data and thus the P-lines and street segments do not match. The differences between P-lines and street segments vary from place to place and they do not show any systematic pattern. Therefore, the street map is not systematically shifted due to the use of a different map projection or any other coordinate transformation. According to their documentation, TIGER files were used in the project and thus the main differences between P-lines and street segments tend to illustrate the low quality of TIGER street maps.

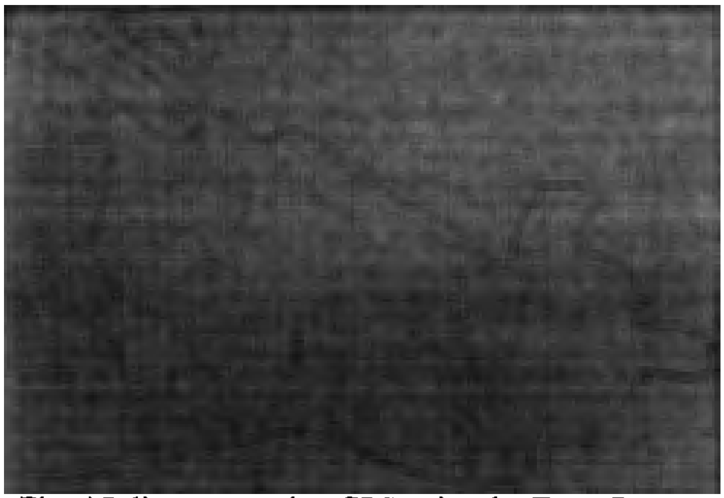


Fig. 4 P-line connecting GPS points by TransCore

Since the software provides neither error-checking function nor error-correcting mechanism, it is not possible to test the appropriateness of the link-matching method using the current included sample data. At the time of evaluation, two questions were considered. First, is it financially feasible for ARB to install differential GPS receivers to a large fleet of vehicles for this project, or to go through post-processing for the huge volumes of vehicle data in the long run. The consideration is not significant for installation of GPS for a small fleet over a short period of time. The funding requirement will be much higher if the GPS data collection is for a large fleet and for a long period of time. Second, are base stations available throughout the state of California for adequate post-processing of GPS data. In California, GPS base stations are available for most urban areas and coastal regions. Inland, rural counties may not be close enough to the coverage of base stations to generate sufficiently accurate GPS data. The quality of post-processed data is a function of distance between the GPS receiver and the base station. In general, if a location is too far away from an existing base station, the processed data may not be accurate. These considerations became unnecessary after May 2, 2000 when the US government released SA and the spatial accuracy of commercial GPS receivers was immediately improved from the previously up-to 100m level to the current 10 m level. As such, differential correction is no longer needed and the concern about GPS data accuracy is no longer necessary.

The main concern with our current project is that the street-matching accuracy may be biased against rural areas. As such, the statewide analysis of vehicle travel data may not be consistent throughout the state and the impact on the result needs to be further assessed.

Trip Smoothing

The trip-smoothing algorithm is an optional function of the TransCore software. The purpose of this function is to filter out point locations that do not match the characteristics of general line features. In common GIS terms, feature smoothing means converting line features of sharp angles into smother line features with additional vertices. The trip-smoothing function of the TransCore software appears to work in the opposite way. In this program, a smoothed line feature becomes one of sharp angle and the number of vertices is reduced instead of increased. Fig. 5 shows a smoothed line from a P-line.

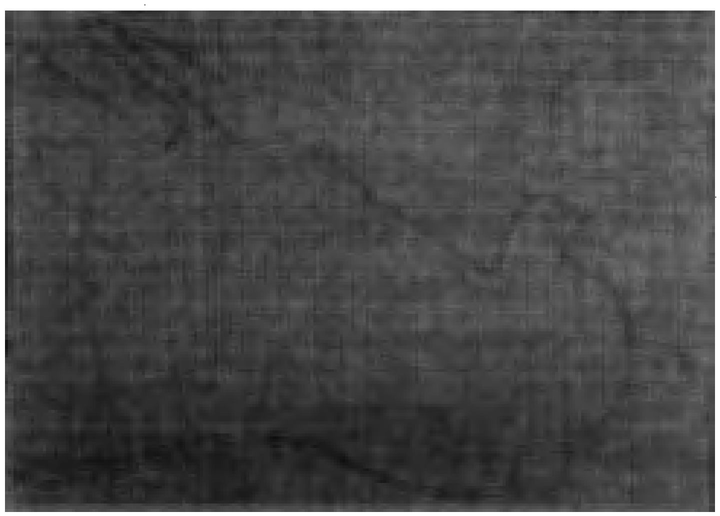


Fig. 5 Line feature smoothed from P-line by TransCore

A possible drawback of the smoothing function is that the procedure results in line features that are more different from their corresponding street segments. In reviewing the smoothed trips, one can tell the worsening match between a P-line and a street segment. Since this function is optional, we may simply skip this step and use the un-smoothed P-lines to match to street segments.

This function could be useful for another important reason, i.e., to define topology among line features. For instance, the GPS data form a series of point locations without differentiating them into multiple sections corresponding to street segments. The line smoothing could be very helpful if the series of points can be divided into multiple sections for more accurate street matching. In this regard, the function does not work appropriately since certain line features are converted in the opposite way. In many cases, a P-line resembles the curve of a street segment quite well and its smoothed shape becomes quite different from the street segment. In the program, line smoothing is an arbitrary series of logical checks. The conditions were hard-coded and cannot be altered by user. As we have no way to modify the source code and generate a better line-smoothing result, we could only evaluate the algorithm from the plots. The evaluation suggested that either the algorithm does not work, or there may exist bugs in the line-smoothing function.

Street Matching

The street matching function is to translate GPS points to the street network. There are numerous discernible errors in the matching results, as illustrated by the highlighted segments in Fig. 6. In the current version of the TransCore software, there must be some major problems with the street matching component. As mentioned earlier, the mathematics of the matching algorithm does not show any major flaw, and that the algorithm is actually a series of logical checks. Assuming that all the logical conditions are set appropriately, then there could be two possible error sources of the matching problem, coding errors or street map errors.

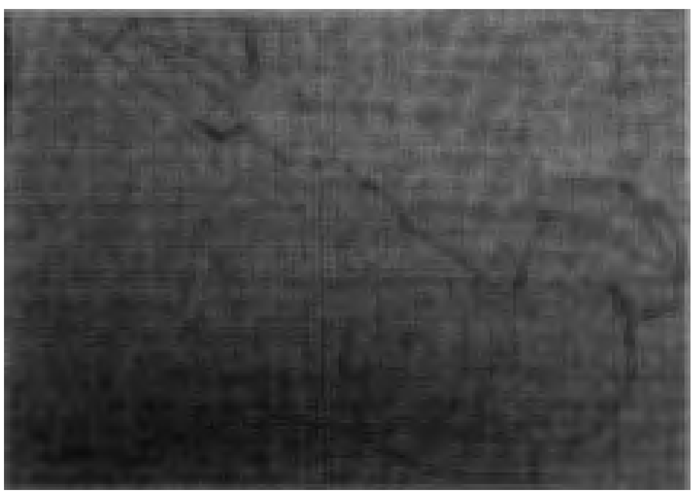


Fig. 6 Street segments matched from P-line by TransCore.

It is possible that the coding of the program may not be consistent with the mathematics described in the documentation, i.e., there could be serious coding errors that are not detected yet. Again, there was no way for the project team to evaluate the source code and thus where the data errors come from cannot be determined. TransCore should check the code carefully to determine if bugs exist to create the low quality matching results.

Assuming that there is no error with the logical conditions or coding, then apparently the data errors associated with the street map are the main source of error. This is because the GPS data seem to be of high quality and thus the only possible major error source is the street map. The street map included in the sample data is TIGER-based. To what extent a better quality street map may improve the matching result cannot be tested at this stage. The obvious problem is not the lack of street segment since in most cases the actual route can be identified from the street map. The major problems of the matching result can be summarized in two areas. First, wrong street segments are selected to form unreasonable routing pattern. Second, more segments are selected than necessary, which creates several loops along a route. This occurs most often near locations where major roads or freeways intersect.

Alternatively, a possible reason for the matching problem is related to the line-feature matching approach. The approach of matching GPS points to line features does not work well at this stage. If both GPS data and street maps are of high quality, the line-feature matching approach may work well. In the current situation, however, such high-quality data may not be available.

Summary Report

The summary function in the TransCore software is designed to clean up the result of street matching and generate link-based statistics for analysis. The accuracy of the summary relies heavily on that of the street matching function. Since the street matching of the sample trips is not quite acceptable, the summary is not suitable for use in the analysis.

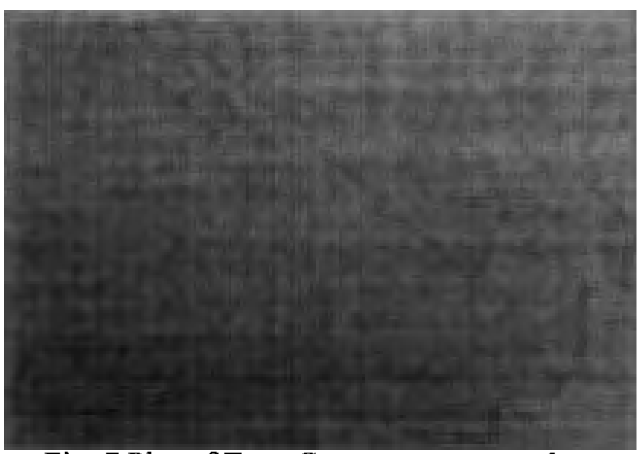


Fig. 7 Plot of TransCore summary result

The plot of TransCore's summary results, Fig. 7, shows unrealistic routes on many occasions. Links are not connected, i.e., several isolated segments are labeled as part of the trip. The generated statistics thus cannot be accurate. The summary report, Fig. 8, contains information very useful for our project, as long as the process is accurate enough. All the statistics generated from the software can be related to the street map through a common link code. An obvious drawback at this stage is that the summary report is generated as an obsolete DOS version dot matrix printer format "prn" which is not appropriate for Windows applications. Changing the output format to a "dbf" or "mdb" should be a relatively easy task for most software developers.

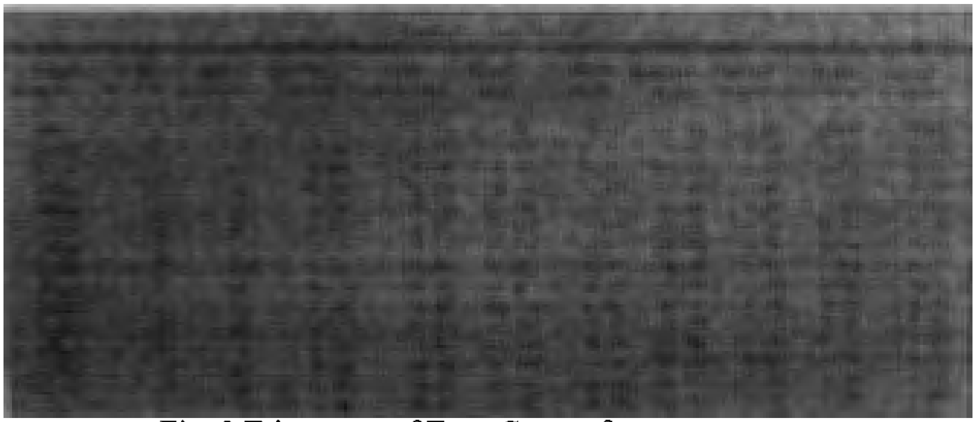


Fig. 8 Trip report of TransCore software

Evaluation of System Requirements

In this section, we evaluated the use of TransCore software in terms of system requirements, based on the assumption that the TransCore software could be applied to the street map of all fifty eight counties in California. Typically, digital street maps purchased from data providers are organized by county while vehicle trips are not confined to county boundaries. As such, restructuring street maps in some ways may be necessary to improve the effectiveness for analysis of vehicle travel data. Once street maps are adequately reorganized, researchers at ARB will be able to apply the software to generate meaningful summary reports and translate such reports into statistics suitable for analysis of mobile source emissions. How ready and easy it is to use the TransCore software was the focus of this section of evaluation. What other steps must be taken, and how much effort is required for each step, was also assessed in this task.

Street Data Organization

To use the TransCore software for this project, the digital street map of the entire state must be purchased and reorganized. The street data must be geographically organized into a database system with at least four levels of spatial resolution for analysis of vehicle travel data: statewide, county, air basin, and 5 km grid. Once the methodology is developed, the size of the grid can be altered for more detailed analysis. For instance, in certain areas it may be desirable to analyze the data at a 1 km grid.

The purchase of digital street maps is the first step toward building this GIS application. Both commercial street map data and the TIGER data are organized by county. Due to the extraordinary size of street segments in California, it is not wise to try to append all the street segments together to form a huge database. If all street segments are organized in the same data layer, it will take a very long time just to answer a simple query, let alone running a program trying to match numerous points with the tremendous number of road segments. The current structure of county-based geographic organization needs to be followed.

However, counties are of different size and different levels of street density. In certain cases it may be necessary to combine small counties in the same metropolitan areas if we can expect a large number of cross-county trips. The organization needs to consider the distribution of air basins as well. Therefore, once the street map data become available, the spatial relationships between counties, air basins, 5 km grid, and possibly grids of other resolution, need to be assessed and the data must be processed accordingly.

In processing the street map data, three steps are necessary for this application. First, an index map must be created to reflect the spatial relationships between different levels of geographic units. The main purpose of the index map is to identify the correct street file for each set of GPS vehicle data.

Second, since in the summary report of the TransCore software data are linked to street segments through a unique code "link", a suitable record ID of the street map needs to be established. Most likely, the street map data of all the fifty eight counties in California need to be processed and a statewide system of link ID must be implemented.

Third, boundary data of the state, counties, air basins, and the 5 km grid, must be established for analysis of vehicle data. Such data are either available through state agencies or ARB and can be converted into the GIS system for this project.

Installation and Training of Vehicle Data Collection Software

In order to use the TransCore software, it is required to install the data collection software in all the vehicles installed with GPS. Appendix B shows the procedures required to install the data collection software for each mobile unit. According to Mr. Doug Laird of FHWA, the software will be free of charge and ARB can install the data collection software in as many vehicles as ARB wants. The software can be installed in common laptop computers or some specific models of palmtop computers. The detailed specifications of the CPU requirements are available in the TransCore documentation.

The installation of the software also requires a brief training for the drivers to use the software. ARB may need to prepare some basic training material to help speed up the process. In most cases, one hour per driver (vehicle) is needed for software installation and setup, GPS testing, and a brief training.

Building the Master Vehicle Database

The GIS application will require the construction of a master database of vehicle travel data. The database must contain the vehicle GPS data obtained from all the vehicles. A simple appending mechanism to allow for new vehicle data to be appended to the master database. Also, a specific field identifying whether a trip is already included in the analysis must be introduced so that the same trip will not be processed more than once.

The master vehicle database must be logically linked to the index map of boundary files so that the processing of a specific trip can be done appropriately from the corresponding street data.

Modification of TransCore Software

According to Mr. Laird of FHWA, minor modifications of the TransCore software can be done without additional cost. The following areas of modifications will be needed for ARB to use the software for analysis of vehicle travel data.

First, the current version of the software requires substantial interaction for running the data of one trip. It is anticipated that there will be a large number of trips to be included in the analysis. To process one trip at a time, and each time going through the multiple steps of operation, will be extremely labor intensive and time consuming. Ideally, a simple text file can be set to run the program in a batch job. All the operator needs to do is to create a file containing the required parameters, such as the names and path of GPS data and the ArcView shape files of street maps. The batch processing will then allow for the operator to run the program and process multiple files of trip data overnight without the need of any interaction.

Second, the output report file created from the summary function is the DOS version prn file format with embedded codes for paging and line returns. Such an output format is not suitable for GIS applications. The software should be revised to generate summary reports in wither a text (ASCII) file format or a database. A database format will be desirable for seamless GIS applications.

In addition, a modification that requires more than minor efforts is that the software be converted into either a COM object or at least a DLL (dynamic link library) will help ARB and other potential users of the program to incorporate it into a GIS application. In this case, a seamless GIS application can be built which incorporates the TransCore software in a GIS environment such as ArcView GIS.

Project Administrative Meetings

Two follow-up meetings regarding Task 1 were held at ARB's El Monte facility after Task 1. In the first meeting on August 4, the Task One report was presented to ARB researchers, followed by a period of questions and answers to clarify related issues. The meeting went from 12:30 pm to about 3 pm, and was participated by both ARB researchers and Mr. Rick Everitt of UCR, who started to work as the graduate research assistant of this project in August, and the principal investigator. The following decisions were made during this meeting.

First, it was decided that the TransCore software is not to be adopted for this project due to the uncertainties and concerns addressed in the previous report. Without using the TransCore software, however, the StreetLinker software of GeoSpatial Technologies (GST) was used for this project.

Second, it was decided that the GIS application should be programmed on the ArcView GIS. Among other reasons, the main consideration as related to this decision is that ArcView has been enhanced to provide the necessary functions for this project and is much more user-friendly for ARB researchers that are not familiar with GIS operations.

Third, among the available street data, Etak street maps should be used for this project. The main consideration with respect to this decision is the overall better quality of the data than other alternatives.

Fourth, the system design was to be started and spatial data that are currently available at ARB, including the boundary files in ArcInfo format should be processed. Several ArcInfo coverages have been provided by Dr. Cicero of ARB in the form of exported files and converted into ArcView shape files.

The second meeting was held from 9 am to 11 am on August 25 at the El Monte facility. The main purpose of this meeting was to discuss the issues about field definitions and data structure related to the GIS application. ARB's ongoing efforts in the installation of GPS-based data loggers were also discussed. A couple of GPS units were presented at the meeting for general discussion of the GPS hardware, including a Trimble Cellular Messenger and a GPS unit built on BellSouth's Mobitex digital communication network.

The following issues were discussed and decisions made at the second meeting.

(1) In the statistical report, speed will be computed for each 5 mph interval.
(2) The number of starts will be calculated as the number of initial movements of a vehicle after the vehicle has stopped for at least 10 minutes.
(3) Soak times are measured for vehicles that have rested for more than 10 minutes during a trip.
(4) Idle times are calculated for vehicles that have rested for at least 2 minutes and no more than 10 minutes.

(5) Calculation of start times, soak times, and idle times will be based on aggregate statistics rather than for individual trips.
(6) Average grade will be calculated from the altitude values of the GPS data, if such data are available.
(7) Facility types will be translated from the ETAK street maps into five classes: freeway, arterial, collector, ramp, and other. Railroads and walkways will be eliminated from the street data to avoid unnecessary confusion.
(8) Vehicle GPS-based trip data will be organized in the GIS by links (street segments) rather than by point locations.

In addition to the above meetings, during this period the main task was to design the system of the GIS application.

Task 2 – Development of Activity Data Allocation Software

Database Requirements

The following theme layers in ArcView shape file format were constructed and maintained in the application.

Spatial Data

1. State Boundary: The state boundary coverage contains one major polygon that delineates the boundaries of the state of California. The database of state boundary in ArcInfo format is provided by ARB and was then converted into an ArcView theme layer. Additional polygons were added to the theme, including simplified outlines of Oregon, Nevada, Arizona, and Mexico. These additional polygons are used mainly for generating statistics about out-of-state trips. The spatial accuracy of boundaries of the neighboring states is not critical.
2. County Boundary: The database of county boundaries was provided by ARB in ArcInfo format and then converted into an ArcView theme. This theme layer contains polygon features of the 58 counties in California. This theme is used for generating statistics of motor vehicle activity by county.
3. Air Basin Boundary: Converted from an ArcInfo coverage, this theme layer contains polygons of the 14 air basins in the state. This theme is used for generating statistics of motor vehicle activity by air basin.
4. District Boundary: This theme layer contains polygons of air districts throughout the state. It is used for generating statistics of motor vehicle activity by district.
5. 5KM Grid: This theme layer contains grid cells of ARB's 5 KM grid to be used for generating statistics within grid cells. The grid is organized as a set of regularly spaced square features, each is of 5 KM length on each side.
6. Street Map: The EtakMap Premium digital street database was used for this project. The map data was pre-processed into a suitable format for the application. Pre-processing of the data includes reclassification of facility types and cleaning of unnecessary data elements. Regarding road type classification, the existing road classification of Etak was combined into five major categories: freeway, arterial, collector, ramp, and other. Cleaning of unnecessary data elements removed railroads and walkways from the database. Furthermore, a statewide unique ID was assigned to each street segment.

Vehicle Data

1. GPS database

   The GPS database contains all the text files obtained from the vehicles installed with GPS data loggers. The input data are in a text ASCII format and each file contains GPS point locations of a vehicle during a trip. The data contain at least the following items: vehicle ID, date, time, latitude, longitude, altitude, speed, and direction. Each file represents a trip and is referenced by a unique trip ID.

2. Vehicle Database

   A vehicle file will be provided by ARB. The file contains vehicle information including Project #, Vehicle ID, Year, Make, Model, Fuel Type, Class, and Tech Group. This file provides the necessary information about vehicles for analysis.

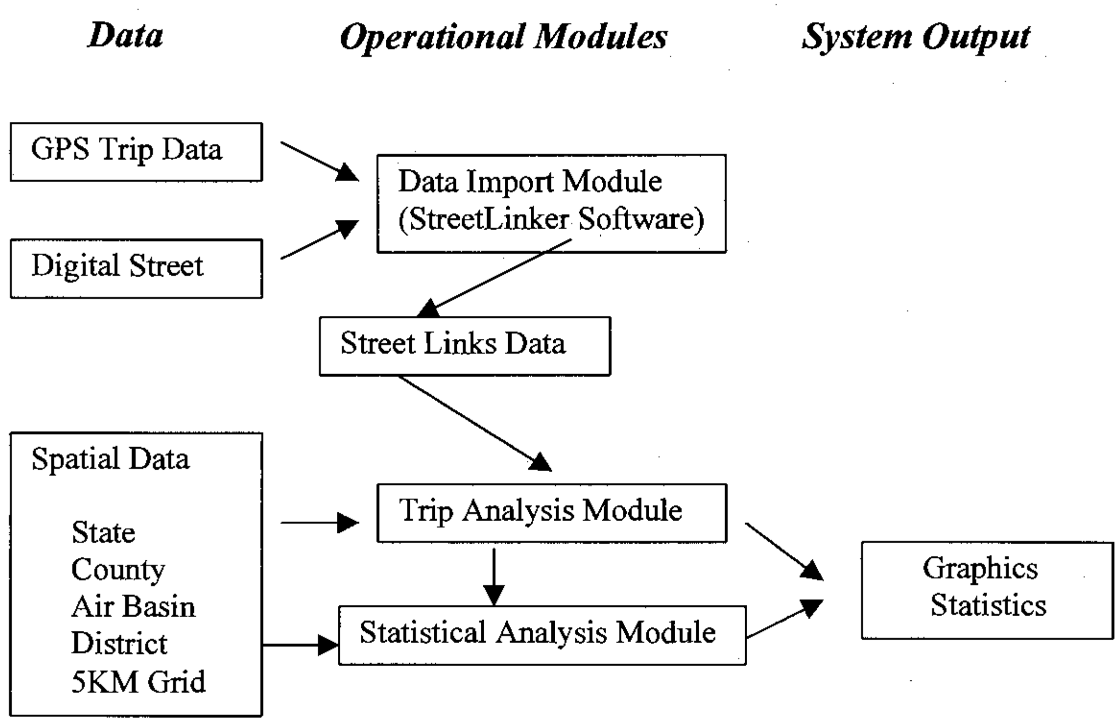


Fig. 9: The Conceptual Framework of the GIS Application

Conceptual Framework

The system to be developed through this project includes three inter-related operational modules: a GPS Data Import Module, a Trip Analysis Module, and a Statistical Analysis Module. The functionality of each module is described below. The conceptual framework is depicted in Fig. 9. The components are listed below.

GPS Data Import Module

The main function of this module is to read GPS-based vehicle activity files and translate them into information sets that are organized by elements of the street network (i.e., street segments or street links) for each trip. This module requires GPS data conversion functions available in the StreetLinker software. As such, when this module is activated through a function call in the application, it automatically executes appropriate algorithms of StreetLinker through a Dynamically Linked Library (DLL) in order to convert the input GPS data into information based on the corresponding street network.

In general, the program will read a raw GPS vehicle data file in text (ASCII) format. A typical file may have several hundred records similar to the following records in the Trimble TAIP format.

Rcvd: >REV000908169508+4433417-0927363806611002;ID=1008;*5A<
>REV000908169568+4432670-0927182206311512;ID=1008;*58<
>REV000908169628+4431650-0927023906614412;ID=1008;*5E<
Rcvd: >REV000908169688+4430461-0926896506016112;ID=1008;*5F<
>REV000908169748+4429070-0926834306115812;ID=1008;*5B<
>REV000908169808+4428056-
0926685706212912;ID=1008;*50<>REV000908169868+4427004-
0926543006214712;ID=1008;*58<
>REV000908169928+4425643-0926509905818102;ID=1008;*5F<
>REV000908169988+4424150-0926510406018002;ID=1008;*5E<

Once activated, the StreetLinker software will perform the following tasks.

First, the raw GPS code listed above must be translated into information suitable for analysis of vehicle activity in a GIS. In terms of position, each GPS record represents just one single point location expressed in Latitude and Longitude. This point location bears no reference to the street network and thus is not GIS-ready for analysis. Furthermore, GPS data always contain both spatial and attribute errors. Depending on the frequency of data receiving and reporting, there could be a substantial amount of redundant data elements in addition to erroneous points. The first StreetLinker function is to filter out unnecessary GPS data elements based on the digital street network. In other words, the internal logic of StreetLinker in determining which data element is erroneous or redundant is based on the corresponding street network. With this first function, each GPS point record will be linked to the street network. There tend to be multiple points assigned to the same node or vertex on a street segment. Therefore, a point filtering algorithm is necessary to reduce the data set and convert each group of points that share the same street location into one single record based on the property of the street segment. After this process, each remaining data point, called a link-based point, is represented as a point location on a street segment and it carries information about that segment, in addition to the information derived from GPS data.

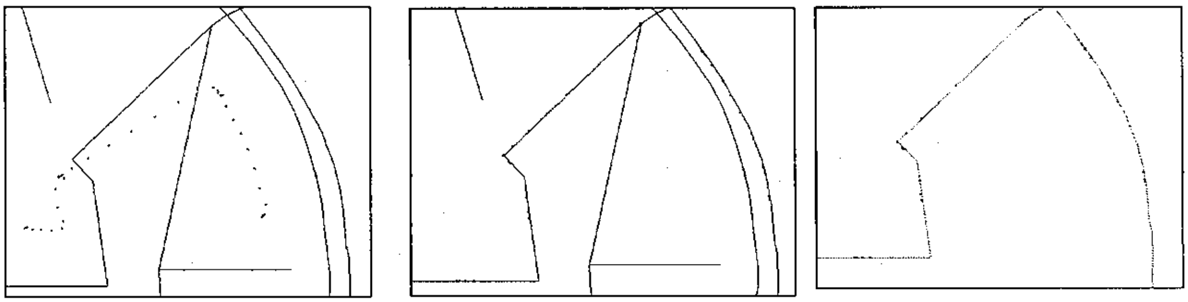


Fig. 10 StreetLinker links GPS to street segments

Fig. 10 illustrates the GPS to street matching in StreetLinker. The left diagram shows the GPS point locations plotted against the street map. It is clear that they do not match well. The middle diagram shows how each GPS point is recalculated in StreetLinker to find the best fit location on street segments. The right diagram shows the result of StreetLinker's street matching algorithm.

Second, the derived link-based points must be further translated into route path (route pattern) so that they may form linear pattern consistent with the street network. This process is extremely complicated since it involves conversion of a series of point data into another series of line features. In this step, StreetLinker introduces several steps of logical checks to make the data conversion as accurate as the data may support. In general, if both the GPS data and street data are of very high accuracy, the resulting link-based points will form a linear feature with an extremely high accurate route pattern. If the GPS points and street data are largely inaccurate, many points will fall into wrong links in many cases and the derived route pattern could be quite erroneous. The result is a series of links that form the route pattern of the trip.

Fig. 11 Route path by StreetLinker freeway linking

Fig. 11 shows an important feature of StreetLinker's linking mechanism. The data are empirical and provided by ARB. The trip goes north according to GPS data. However, GPS points fall to the west of the freeway. The TransCore software would match the points either to the local streets on the west of the GPS points, or match to the south bound lanes of the freeway. StreetLinker evaluates the pattern of the GPS points and correctly match them to the nouth bound lanes of the freeway.

Third, once the route pattern of a trip is formed, another function of attribute data organization will carry the information about the street segment to the street segments that form the route. Such segments are called street links in StreetLinker. The resulting street links provide the building blocks for both trip analysis and statistical analysis of the system.

StreetLinker has a built-in checking mechanism for evaluating the accuracy of the street linking process. The output of the data import module is a set of street links that can be displayed in ArcView for analysis of the derived trip pattern. The accuracy of the street linking algorithm can be evaluated by comparing the raw GPS points and the final street links. As the street links are saved as an ArcView shape file, it can be edited for more accurate trip analyses.

Trip Analysis Module

The trip analysis module is designed for analysis of one or more trips for detailed analysis of the trip pattern. However, due to the large size of the output file for each trip, it is not recommended to run the module for a large number of trips. In other words, this module is designed for close examination of a small number of trips.

Once activated, this module reads a trip file organized as street links and generates both graphic output and statistics for the selected trips. In this application, the user will be able to select a single trip or any number of trips for analysis. Street links derived from the previous module will be reorganized according to the specified spatial and temporal resolution.

The result of this module will still be link-based, i.e., each link will be a record in the output file. A header will provide vehicle information associated with the trip. In this module, specification of spatial and temporal resolution is for trip selection rather than for generation of statistics. The user may select trips that occurred in any county, 5KM grid cell, or air basin, and during any given time window. If the desired output is statistics over a spatial configuration or temporal specification, then the Statistical Analysis Module would be a more appropriate tool.

Statistical Analysis Module

The main purpose of this module is to generate aggregate statistics based on user-specified spatial and temporal resolution. In this case, all the trips that occurred within the selected spatial units and within the user-specified temporal resolution will be aggregated. The result will be the aggregate statistics of all trips and can be displayed as ArcView theme layers.

Once activated, this module reads the street-links file of each trip, re-organizes street links into sections by the user-specified spatial and temporal resolution, and then reads the next file for the next selected trip, until all the trip files have been read and re-organized. Statistics will then be generated according to the appropriate spatial resolution. As the aggregate statistics for each spatial unit are generated, a temporary theme table can be created to display the spatial pattern of the distribution according to the specified spatial resolution.

User Interface

The user interface was developed as an ArcView application. When the program is executed, it automatically activates ArcView and displays a front-end window for the user to specify spatial and temporal resolution of the analysis.

Spatial Resolution

The user will be asked to choose from the geographic units listed below for analysis. For instance, one may conduct an analysis by county and another analysis by air basin. Once the geographic unit is determined, the user may select one or more such units for analysis. For instance, the user may analyze trips that occurred in Los Angeles county only, or those that took place in both Los Angeles and Orange counties.

(1) County: One or more counties can be selected for analysis. Depending on the user's specification, the program will generate county-based statistics as the result of the analysis.
(2) 5 KM Grid: The user may specify one or more cells of the 5KM grid system for analysis.
(3) Air Basin: The user may specify one or more of the 14 air basins for analysis.
(4) District: The user may specify one or more districts for analysis.
(5) Statewide: If this option is selected, the analysis will be statewide and statistics will be generated for the state of California and the neighboring states and Mexico.

Temporal Resolution

A window listing possible temporal resolutions for analysis will be displayed. The default will be all unless otherwise specified.

(1) Date Range: The user will be able to specify the "from date" and "to date" range and only those records that fall into the specified range will be identified and analyzed.
(2) Hour Range: The user may specify a "from time" and "to time" range on an hourly basis. Only those records that satisfy the specified range will be analyzed.

In addition to the above specifications of temporal resolution, the users may specify the desired statistics either as hourly average, daily average, per hour total, or per day total.

Trip Selection

The analysis can be conducted for a single trip, multiple trips, or all trips available in the trip database. Each trip will be specified by either a unique trip identifier if such an identifier is available, or a combination of vehicle ID and date of trip. If the user is to analyze a single trip, then the trip must be specified and statistics will be generated for the selected trip only. If multiple trips are to be analyzed, then the user must choose two or more trips for the analysis and the statistics will be based on the sum total of all the selected trips. The default will be all trips in the trip database and the statistics will be based on the sum total of all.

System Output

The system generates both graphic output in ArcView shape file format and trip statistics in ArcView dbf format as a result of each analysis.

Graphic Output

In the application, the GPS trip data are translated into GIS information associated with the corresponding street network. The graphic output of the system will consist of the following ArcView shape files that can be displayed against the street network.

(1) GPS data points: This theme layer is the first output set of the StreetLinker software and contains the raw GPS data points. This set is the typical output for a single trip analysis where the user may select a trip and display the GPS point locations that have no relation with the street map. The accuracy of GPS data and street network can be examined from this set. In order to avoid unnecessarily large data storage requirement, normally this set will not be saved and thus each time the user may specify a trip and generate this theme layer directly from the raw data. The disadvantage is that this layer is not always ready for display and the user will need to run the program whenever such a layer is desired. The user may save the output in a separate directory if so desired.
(2) Link-based Points: This theme layer is the second output set of the system. This layer contains point locations that are translated from the original GPS data points into point locations on the street network. This set is part of StreetLinker output where GPS points are filtered and linked to street segments. In other words, this set no longer represents the original GPS locations. This set is available for the single trip

analysis. The user may choose to save the output as a separate point theme layer if so desired.

(3) Street Links: This set represents the third output of the system. This is the final result of the StreetLinker software where GPS point locations are translated into street links (segments). This theme layer is also designed for single trip analysis although the user may choose to save the output in a separate directory for future use, if so desired. The resulting Street Links will carry all the necessary information for trip analysis. For single trip analysis, the output can be used to generate the summary statistics of the trip. For multiple trip analysis, Street Links will be aggregated and the statistics will be based on the selected multiple trips.

(4) County Summary Map: Typically, this layer shows the summary statistics by county as a result of a multiple-trip statistical analysis. The map illustrates the spatial distribution of aggregate result of an analysis. All the summary statistics listed in the Summary Report will be available.

(5) Grid Summary Map: This theme layer is the same as the county summary map except that this layer is based on the 5 KM grid.

(6) Basin Summary Map: This theme layer is the same as the county summary map except that this layer is based on Air Basins.

(7) District Summary Map: This theme layer is the same as the county summary map except that this layer is based on districts.

(8) State Summary Map: This theme layer is the same as the county summary map except that this layer is based on the state boundary. This map is typically generated for analysis of out-of-state trips, including trips that crossed the international border with Mexico.

Statistical Output

Two types of statistical outputs will be available, one for single-trip analysis and the other for multiple-trip statistical analysis.

(1) Single-Trip Analysis: The header information of each single-trip summary will include Project #, Vehicle #, Year, Make, Model, Fuel Type, Class, and Tech Group. The data will be retrieved from a file of vehicle information provided by ARB. In the GPS data, a vehicle identifier will become part of the raw GPS output. The vehicle identifier will be used to retrieve the required information from the vehicle data file. If a matching vehicle is not found from the data file, then the header will report a "0" in the vehicle # field and the remaining fields will be empty.

The statistical output for a single-trip analysis will contain the following fields based on street segments (links):

- Link ID: The statewide unique ID of street segments (links).
- VMT (miles): travel distance in miles on the link.
- VHT (minutes): travel time in minutes on the link.
- Speed (mph): average speed in mph on the link.
- MaxSpd (mph): maximum speed in mph on the link.

- MinSpd (mph): minimum speed in mph on the link.
- Facility Type: classified into freeway, arterial, collector, ramp, or other.
- Average Grade: derived from GPS altitude data.
- Stops: the number of stops on the link for the trip.
- StopTime: Average time per stop.

(2) Aggregate Vehicle Activity Statistics: For multiple-trip statistical analysis, the following summary statistics related to vehicle activity will be generated at the specified spatial and temporal resolution. For each geographic unit at the specified temporal resolution, the aggregate statistics will include the following items:

- Trips: total number of trips that occurred within the selected geographic unit.
- VMT by speed by facility type: VMT at the interval of 5 mph per facility type.
- Starts: the number of starts defined as initial movements of a trip after at least 10 minutes of stop.
- Soak Times: the number of times and the total amount of time that a vehicle has rested for more than 10 minutes on the trip.
- Idle Times: the number of times and the total amount of time that a vehicle has rested for less than 10 minutes but longer than 2 minutes.

Vehicle Activity Stats (County) By Facility

Stats	Count	Sum_Vmt	Sum_Num_starts	Sum_Soak_time	Sum_Idle_time
Orange - Arterial	17	3.3837	0.0000	0.0000	0.0000
Orange - Collector	4	1.0994	0.0000	0.0000	0.0000
Orange - Freeway	135	55.3013	0.0000	0.0000	0.0000
Orange - Other	2	0.2054	0.0000	0.0000	0.0000
Riverside - Arterial	28	4.5199	0.0000	0.0000	0.0000
Riverside - Collector	21	5.5971	0.0000	0.0000	0.0000
Riverside - Freeway	86	38.1035	0.0000	0.0000	0.0000

Table 2. Statistics by county by facility type

System Requirements

Computer Hardware

1. CPU: minimum of Pentium II 400 MHz with 128 MB RAM (256 MB or more recommended)
2. Hard Drive: Minimum of 30 GB HD, partitioned or separated for spatial data and GPS data.
3. Printer: 11x17 color inkjet printer

Software

1. ESRI's ArcView 3.2 or higher. ARB's current ArcView licenses can be used for this project.
2. GSTI's StreetLinker 2.0. A copy of the StreetLinker software for linking GPS data to digital street maps will be included in the system. ARB does not need to purchase the license from GSTI and will be able to continue using StreetLinker with the GIS application developed through this project.
3. The ArcView application developed for this project. The application is written in the Avenue programming language and is compiled into an ArcView extension.

Operational Procedures

The basic operational procedures of the application developed for this project, called Motor Vehicle Activity GIS, are described in the following section.

MOTOR VEHICLE ACTIVITY GIS: OPERATIONAL PROCEDURES

To start Motor Vehicle Activity GIS program, click on the button 

located at the end of button bar below the menu. A dialog appears with the following options:

GPS Data Manager

- Import GPS
- Batch Import
- Delete Trip
- Edit GPS Link

Trip Analysis

- Single Trip
- Multiple Trips

Statistical Analysis

- County
- Basin
- District
- 5Km Grid
- State
- Facility

The functionality and operational procedure for each option are described below in the same order as these options are arranged in the pull-down menu.

GPS Data Manager – Import GPS

 Clicking on this button will also start the import procedure.

Select GPS Format: *NMEA, JHK* or *GST.*

Note: NMEA is the standard raw GPS code, JHK stands for the GPS receiver that ARB used in generating the sample files. GST is the standardized GPS data format of GeoSpatial Technologies, Inc.

Select GPS Filter: *Time*, *Speed*, and/or *Distance*.

1. Enter second(s) if time is checked.
2. Enter mph if speed is checked.
3. Enter meter(s) if distance is checked.

Note: The GPS Filter eliminates records that are either redundant or unnecessary. If the time parameter is specified, consecutive GPS records that are recorded within the time threshold will be eliminated. If Speed is specified, the GPS records that show a speed less than the specified threshold will be eliminated. If Distance is specified, the consecutive GPS records that occur within the specified distance will be eliminated.

Select a linking method from the following:

1. Single point to point – Find the closest vertex on the street network to the GPS point under consideration.
2. Single point to line – Find the closest street segment to the GPS point.
3. Double points to line – Find the closest street segment to each GPS point while the matched segment must in the same direction as that of the GPS point and the next GPS point. In this case, direction of the street segment is considered in the matching algorithm.
4. Double points to street – Find the closest street segment to the GPS point considering direction, connectivity, street name, and street oneway restriction if such data are available.

Enter the following for StreetLinker Settings:

1. Initial search distance – StreetLinker uses this initial search radius of GPS points to build a list of line segments for consideration. The larger the number is, the slower StreetLinker runs. On the other hand, a small number may rule out some possible candidates.
2. Maximum search distance – The maximum distance that the search algorithm will apply. This is to prevent unlimited searching when a GPS data point is corrupted and the search may take a very long time to run.
3. Maximum intersection distance – This is to define the maximum radius that a GPS point will be considered at a street intersection. Any GPS point within

the distance of the end point of a selected segment will display street as intersection such as: Main St ; Slater Ave.

4. Maximum angle – This parameter is used in the linking method with directional consideration, such as double points to line or street. This parameter specifies how wide an angle will be considered acceptable in street linking. A angle of 180 degrees will cause no effect of this function since it covers the entire possible range of line angles.

Enter Input GPS File. This file is the GPS file.
Enter GPSpnt File. This file is reformatted and filtered from the GPS file.
Enter LnkPnt File. This file is the final output linking streets to GPS points.
Enter Trip ID. This should be unique.
Select a County where the GPS file is geographically located.
Check on *Beep when done* to beep when StreetLinker finished linking all the GPS points.
Press *Apply* to run or *Cancel* to exit.

Once you press *Apply*, it creates another process outside of ArcView to process the GPS points. It's minimized but you can see it in the task bar on the bottom of your screen call ed StreetLinker. Once the processing is complete, it instructs ArcView that it's done. ArcView then creates three shape files based on the trip id (ex: TripLA001) you entered:
TripLA001_pnt.shp – orginal GPS points
TripLA001_lnk.shp – closest points on segments from the orginal GPS points
TripLA001_str.shp – closest streets from the original GPS points.

Reset StreetLinker. If the process outside of ArcView abruptly ends and you press the icon above to run StreetLinker, you will get a message saying, "Street Linker is running." Use this icon and you should be about to run StreetLinker again.

Note: Make sure that program outside of ArcView is not running. You can select StreetLinker on the task bar. If it doesn't open, it means that it's still processing so please wait.


Draw lines connecting GPS points to closest points StreetLinker found.


Remove lines drawn.


Remove all themes whose name contains _pnt, _lnk, and _str.

GPS Data Manager – Batch Import

Add – Add an entry to the batch import. Follow the procedure described in *Import GPS.* To add more entries, change the parameters and press *Apply.* Press *Cancel* when done.

Edit – Edit the selected entry. Follow the procedure described in *Import GPS.*

Del Sel – Delete selected entries.

Del All – Delete all entries.

Read File – Restore entries from a file.

Write File – Save entries into a file.

Apply – Start batch import.

Cancel – Cancel batch import dialog.

GPS Data Manager – Delete Trip

Select one or more trips. To select all trips, check on *All* checkbox and press button >>>. Press Apply to delete trip(s).

Note: Trips are deleted from trip.dbf and tripmaster.dbf. <trip>_pnt.*, <trip>_lnk.*, and <trip>_str.* are not deleted.

GPS Data Manager – Edit GPS Link

Select a trip from the pop-up list to start edit. The selected trip and related themes will be added to View1. Use the following tools to edit: 

 Click on this tool and theme <trip>_lnk.shp will be active. Select GPS point(s). Use the shift key to add more points.

 Click on this tool and theme <trip>_str.shp will be active. Select a street.

 Click on this tool to assign select GPS point(s) to the selected street.

 Undo.

 Redo.

 Save current edits.

 Save current edits and apply changes to trip.dbf and tripmaster.dbf

Trip Analysis – Single Trip

Select a trip to add theme <trip>_str.shp to view *Single-Trip View.*

Check *Display Map* to add county street map for this trip to view *Single-Trip View.*

Check *Display Table* to open statistics for this trip. This is extracted from tripmaster.dbf.

Trip Analysis – Multiple Trips

Add themes the following criteria to view *Multi-Trips View.*
Select a county. Check on *All* checkbox and press button >>> to select all counties.

Select a basin. Check on *All* checkbox and press button >>> to select all basins.

Select a district. Check on *All* checkbox and press button >>> to select all districts.

Enter a grid id and press button >>> to the right. To enter a block of grid id, enter from row and column, and to row and column and press button >>> to the right.

Enter date from and to.

Enter time from and to.

Select a state. Check on *All* checkbox and press button >>> to select all states.

Select a trip. Check on *All* checkbox and press button >>> to select all trips.

Check *Display Map* to add county street map for this trip to view *Single-Trip View.*

Check *Display Table* to open statistics for this trip. This is extracted from tripmaster.dbf.

Press Apply to search.

Note: If more than one criterion is used, such as counties and 5km grid, then all trips in specified counties and within specified grid will be selected. For example:

1. Counties selected are *Los Angeles* and *Orange*.
2. Grids selected are 1001 and 1001.
3. Selection is made base on the following criteria:

((County='Los Angels') or (County='Orange')) and ((Grid=1001) or (Grid=1002))

Statistical Analysis

Select county, basin, district, 5km grid, or state as the geographic base region.

Press button *Run Statistic.*

A dialog opens. Follow the instruction under the section: *Trip Analysis – Multiple Trips.*

Checking no Facility creates another table that divides the base geographic region further by facility: Major Rd, Freeway, Minor Rd, and Other.

Result will display in view *Statistical Analysis.*

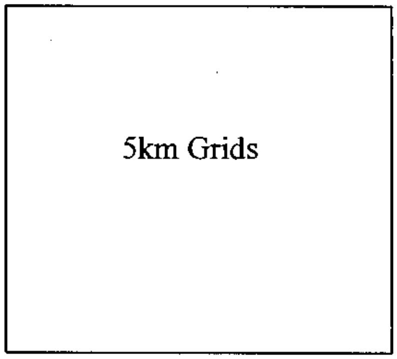
Click on view *Statistical Analysis*. Select the following under main menu *Map Display* to view different statistic:

- VMT – Total vehicle miles traveled
- VHT – Total time traveled (minutes)
- Speed – Average speed
- Soak Time – Total soak time
- Idle Time – Total idle time
- Stops – Total stops
- Starts – Total starts
- Grade Up – Average grade up (if data available)
- Grade Down – Average grade down (if data available)

5Km Grids – Five kilometer grids

Attribute	*Value*
Icell (column)	1 - 200
Jcell (row)	1 - 220
Km5_id	Icell*1000+Jcell

Icell=1,Jcell=220 Icell=200,Jcell=220
Km5_id=1220 Km5_id=200220



Icell=1,Jcell=1 Icell=200,Jcell=1
Km5_id=1001 Km5_id=200001

Trip.dbf

Name	Field Type	Width	Precision	Description
Trip_id	CHAR	30	0	Trip ID (unique)
Datefr	CHAR	8	0	Date from (JHK data has no date)
Dateto	CHAR	8	0	Date to (JHK data has no date)
Timefr	CHAR	5	0	Time from
Timeto	CHAR	5	0	Time to
Leninst	DECIMAL	12	2	Trip length inside California in miles
Lenoutst	DECIMAL	12	2	Trip length outside California in miles
Source	CHAR	60	0	Path to trip ex: d:\strlnkr\tripdata\org_str.shp
Fips	CHAR	10	0	Future usage

Tripmaster.dbf

Name	Field Type	Width	Precision	Description
Tripid	CHAR	30	0	Trip id
Facility	CHAR	10	0	Convert from field class in street shapefile: A, C – Major Rd F – Freeway O, V, W – Minor Rd
County	CHAR	16	0	County
Basin	CHAR	16	0	Basin
District	CHAR	16	0	District
Km5_id	DECIMAL	16	0	Five kilometer grid id
State	CHAR	16	0	State
Vmt	DECIMAL	10	6	Travel distance in miles on the link
Vht	DECIMAL	9	2	Travel distance in minutes on the link
Arc_id	DECIMAL	16	0	Arc ID (unique for each county)
Speed	DECIMAL	5	2	Average speed in mph on the link
Max_speed	DECIMAL	5	2	Maximum speed in mph on the link
Min_speed	DECIMAL	5	2	Minimum speed in mph on the link
Num_stops	DECIMAL	4	0	Number of stops on the link for the trip
Soak_time	DECIMAL	10	0	Total amount of time that a vehicle has rested for more than 10 minutes on the trip
Idle_time	DECIMAL	10	0	Total amount of time that a vehicle has rested for less than 10 minutes but longer than 2 minutes
Avegrade	DECIMAL	6	2	Not used
Gradeup	DECIMAL	6	2	Average up grade (for NMEA data only)
Gradedown	DECIMAL	6	2	Average down grade (for NMEA data only)
Numpnt	DECIMAL	16	0	Number of points on the link
Num_starts	DECIMAL	4	0	Number of starts defnied as initial movements of a trip after at least 10 minutes of stop
Length	DECIMAL	16	5	Length of the link
Fips	CHAR	20	0	FIPS of the link

TripID_pnt.shp – GPS point

Name	Field Type	Width	Precision	Description
Shape	SHAPEPOINT	6	0	Standard ArcView shape field
Vehid	CHAR	30	0	Vehicle ID (UNKN for NMEA data)
Date	CHAR	11	0	Date (JHK data has no date)
Time	CHAR	9	0	Time
X	DECIMAL	12	6	X-coordinate of GPS point
Y	DECIMAL	12	6	Y-coordinate of GPS point
Alt	DECIMAL	7	2	Altitude (for NMEA data only)
Speed	DECIMAL	7	2	Speed in mph
Numstop	DECIMAL	4	0	Number of stops on the link for the trip
Numstart	DECIMAL	4	0	Number of starts defnied as initial movements of a trip after at least 10 minutes of stop
Idletime	DECIMAL	11	0	Total amount of time that a vehicle has rested for less than 10 minutes but longer than 2 minutes
Soaktime	DECIMAL	11	0	Total amount of time that a vehicle has rested for more than 10 minutes on the trip
Dir	DECIMAL	7	2	Direction in degrees (JHK data has no direction)
County	CHAR	30	0	County
Path	CHAR	60	0	Path to GST mdb file, ex: D:\Strlnkr\street\c06059\c06059_str.mdb
Arcid	DECIMAL	11	0	Not used
Closestx	DECIMAL	12	6	Not used
Closesty	DECIMAL	12	6	Not used
Street	CHAR	60	0	Not used

TripID_lnk.shp – Closest point on a street to a GPS point

Name	Field Type	Width	Precision	Description
Shape	SHAPEPOINT	6	0	Standard ArcView shape field
Vehid	CHAR	30	0	Vehicle ID (UNKN for NMEA data)
Date	CHAR	11	0	Date (JHK data has no date)
Time	CHAR	9	0	Time
X	DECIMAL	12	6	X-coordinate of GPS point
Y	DECIMAL	12	6	Y-coordinate of GPS point
Alt	DECIMAL	7	2	Altitude (for NMEA data only)
Speed	DECIMAL	7	2	Speed in mph
Numstop	DECIMAL	4	0	Number of stops on the link for the trip
Numstart	DECIMAL	4	0	Number of starts defnied as initial movements of a trip after at least 10 minutes of stop
Idletime	DECIMAL	11	0	Total amount of time that a vehicle has rested for less than 10 minutes but longer than 2 minutes
Soaktime	DECIMAL	11	0	Total amount of time that a vehicle has rested for more than 10 minutes on the trip
Dir	DECIMAL	7	2	Direction in degrees (JHK data has no direction)
County	CHAR	30	0	County
Path	CHAR	60	0	Path to GST mdb file, ex: D:\Strlnkr\street\c06059\c06059_str.mdb
Arcid	DECIMAL	11	0	ArcID of closest street to GPS point
Closestx	DECIMAL	12	6	Closest X-coordindate on the street segment
Closesty	DECIMAL	12	6	Closest Y-coordindate on the street segment
Street	CHAR	60	0	Street name of closest street to GPS point
Vmt	DECIMAL	10	6	Travel distance in miles on the link
Vht	DECIMAL	9	2	Travel distance in minutes on the link
Altup	DECIMAL	7	2	Grade up (for NMEA data only)
Altdown	DECIMAL	7	2	Grade down (for NMEA data only)

TASK 3 – INSTALLATION, DOCUMENTATION, TRAINING, AND SUPPORT

The GIS application is scheduled to be installed at ARB's El Monte facility on August 15, 2000. The User's Guide and Programmer's Reference Manual will be submitted to ARB upon the installation and training of the system. A one-day training will be held either at the El Monte facility on the same day, or at a different location to be decided later. Technical support will be provided to ARB for designated researchers for a period of six months after completion of this project. The technical support will include phone and e-mail support, fixing of any bugs identified by ARB researchers, and free system upgrade during the six-month period.

DISCUSSION

Through the cooperation between ARB researchers and the investigator, this project has successfully developed a powerful tool for intensive analysis of motor vehicle travel data. One of the previous concerns is about the accuracy of the GPS vehicle location data. During the course of the development, on May 1, 2000, the US government announced the removal of Selective Availability (SA), which was introduced by the Department of Defense on purpose. The removal of SA makes the project even more useful as now the vehicle data become much more accurate and thus the result of analysis will be much more reliable. The current GPS accuracy is within 10 meters, a level that is quite adequate for analysis of motor vehicle data since roads tend not to be within 10 meters from each other.

Data errors are inevitable, especially for large databases. In this project, a tremendous amount of spatial data is processed and it is not unexpected to contain data errors in the current spatial database. Any researcher using the developed system should be aware of the possible data errors. This project employs the latest version of Etak street maps, considered of relatively high quality and up to date, yet problems such as new streets not included, existing streets geocoded with spatial errors, and roads mis-classified, are not uncommon even for the top quality digital street files. Nevertheless, the overall accuracy of the street information incorporated in this project should be quite acceptable.

The GIS application developed in this project is designed to be user-friendly and requires no major GIS background for ARB researchers. Since the spatial database is huge in size, the system requires careful organization of the data to ensure appropriate processing and manipulation in the algorithms. Currently the spatial data are organized by county, thus there are 58 sub-directories in the structure of the spatial database. It will be beneficial for ARB researchers that will be using this analytical tool to be familiar with how the data are organized. The spatial data organization could be the foundation for ARB to further expand this system for other types of related modeling and analysis.

SUMMARY AND CONCLUSIONS

This project involves development of a customized Geographic Information System (GIS) software application to spatially and temporally allocate Global Positioning System (GPS) instrumented vehicle activity data and provide summary statistics at a range of spatial and temporal resolutions. The ultimate object is to further enhance emission inventory models based on empirical, accurate vehicle travel data in order to assist in the development of air pollution control strategies and regulations in California.

Through this project the research team thoroughly evaluated an existing software package developed by TransCore for translating GPS points into trip summary data. The package was not quite ready for this project and thus another similar package called StreetLinker is adopted for this project. Assessment of the available GIS tools and digital street maps led to the decision to build the system on ESRI's ArcView GIS, using Etak's latest digital street maps.

The developed system allows for ARB researchers to (1) convert vehicle GPS data into a database for trip and statistical analyses, (2) generate multiple map layers including original GPS point locations, point locations interpreted by street network, and street segments that form each trip, (3) conduct trip analysis to show basic statistics associated with each trip and any set of selected trips, and (4) conduct statistical analysis to obtain statistics of vehicle travel data both spatially and temporarily.

RECOMMENDATIONS

In this project a GIS application is built for ARB researchers to analyze the GPS-instrumented vehicle travel data throughout California. The system provides a powerful analytical tool that enables vehicle travel patterns to be extensively analyzed. At this stage there are very few empirical data collected for this kind of research. The next step is certainly to build the database for meaningful statistical analyses of vehicle travel patterns.

Currently, vehicle GPS data are collected through a hardware mobile unit. As now the analytical tool becomes available, collection of data is most urgent and meanwhile most difficult. It will require a researcher to take the mobile unit box to a vehicle and install it for the next day, and retrieve the box the next evening. The operator will then download the vehicle GPS history file (or log file) into a computer before taking the box to the next vehicle.

One possible way to make the procedure more effective is to use a mobile unit with a radio modem installed. The radio modem must be affordable and with as complete coverage within the state of California as possible. A possible solution is to use Mobitex of BellSouth which has good coverage throughout Califonria except for rural areas. Typically the real-time transmission of the vehicle GPS data is not necessary for ARB's project. Therefore, communication between the vehicles being tracked and the ARB host station does not need to take place in the day time when wireless communication is most expensive. This solution has important advantages. First, the vehicle data can be collected for multiple days, such as one week, so that the travel data can be more correctly representing the actual travel pattern. Second, ARB researchers do not need to download the data on a daily basis, in fact, there is no need to download the vehicle data at all since the data can be automatically transmitted to the centralized facility such as at El Monte. The radio modem will automatically transmit the vehicle travel data every night when the price of air time is minimum. Third, since the researchers do not need to physically download the data, researchers at El Monte can collect vehicle data from many vehicles throughout the state without the need of physically go the vehicle. The use of a radio modem in the mobile unit can make data collection much easier and more effective.

In addition to data collection, currently the researcher has to download the data, bring the data to the workstation, and then copy the data to the hard disk, and then run the system developed for this project to get the statistics. The operation can be made much more efficient by implementing an automatic data loading utility so that all the vehicle travel data can be automatically loaded into the database through radio modems. The saving on the trips made by ARB's researchers should exceed the cost of building an automated data loading system.

REFERENCES

Chou, Y. H. 1998. Exploring Spatial Analysis in Geographic Information Systems, Onword Press.

Chou, Y.H. and C.B. Baker. 1999. Tracking Buses with GPS, Business Geographics, July 1999, pp. 29.

Chou, Y.H., E.I. Rudd, and J. Pennington. 1998. Emergency 911: Integrated GPS/GIS to the Rescue. GIS World, August 1998. Pp. 48 – 52.

Logsdon, T. 1992. The Navstar Global Positioning System, Van Nonstrand Reinhold.