Technical Report
Research Project T9903, Task 62
FLOW Evaluation

FLOW EVALUATION DESIGN TECHNICAL REPORT

by

John M. Ishimaru
Senior Research Engineer

Mark E. Hallenbeck
Director

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
University District Building
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation
Technical Monitor
Leslie Jacobson
Assistant Regional Administrator for Traffic Systems

Prepared for

Washington State Transportation Commission
Department of Transportation
and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

March 1999
(revised May 1999)

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Transportation Commission, Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Table of Contents

Section

List of Figures and Tables

Figures

Tables

Purpose of This Report

This report describes an evaluation approach, process, and analytical tool set that was developed for the analysis of freeway usage and performance in the central Puget Sound region. It also functions as a user's guide, describing the range of performance measures and the methods and tools available to estimate them, along with step-by-step instructions. This report also serves the important purpose of documenting the analytical assumptions and limitations of the evaluation method; these considerations should be carefully taken into account by prospective users.

This report is one of three products of a Washington State Department of Transportation (WSDOT) project to enhance the department's ability to monitor and improve its traffic management efforts on Seattle-area highways, and to provide useful information to the public and other decision makers about the status of the freeway system and state traffic management activities in the area. In addition to this report, this project has produced a set of software tools to assist in freeway data analysis, as well as an interim report documenting the level of traveler usage and travel performance on the principal urban freeways in the central Puget Sound area for 1997. The freeways studied in this project are managed by the Washington State Department of Transportation using its FLOW system, a coordinated network of traffic monitoring, measuring, information dissemination, and control devices that operates on urban state and interstate highways in the central Puget Sound region.

The data analysis procedures described in this report are intended to facilitate a series of periodic evaluations of the central Puget Sound urban highway network and the WSDOT FLOW system. They were also designed to be general capabilities, so that they can be employed at any freeway location of interest, provided that the appropriate data have been collected. This evaluation process focuses on mainline (GP and HOV) performance measures; plans call for eventual expansion of the process to include other aspects of the WSDOT FLOW system, such as ramp meters, safety aspects, and traveler information systems.

What Is In This Report

This report summarizes the methodology, tools, and procedures developed for the evaluation of the Seattle area freeway network. The report is divided into five sections:

1. **Introduction.** The objectives of this report are described.

2. **Evaluation Methodology**. The overall evaluation process is described. Data collection, sampling, processing, and display issues are summarized.

3. **User's Guide: FLOW Evaluation Analysis Procedures and Software Tool Set**. The process by which freeway data is collected and analyzed is described in more detail. The capabilities of the software tools CDR (Compact disc Data Retrieval), CDR Analyst, and CDR Auto are described, including a catalog of performance evaluation options, illustrations of types of output, and ancillary macros and template files. Performance measures and the software tools used to compute them are presented, including step-by-step instructions for their use as well as specific data requirements. CDR Analyst algorithms are also described.

4. **Using the FLOW Evaluation Analysis Process**. Three examples of the use of the FLOW evaluation process are described:

   Example I: Producing the Central Puget Sound Freeway Usage and Performance Report. The process used to produce the "Central Puget Sound Freeway Usage and Performance Report" is summarized.

   Example II: Producing Supporting Data for the WSDOT Ramp and Roadway Traffic Volume Report. Guidelines for the use of this evaluation method to produce data that can assist in producing the "Ramp and Roadway Traffic Volumes Report" for WSDOT Northwest Region are described. (The ramp and roadway report summarizes average peak hour and daily weekday volumes for locations throughout the regional freeway network.)

   Example III: System Operations Diagnostics. Examples of the application of FLOW evaluation analysis tools to study system operations issues, diagnose data collection equipment status, and interpret the validity of summary statistics are described.

5. **Analytical Assumptions and Limitations**. The assumptions upon which performance measures, software, and analysis methods are based are described. Limitations and caveats associated with the software and algorithms are also summarized.

Appendix: CDR User's Guide. A user's guide has been developed for one of the three primary analytical tools described in this report. Because much of the discussion of analytical methods, data quality issues, and naming conventions in the CDR User's Guide is also relevant to the other principal tools in the FLOW evaluation tool set, this appendix is included with this report.

About This Project

This report is a product of a WSDOT-sponsored project, FLOW Evaluation Framework Design. The overall objectives of this project are to 1) develop a methodology, framework, and detailed procedures for conducting an ongoing series of evaluations of the performance and effects of the FLOW traffic management system now in operation on Puget Sound area freeways; 2) create tools for performing those evaluations; and 3) supplement earlier evaluation data with updated results by using the developed framework to evaluate selected portions of the FLOW system. This report documents the results of work on the first two objectives. A separate report, "Central Puget Sound Freeway Usage and Performance: Interim Report", addresses the third objective.

1. Introduction

This report describes the analytical methods and tools that were developed for a WSDOT-sponsored project, FLOW Evaluation Framework Design. These methods were designed to address the need for a cost-effective process by which timely evaluations of freeway usage and performance on the central Puget Sound freeway network could be prepared.

The analytical approach outlined in this report focuses on the usage and performance of mainline general purpose and high-occupancy freeway lanes in the central Puget Sound region. Potential future activities include the extension of this methodology to other aspects of the freeway system, such as freeway ramp traffic management and traveler information systems. Also under consideration are extensions of the methodology to data collection systems in other regions.

The algorithms used in this project were chosen in an effort to balance considerations of accuracy, calibration requirements, and usefulness of the resulting performance measures. These techniques involve analytical assumptions and limitations, which are noted in this document. The tool set developed for this project is a work in progress. It represents the initial software implementation of the analytical techniques used in this evaluation process; its principal purpose is to automate as much as possible the process by which an evaluation is performed. Potential future software development activities include an enhanced user interface, additional functionality, and greater integration with other tools and databases. All descriptions of tools are based on V 1.1 of the software (March 1999).

2. Evaluation Methodology

Overview

This section describes an evaluation approach that was developed to perform a systemwide evaluation of the usage and performance of the major corridors of the Seattle area freeway network. The discussion in this section begins with an overview of the guidelines that were used to determine the evaluation method, data collection approach, and analytical tool development process. This is followed by a discussion of the measurement sampling approach used, and criteria that were used to select the locations that were included in the measurement sample.

Evaluation Design Guidelines

The design of this evaluation methodology was driven by three sets of issues: overall project objectives, data collection issues, and analytical tool issues. The overall project objectives were established in the original project proposal, based on the initial goals of the WSDOT. Data collection issues revolve principally around the pragmatic considerations associated with the tradeoff between comprehensive, high quality data and cost (in both dollars and time). Analytical tool issues involve the user-friendliness of the analysis process. The following is a discussion of the objectives and concerns that guided the development of the evaluation methodology, as well as the ways in which the methodology addressed these issues.

Overall Objectives

The initial proposal for the FLOW evaluation design project was based on a recognition that a systematic, cost-effective method of evaluating the usage and performance of a freeway network was desired. The principal goals at that time were the following:

1) **Easier evaluations.** Previous evaluations were of a "custom-made" nature; they did not have the benefit of already-developed methods or analytical tools. There was a need for a more automated approach that would make the evaluation process faster and easier to implement.

2) **More frequent evaluations.** There have only been two formal evaluations of the FLOW system and the associated freeway system since 1981. There was a need for an evaluation system that would facilitate more frequent evaluations without the need to re-invent procedures and analytical methods.

3) **More effective data analysis methods.** There was a desire for a set of performance measures that would allow the planner or engineer to extract additional insights about the potentially complex interactions that affect the freeway system.

4) **More effective communication methods.** There was a need for data presentation techniques that would more effectively and succinctly summarize the usage and performance of the

freeway system in ways that would be useful for technical professionals, policy decision makers, and general audiences.

5) **Comparable results.** A useful element of an ongoing evaluation process is the ability to monitor trends over time. This requires that the evaluation method produce information that is consistent and comparable from one evaluation to the next. There was therefore a need for an evaluation process that used a consistent set of measures, so that comparisons over time could be made.

Project Response: The project focused on five principal ways in which the evaluation process could be made easier and more frequent. First, traffic data had to be made available in a convenient and timely fashion. This was addressed by the WSDOT in 1997 when it began to record detailed 5-minute traffic data for the central Puget Sound freeway network onto compact disks and update the CD set on a regular basis. Second, tools to facilitate the analysis of that data had to be developed to allow users to produce results more easily. Furthermore, given the significant quantity of data that needed to be processed, it was important that those tools offered the option for automated operation as much as possible. The development of such a tool set was thus included as an objective in this project. Third, a repeatable step-by-step process to perform an evaluation in a consistent way had to be developed, so that successive evaluation results could be compared with one another and trends could be monitored. The development of an evaluation methodology and associated user guide as part of this project addressed that objective.

The development of more effective analysis methods and communications techniques was also addressed in this project. The analysis methods eventually developed combined traditional traffic engineering measures such as average daily volume with more detailed measures that focused on the influence of factors such as time of day, lane type, and corridor location on traffic patterns. Also included were measures that more directly measured the effects felt by the individual traveler, such as travel time and speed; the variability or reliability of travel was another performance consideration that has significance for the traveler. These measures benefited from new methods of presentation, including the use of graphical formats and color to better convey the meaning of the analysis to a wider range of audiences. Finally, the ability to compare results was addressed through the development of standard sets of measures, evaluation tools, and measurement locations.

Data Collection Issues

The process of collecting raw traffic data for further analysis offers several challenges. First, there is the practical aspect of establishing a mechanism to collect traffic data in a timely and cost-effective manner. Second, there is the challenge of collecting measurements that are at a sufficient level of detail to accommodate a range of analyses without becoming overwhelmed with data. The following data collection issues were considered in the evaluation method design:

1) **Data collection limits.** An important issue for traffic data collection is one of tradeoffs, namely the challenge of collecting a sufficient quantity of high quality traffic data within the restrictions imposed by time, available human and equipment resources, and funds. Given these data collection trade-offs, the evaluation method needs to include criteria to guide the allocation of scarce resources to freeway locations of high utility and interest.

2) **Readily available data.** Because resources are limited, it is important that readily available data sources be used whenever feasible.

Project Response: The limitation of data collection resources was addressed by establishing a core set of representative measurement locations that could be used at each evaluation (the selection of locations was based on a set of criteria that will be discussed shortly). The analysis tools were designed to allow other locations to be studied as well (as long as data were available), but the core set of locations allowed for consistent measurements from one evaluation to the next. The project was also fortunate to have access to a substantial database of WSDOT traffic data that was converted to CD as this project was being developed; this data set is a key resource for the evaluation process. This data set is updated using a collection mechanism (a network of electronic sensors) that is already in existence, thus saving this project (and future evaluations) considerable time and money.

Analytical Tool Issues

It is also important to supplement an evaluation process with analytical tools and methods that facilitate the evaluation approach by providing fast, convenient results in a cost-effective manner. The following are issues associated with the design and operation of analysis tools that were considered for development in this project.

1) **Emphasis on automation.** The evaluation of a freeway network covering a region the size of the central Puget Sound area requires that analytical tools operate as quickly as possible to process the significant quantities of associated traffic data that are involved and to allow more frequent evaluations to take place. As part of this project, an interim evaluation report was produced using an initial evaluation procedure that was only partially automated. From this tedious and repetitive experience, it became abundantly clear that insufficient automation significantly inhibits the ability and desire to perform frequent or timely evaluations. Tools should therefore operate in an unattended mode as much as possible, especially when lengthy, highly repetitive tasks are involved.

2) **Blending custom and off-the-shelf software.** In the first phase of this project, prototypes of analytical tools were developed using commercial spreadsheet software and custom macros. While this method was useful as a proof-of-concept exercise, it became clear that software written in a high-level computer language would offer the performance (speed) needed to process the large quantities of data involved. At the same time, rather than write certain functions such as graphical output from scratch, it was noted that those functions could be provided more cost-effectively by using off-the-shelf software that already included such capabilities. As a result of these early project results, it was determined that the use of a blend of custom-built and off-the-shelf software would be the most cost-effective way to build analytical tools.

3) **Minimal hardware and software requirements.** To enhance the ease of use of the analytical tools, minimal hardware and software requirements were desired, with an emphasis on ubiquitous hardware and inexpensive, readily available off-the-shelf software.

Project Response: The approach taken to address these issues involved a blend of commercial packages, custom macros, and custom high-level code, all operating on a readily available computer platform. First, it was decided that certain data processing functions were already well-developed in available commercial packages and therefore did not need to be reinvented. Specifically, some post-processing tasks and graphics output processes are performed within Microsoft Excel using its built-in macro language. Excel was chosen because of its built-in graphics capabilities, automation options via user-written macros, and widespread availability at WSDOT. Second, tasks that required significant data processing and/or the computation of new performance measures were custom-coded to increase speed of processing; this code was written in the industry standard ANSI C language. These software analysis tools were designed to allow unattended operation for selected performance measures that involved significant processing time and/or repetitive tasks. By providing an operation mode that did not require constant user participation, yet was controllable by a user-specified batch file or command file, extended analysis could be performed in parallel with other activities, or even overnight as needed. Third, the standard platform requirement for these tools was an IBM PC-compatible computer operating with Windows 95, Windows 98, or Windows NT, and including a CD-ROM drive, color monitor, and optional color printer. Recommended specifications include a Pentium II 200 Mhz processor (faster is preferred), 64 Mb of RAM, and at least 1 Gb of free disk space (more space is preferred).

As suggested above, the measurement approach taken for the FLOW evaluation is to measure the state of the system by sampling performance at a core set of locations, where each location is in some way representative of system performance in that geographic vicinity, for as many sampled time periods as is feasible. This approach addresses the reality that though a comprehensive evaluation would ideally include measurements performed at as many network locations as possible, at different times, and under a wide range of conditions, in practice the ability to perform a comprehensive review of this sort is limited by resource constraints as well as the fixed number of measurement sites for which usable measurement equipment is already in place. The next discussion describes the guidelines used to select the locations at which performance of the system will be sampled.

Performance Sampling Considerations

The selection of representative freeway locations requires a balance between the desire for highly detailed comprehensive measurements of the system vs. limited time, staffing, and equipment resources. Ideally, one could select and evaluate an exhaustive collection of measurement points for a broad range of conditions, to determine in detail the variations in system performance as a function of both location and ambient conditions. A benefit of this approach is that it focuses on the accumulation of detailed data at many locations to minimize the possibility that important locations or conditions will be missed.

Clearly, though, time and resources impose limits on the extent to which a comprehensive road network evaluation of this type can be performed. At the other extreme, one could use an aggregate index, or set of indices, that somehow attempts to capture the essential elements of the highway system's performance. This measure could be determined by combining data collected at a limited number of discrete locations and producing an overall index or metric of the system's performance. The level of detail of that metric could vary depending on the nature of the measure of interest; a simple example (that is, simple to describe if not to calculate) would be a "congestion index", that uses a single value to "rate" the overall level of congestion on the system. While such an approach might be less data-intensive and less costly to calculate, there remains the problem of developing an aggregate measure that is accurate, meaningful, and intuitive. In addition, important variations between locations could be masked by the use of only one or a few summary measures that "average out" noteworthy differences.

As a practical matter, it was decided that the measurement of the performance of a freeway network would take place at a limited number of discrete points. At the same time, there is potentially significant analytical value to be gained from sensitivity analyses that study the variations in traffic as a function of changes in location and other conditions. Therefore, for the FLOW evaluation framework, a middle ground was chosen to provide an evaluation that was easier to implement, but comprehensive enough to be useful for overall system analysis. First, a core set of locations was selected to represent significant elements of the system. (The use of a fixed set of locations also allows trends over time to be monitored while holding location-dependent conditions as constant as possible.) Second, a range of analytical options is provided to make the most of the data that is collected by measuring traffic performance in various ways (traditional measures like daily volume, as well as other measures such as travel reliability) and exploring the relationship between those measures and the freeway's environment; examples of the latter include variations in traffic performance by lane type, time of day, and direction of travel.

This approach, based on performance sampling at a representative core set of locations, tries to provide an appropriate balance by avoiding inaccuracies that might result from evaluating a system with too few measurement points, while also avoiding an expensive and time-consuming, data-intensive analysis effort. It should be noted that this approach still allows the analytical tools to be used at additional locations above and beyond the core set of locations, as resources and data availability allow, and analytic needs dictate.

Sampling Location Selection Criteria

Measurements taken at a specific location in a road network provide information on typical traffic conditions in the immediate vicinity. The degree to which measurements at one site are representative of those at a nearby site depends on such issues as the location of intervening traffic sources and sinks (i.e., on-ramps and off-ramps), variations in road geometry (e.g., elevation changes, lane widths, number of lanes) and differences in physical conditions (roadway or facility changes, such as tunnels or bridges). Measurements also depend on the conditions under which the measurements are taken (weather; time of day, week, month, or year; how "typical" the traffic conditions were at the time of the measurement). Variations in those conditions can cause two measurement sites in close proximity to one another to produce very different measurement results. Therefore, in order to select individual locations that reflect nearby sites as much as possible, it is important that issues such as those mentioned above be considered to avoid errors of commission (picking a nonrepresentative site) or omission (not sampling road segments that are potentially significant). The following is a description of criteria used to select sampling locations.

Picking Measurement Locations

Representative measurement locations for the FLOW system were chosen according to the following criteria:

- Measurement locations should be located at or near sites that are typically of interest from a traffic management and/or traveler viewpoint. This includes high volume locations, key facilities (e.g., bridges), and routes that lead to popular destinations such as major employment centers, universities, or central business districts.
- Measurement locations should be at or near sites that are expected to be of future interest from a traffic management viewpoint. Examples include perimeter locations that are not heavily traveled now, but are expected to be impacted by future residential or commercial growth, or growth management directives.
- Each major interstate and state freeway facility should be represented by at least one measurement site.
- If a major corridor has more than one major segment, each segment is represented by at least one measurement site. The determination of what constitutes a "segment" depends on traffic flow issues such as typical travel patterns (e.g., commute patterns) and major intersections between corridors, as well as practical matters such as available measurement installations. An example is the I-5 corridor, which can be divided into north, north central, central, and south central segments for the purposes of an evaluation; the endpoints of the segments could be determined by noting the location of major interchanges (I-5/I-405 Swamp Creek, I-5/SR 520, I-5/I-90, I-5/I-405 Tukwila). Each segment would contain at least one measurement location.
- Whenever feasible, measurement locations should be in corridor segments that are part of a trip route selected for travel time analysis. (The travel time analysis will be discussed in the next section.)
- Whenever feasible, measurement locations should be located in corridor segments that are designated by the Puget Sound Regional Council as congested segments that are the focus of their Congestion Management System.

Picking Measurement Sites

Once a measurement location (i.e., a general area description such as "I-5 near the University District") has been selected, one or more specific measurement sites at that location are chosen from among those sites that have existing data collection sensors maintained by the WSDOT (e.g., mainline northbound lanes on I-5 at the NE 45th Street overpass). These sites use electronic sensors embedded in the pavement to record basic data on volume and lane occupancy percentage; some sites also collect speed and vehicle length data. The selection of specific measurement sites is based primarily on the presence of existing WSDOT data sensors; however, locations of supplementary measurement sites can also be taken into account. Examples of supplementary sites include locations where average vehicle or bus occupancy data are collected; such data are available from the WSDOT-sponsored HOV Lane Evaluation project or transit agencies. Whenever possible, primary and secondary measurement sites should be coincident, so that the volume, lane occupancy, and vehicle occupancy data for a measurement location are as compatible as possible. When this is not possible, primary and secondary measurements should be chosen to be as close to one another as practical.

- Measurement sites are restricted to those corridor locations for which the WSDOT has active inductive loops in place in each and every mainline general purpose lane as well as each HOV lane (if any exist there). These loops are identified with WSDOT loop identifiers that have the _Mx/MMx (general purpose) or _MxH (HOV) designations, where x refers to the direction (N, S, E, W, R). These data are accessible on CD via a set of analysis tools (described in the next section). The loop identifier system is described in detail in the CDR User's Guide, included as an appendix to this report.

- Whenever possible, measurement sites are located at or near supplementary measurement sites used for vehicle occupancy data collection by the WSDOT HOV Evaluation project and/or transit agencies. These data are used to estimate person throughput.

- Whenever the vehicle sensor data, car occupancy data, and transit occupancy data cannot be at exactly the same location, the priority is to match the transit data location with the loop location, and then to use the nearest available car occupancy data.

Selecting a Sampling Time Period

In addition to selecting analysis sites, a time period for the analysis should be selected. Typically, annual average performance measures are used. However, some studies might require a time interval other than one calendar year. Examples of such studies include before-and-after performance changes relating to construction projects or operational changes, or an analysis of seasonal or special event-related variations of freeway performance. In addition, it may be appropriate to pick only weekday, only weekend, or seven-day-a-week data, depending on the analysis.

Performance Sampling Example: Central Puget Sound Freeway Usage and Performance Interim Report

The performance sampling approach described in this section was used to prepare the Central Puget Sound Freeway Usage and Performance Interim Report. This report includes 1) an introduction that describes the FLOW system and establishes the context for its implementation; 2) a description of freeway system usage; and 3) a range of performance analyses for selected aspects of the FLOW system. Information about the measurements and sites that were analyzed is presented later in this document.

3. User's Guide: FLOW Evaluation Analysis Procedures and Tools

Introduction

The remainder of this document is a user's guide to the capabilities, operating instructions, and underlying analytical bases for the software tools and algorithms used to analyze the usage and performance of the freeway system. In this section, the functionality of the analysis tool set and procedures for their use is presented. In section 4, a series of example applications of the evaluation process are summarized, including 1) the process by which a "Central Puget Sound Freeway Usage and Performance" interim report was prepared; 2) the potential use of the evaluation tool set to provide supporting data for the WSDOT Northwest Region's "Ramp and Roadway Traffic Volumes" summary report; and 3) the use of the tool set to perform various system operations diagnostics and interpret performance analyses. In section 5, known analytical assumptions, limitations, and caveats of the tool set are summarized.

The following topics are discussed in this section:

a) **Data overview.** The data set upon which the evaluation tool set is based is described, including contents, level of detail, and availability.

b) **Tool set overview.** Each analysis tool in the evaluation tool set is described, including functions and operating platform requirements.

c) **CDR (Compact disc Data Retrieval) software.** A catalog of data processing options available for evaluating the freeway network using the CDR data retrieval tool is outlined. Illustrations of sample output are shown. (A complete user's guide to CDR is available in the Appendix of this report.)

d) **CDR Analyst software.** The range of analysis options available for evaluating the freeway network using the CDR Analyst post-processor program is presented, along with step-by-step operating instructions. Pre- and post-processor macros, templates, and related data sets are also described. Sample illustrations of output from each analysis option are included.

e) **CDR Analyst algorithms.** The algorithms used to implement the CDR Analyst analysis options are described.

f) **CDR Auto software.** The purpose and operating instructions for the CDR Auto pre-processor program are presented.

g) **CDR Analyst utilities.** The CDR Analyst Excel-based macros and template files are described.

h) **Data Quality Mapping utility.** The traffic data file data quality macro is described.

The analysis options developed to date focus specifically on the evaluation of the mainline general purpose (GP) and high-occupancy vehicle (HOV) lane network in the central Puget Sound region.

Data Overview

The data set used by the analysis tools described in this report is available on compact discs produced by the WSDOT. These CDs include traffic data collected from electronic inductance loop sensors installed at approximately 0.5-mile intervals on mainline lanes and ramps of freeways and state highways in the central Puget Sound region, including I-5, I-405, I-90, SR 520, SR 18, SR 522, and SR 99. Vehicle presence is detected by the sensors, and the resulting detection data are collected at 20-second intervals and transmitted to the WSDOT Transportation Systems Management Center for processing and archiving. The data are archived at 5-minute intervals (i.e., 15 consecutive 20-second values are combined to produce a single 5-minute value). CD archives are available for data starting from mid-1993, with 2 to 4 CDs required to hold a year of data from all sensor locations in the central Puget Sound freeway network.

The principal values recorded by the sensors are vehicle volumes and average lane occupancy percentage. Vehicle volumes at a particular roadway location are estimated by recording the number of times that an inductive loop embedded in an individual lane of a road or ramp is "triggered" by a passing vehicle. This count is recorded 24 hours a day, except when equipment is turned off, being serviced, or inoperative. Vehicle volumes can be a useful measure of facility usage; they can also be combined with information about per-vehicle person occupancy to estimate person throughput. Five-minute vehicle volumes can also be aggregated to produce data for other time intervals (e.g., hourly, daily, yearly averages).

The loop sensors are also used to estimate lane occupancy. Lane occupancy refers to the percentage of time that a given loop is in a triggered ("on") position, where the triggered state indicates a vehicle's presence within the loop's detection range. For example, if a loop recorded a lane occupancy of 10 percent for a 5-minute period, this would mean that vehicles were sensed within the loop's detection range for a total of 30 seconds during the 5-minute interval (10% of 5 minutes = 30 seconds). Lane occupancy can be considered a surrogate measure for the density of vehicles on a roadway; it can also be used as a measure of congestion. Lane occupancy can also be combined with vehicle volume estimates to derive estimated vehicle speeds.

The estimated validity of each 5-minute data value is also recorded on the CD archive in the form of a code that summarizes the data quality of its constituent 15 20-second values. The 5-minute data validity codes include "good" (all 15 constituent values are considered valid), "bad" (all 15 values are considered invalid), "disabled" (all 15 values were collected when the data collection equipment at that sensor site was not operational), or "suspect" (all other combinations of 15 data point conditions). Additional information about the data set is available in section 5 (Analytical Assumptions and Limitations: Input Data Considerations) and the Appendix (CDR User's Guide).

At selected locations, average vehicle speed and vehicle length information are also estimated, using pairs of mutually operating sensors.

Tool Set Overview

The evaluation tool set includes software that performs the following functions: 1) summarize raw traffic data and compute performance measures (CDR and CDR Analyst); 2) present the analyzed data in text and graphical formats (CDR, CDR Analyst, and CDR Analyst utilities), and 3) reformat raw traffic data for use by CDR Analyst (CDR Auto). There are five categories of analysis options; they include 1) basic traffic statistics, 2) site performance measures, 3) corridor performance measures, 4) travel time performance measures, and 5) system operations diagnostics measures.

Three analysis tools will be discussed in the remainder of the report:

- **CDR (Compact disc Data Retrieval):** This program, developed by WSDOT Northwest Region, accesses 5-minute traffic data stored on compact disc, computes summary statistics based on that data, and produces text output files. Output options include volume, lane occupancy percentage, and speed and vehicle length category (at selected locations). Data can be summarized at various levels of detail ranging from peak-hour to yearly statistics. Minimum data quality thresholds used to compute summary statistics are user-specified.

- **CDR Analyst (and associated utilities):** Like CDR, CDR Analyst also uses traffic data stored on WSDOT-produced compact discs. CDR Analyst can produce additional performance measures, including 24-hour volume and speed profiles, congestion frequency statistics, corridor congestion summaries, travel time estimates, and travel time reliability measures. CDR Analyst output can be post-processed using Analyst utilities and templates to produce color graphics output.

- **CDR Auto**: This utility is a pre-processor for CDR Analyst that converts compact disc traffic data to a file format that can then be used by CDR Analyst.

There are several utilities and templates that post-process CDR and CDR Analyst output to produce graphical output that can be used for analysis and report preparation. These ancillary programs are written using the macro language of Microsoft Excel, and use the graphics capabilities of that program. They will be described later in this report.

CDR has been revised a number of times over the past two years, and now features a user-friendly interface and operating process. CDR Analyst, CDR Auto, and the associated utilities and templates are still in active development; for the most part they utilize a command line interface, and require some manual modifications of the data to produce graphics output. Tentative plans call for enhancements to the user interface of these command-line-based tools in a future phase of this project.

Please note that the evaluation techniques described here are independent of the measurement sites selected (data availability notwithstanding). Although they were initially used on a set of core locations in the freeway usage and performance interim report (described later in this document), the tools are designed to perform detailed analyses at any location and for any time period, provided that the necessary data is available in an appropriate format.

The following are summary descriptions of each tool. Most of the discussion centers on CDR Analyst. CDR is discussed in detail in the Appendix, while CDR Auto is a data re-formatting program that requires minimal discussion.

CDR (Compact disc Data Retrieval)

CDR is a program developed by WSDOT Northwest Region to access 5-minute traffic data stored on CD, and produce a summarized version in a text file format. Users can specify specific dates of data collected, specific locations (actually, specific lanes), and various levels of summarization. The output is in the form of a text file that can also be read directly into a spreadsheet program such as Microsoft Excel. Standard output from the program includes 5-minute raw traffic data that includes traffic volumes and average lane occupancy percentage as well as a data quality/validity indicator. Data can also be aggregated to a 15-minute, hourly, peak hour, peak period, daily, weekly, monthly, or yearly level. At some locations, average estimated speed and vehicle length information can be provided. Figure 1 shows the user-interface for CDR, including the three general levels of output: raw data (left one-third), daily summaries (middle one-third), and multi-day or yearly summaries (right one-third). Figures 2 and 3 show typical output.

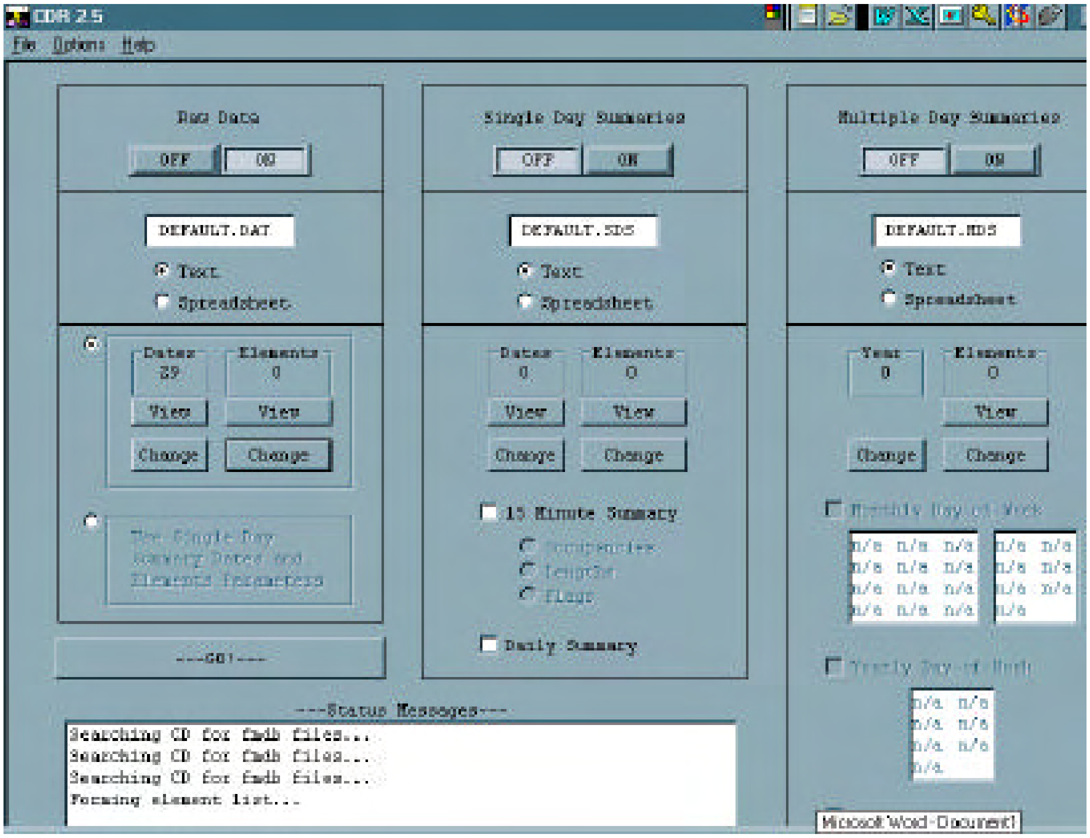


Figure 1. Main CDR Screen

****

Filename: 5TO15.DAT

Creation Date: 02/2/98 (Wed)

Creation Time: 03:16:59

File Type: SPREADSHEET

****

ES-145D:_MS___1   I-5 Lake City Way   170.80

09/01/97 (Mon)

---Raw Loop Data Listing---

Time	Vol	Occ	Flg	nPds
0:00	49	3.80%	1	15
0:05	37	2.90%	1	15
0:10	38	3.50%	1	15
0:15	34	2.60%	1	15
0:20	48	4.40%	1	15
0:25	44	3.60%	1	15
0:30	35	2.80%	1	15
0:35	33	3.30%	1	15
0:40	28	2.50%	1	15
0:45	30	2.30%	1	15

Figure 2. Example of 5-minute Output (Volume and Occupancy)

****

Filename: AADT.MDS

Creation Date: 02/2/98 (Thu)

Creation Time: 10:54:09

File Type: SPREADSHEET

****

ES-145D:_MS___1 I-5 Lake City Way 170.80

Monthly Avg for 1996 Jan (Sun)

---Multi-Day Loop Summary Report---

Summary	Valid	Vol	Occ	G	S	B	D	Val	Inv	Mis		
Daily	VAL	19392	7.50%	1133	18	1	0	4	0	0		
AM Peak	VAL	1493	3.50%	142	2	0	0	4	0	0		
PM Peak	VAL	5069	15.60%	190	2	0	0	4	0	0		
AM Pk Hour	VAL	1381	10.00%	47	1	0	0	4	0	0	10:45	11:45
PM Pk Hour	VAL	1576	11.90%	48	0	0	0	4	0	0	13:45	14:45

Figure 3. Example of Multi-Day Output

Additional operating and background information on CDR can be found in the CDR User's Guide, which is installed in PDF format on each WSDOT traffic data CD since early 1998. It is also included as an appendix to this report.

CDR Analyst

The discussion of CDR Analyst is organized as follows. First, an overview of the program and the data that it uses are provided. Next, operating instructions to generate each of the major performance measures produced by CDR Analyst are presented. This is followed by a discussion of each of the primary algorithms used by CDR Analyst to compute those measures.

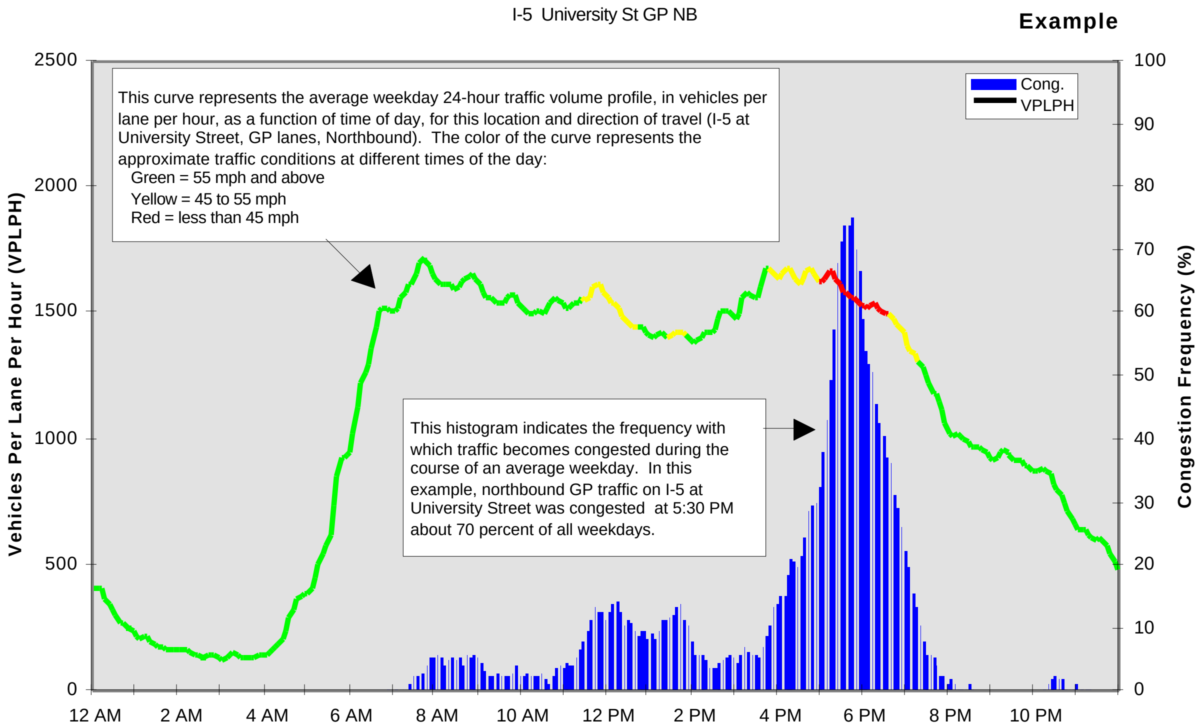
Overview

CDR Analyst is a program developed by TRAC to access 5-minute traffic data stored on CD, and produce performance measures that are not normally available from CDR. Because CDR produces statistics one lane (sensor) at a time, it does not directly produce statistics that reflect all the lanes in a given location traveling in a particular direction (e.g., the total average daily northbound volume on all GP mainline lanes on I-5 at University Street). CDR Analyst addresses this limitation by allowing the user to process all relevant lanes at a specific location when producing statistics. CDR Analyst also computes peak hour, peak period, and daily (e.g., Average Weekday Daily Traffic) values similar to those produced by CDR, but does so for all relevant lanes, e.g., all northbound GP lanes at a specific location, rather than only one lane at a time. In addition, CDR Analyst produces supplementary performance measures beyond those computed by CDR. These measures include the following:

- **Average daily site profiles.** CDR Analyst can process user-specified days of traffic data and produce three site-specific traffic profiles as a function of time-of-day. The first of these is an average 24-hour profile of volume per lane per hour at a selected location for a specified direction of travel (across all lanes), at 5-minute increments. The user can specify that only weekdays will be used, or all days of the week. Second, a corresponding 24-hour estimated speed profile is produced. Third, the program computes a 24-hour reliability profile which estimates the percentage of time that the location is congested, as a function of time of day. The definition of what constitutes congestion is based on a lane occupancy percentage threshold value; this threshold can be changed by the user.

  The resulting output data can then be processed by a spreadsheet-based utility that will produce a graphic that is suitable for analysis or report preparation. As can be seen in Figure 4, the resulting graphic shows a line graph of vehicle volume (measured per lane per hour) as a function of time of day, at a particular site for a given travel direction and lane type (GP or HOV). This graph is then supplemented with the speed estimate profile by adding a color to each data point on the volume curve, where the color is based on the corresponding speed for that time interval. To determine the color for each data point, the speed information is mapped to several speed ranges; each range then corresponds to a different color. For example, if the calculated speed is estimated at above 55 mph at 9:00 am, the corresponding volume line segment is green at that time; if the speed is between 45 to 55 mph, the line is yellow; and if the speed is below 45 mph, the line is red. While the specific speed ranges can be changed, the green/yellow/red color system is intended to approximately represent free flow, restricted flow, and congested traffic conditions, respectively. By combining the volume and speed profiles in this way, a single graph can allow the viewer to distinguish between, for example, low traffic volumes associated with free-flow traffic, and low volumes that are the result of congested conditions.

Figure 4. Estimated Weekday Volume, Speed, and Reliability Profiles



This graph is then supplemented with an overlay of the reliability profile, in histogram or column graph form. When displayed in tandem with the color-coded volume line graph, the reliability histogram can help to highlight the relationships between the averaged values of volume and speed, and the frequency of congestion. By measuring the frequency of congested conditions, the histogram indicates the extent to which there is significant day-to-day variability from the average values.

- **Average daily corridor profile.** CDR Analyst can also process a series of user-specified locations along a corridor and produce an average 24-hour profile of lane occupancy percentage at each of those locations for a specified direction of travel, at 5-minute increments. The user can specify that only weekdays will be used, or all days of the week.

  The output file can then be displayed using a spreadsheet-based template that will produce a graphic that is suitable for analysis or report preparation. As can be seen in Figure 5, the resulting graph shows a contour map of the lane occupancy percentage information that is color-coded according to estimated congestion level. The result is a corridor overview of traffic conditions as a function of location along the corridor as well as time of day and direction of travel. The graphic is in a topographic map style, where elevation is replaced by average traffic conditions. Note: The template that is used to produce this graphic will produce one contour map for one direction of travel. In order to produce a two-direction graphic such as the one shown in Figure 5, each contour map and any descriptive map art must be produced separately, then brought together using a standard drawing program such as Corel Draw.

- **Average travel time profile.** CDR Analyst can process corridor information to produce three 24-hour profiles related to a specific trip. First, it can estimate the average travel time from one point to another on one corridor (i.e., a particular interstate freeway or state highway that has vehicle sensor installations) as a function of the time that the trip starts (in five-minute increments, throughout an average 24-hour day). Second, a 90th percentile travel time is computed as a function of trip start time. The 90th percentile travel time is a travel time N such that 90 percent of the time the trip will take less than or equal to N. For example, if the 90th percentile travel time is 15 minutes for a trip starting at 7 AM, this means that 90 percent of the time a trip that starts at 7 AM is estimated to take no more than 15 minutes, based on available data. Third, trip travel time reliability is estimated as a function of trip start time by computing the likelihood (as a percentage) that the overall trip speed is less than 45 mph.

  The output file can then be displayed using a spreadsheet-based template that will produce a graphic that is suitable for analysis or report preparation. As can be seen in Figure 6, the resulting graph shows a line graph of average travel time, a line graph of 90th-percentile travel time, and a superimposed histogram (column graph) that shows the travel time reliability measure. All three measures are displayed as a function of trip start time for a specific origin and destination.

Selecting Data to be Processed by CDR Analyst. CDR Analyst uses all the data contained in the input files specified by the user. The basic set of data files created by the CDR Auto pre-processor program to be used by CDR Analyst contain seven-day-per-week data for an entire calendar year. Each data files contains one combination of location, travel direction, and lane type, e.g., cabinet 100, northbound, GP. The user has the option to extract two types of subsets from this collection of data files. First, the user can decide which location(s), direction of travel, and lane type (GP or HOV) will be analyzed by the program, by indicating the specific data files of interest, either individually or in a batch file. Second, the user can request that only weekday data be processed (Monday through Friday); otherwise, all seven days of the week are processed. In either case, data for an entire calendar year is processed.

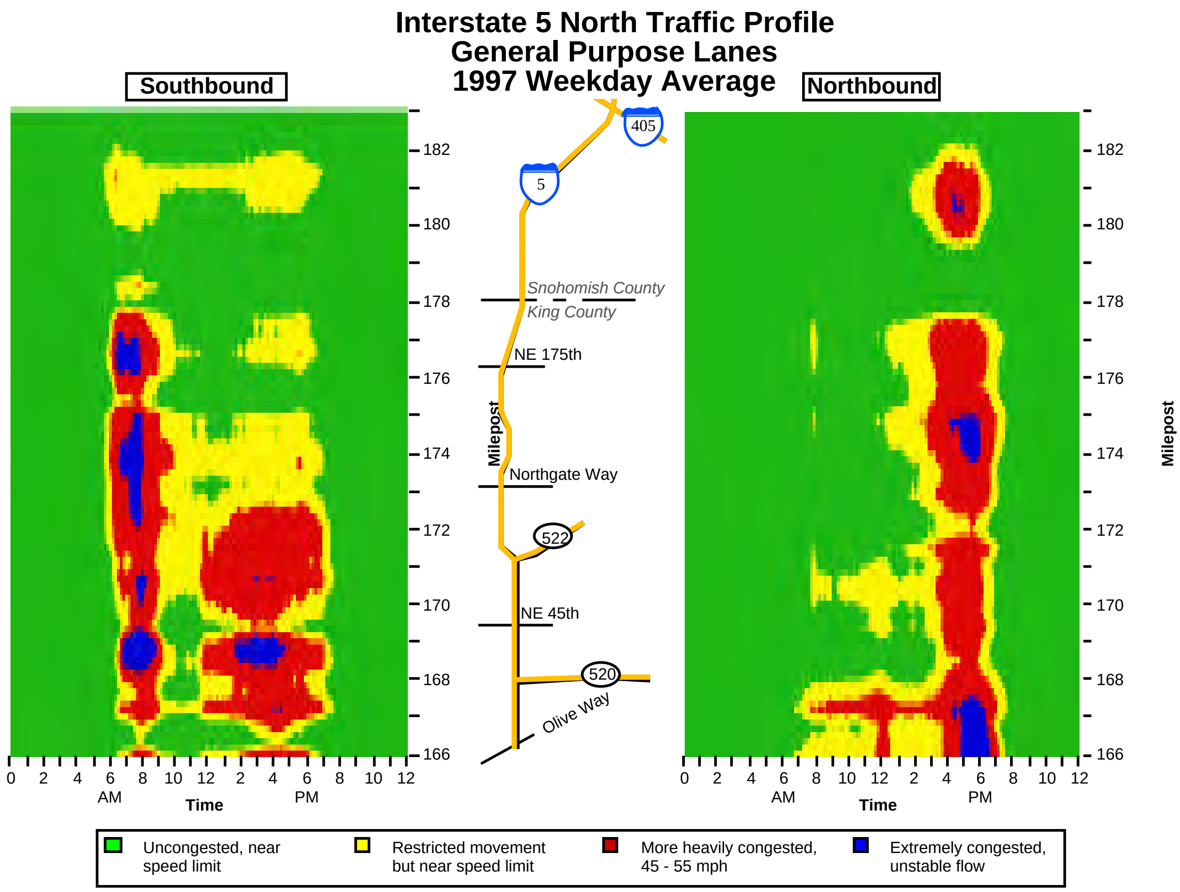
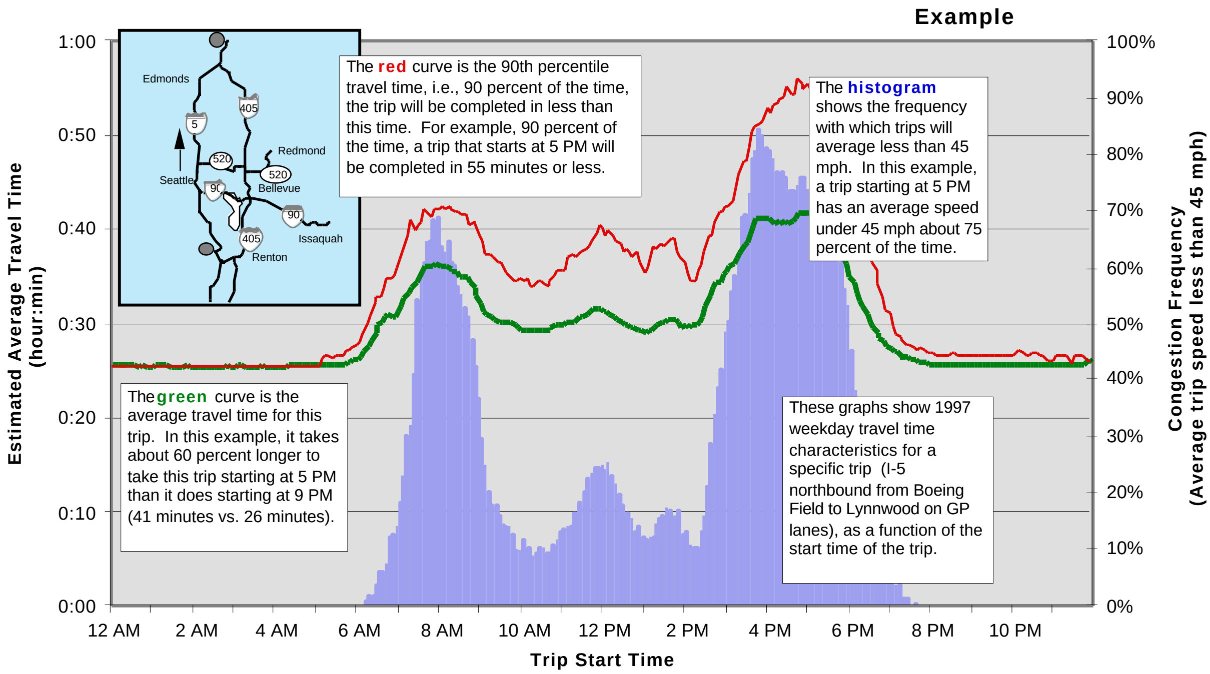


Figure 5. Corridor Traffic Profile

These two contour maps indicate the average traffic conditions along a freeway corridor at different of the day, for each direction of traffic. The horizontal scale represents an average weekday 24 period, while the vertical scale represents the milepost locations along the corridor. The center provides a rough approximation of the location that corresponds to a given milepost.

Figure 6. Estimated Average Weekday Travel Time

Northbound Interstate 5 GP Lanes, Boeing Field to Lynnwood (25.4 mi) (1997)



If the user wishes to pick some other subset of the year (e.g., one particular month, weekends only, Tuesdays through Thursdays only), these data files can be created manually by the user using the CDR program (**not** CDR Analyst). The user must run CDR once for each combination of location, travel direction, and lane type.

Operating Instructions

The following are step-by-step instructions on the use of CDR Analyst to produce each of the three profile types described above. Before using CDR Analyst, the user should edit the text file **CDR_pref.txt**. CDR_pref.txt allows the user to specify a number of parameters associated with data quality thresholds, days of data to be processed, etc. These are values that typically do not vary from one run to the next; by putting them in a preference file, the user does not need to answer the same questions repeatedly when processing, for example, a series of locations along a corridor or in a region. CDR_pref.txt can be edited by any text editor (e.g., WordPad). It should be saved in a text format. **See the CDR-Analyst readme file for more information on preference file options.**

To produce average daily site profiles:

(Note: This process can also be used to create ramp and roadway report statistics. See Section 4, Example B for more details.)

1. Double-click the program icon to start. The most recent version is Rev 1.1 (as of May 1999).

2. Do you want to extract weekday data? Answer Y (yes) or N (no).

   All data files include both weekday and weekend data. If you want to produce statistics based only on weekdays, indicate your choice. Otherwise, all seven days of the week will be used to produce the desired statistic.

3. Do you want to run a batch job? Answer Y (yes) or N (no).

   The batch job option is used to process multiple locations for corridor profiles and travel time profiles. It is not used for a single site profile, so answer N (no).

4. Do you want AWDT/AADT information? Answer Y (yes) or N (no).

   If AWDT/AADT statistics are desired, indicate that here.

5. Do you want peak period information? Answer Y (yes) or N (no

   If peak hour and peak period statistics are desired, indicate that here.

6. Input the file name.

   At this point, type the name of the file that contains the data for the site that you are interested in. The following is the (CDR Auto) file naming convention:

   CCCCLDYY.dat

   where

   CCCC = cabinet number (i.e., the measurement site); leading zeroes are used if < 4 digits

   L = lane type (G = GP, H = HOV)

D = direction of travel (N=northbound, S=southbound, E=eastbound, W=westbound)

YY = year (e.g., 97 = 1997)

NOTE: CDR Analyst assumes that all files associated with the analysis (the CDR Analyst program, the preference file CDR_pref.txt, the input data files, batch files if any) are in the same directory.

7. Accept suspect data? Answer Y (yes) or N (no).

   Each 5-minute data point has a data quality indicator flag associated with it. This data flag indicates whether the data point is considered "good", "bad", "suspect", or "disabled" (i.e., the equipment is off-line). CDR Analyst uses all available good data, and attempts to replace bad, suspect, or disabled data points with nearby good data. (The replacement method is described later in this section; see "CDR Analyst Core Algorithm.") However, the user has the option to accept suspect data as good data. This option is available because the "suspect" label is based on a conservative threshold. In many cases, suspect data is very much consistent with the good data that surrounds it. However, if there is some question about this, you can specify that suspect data be considered the same as bad data and therefore subject to replacement by nearby good data.

8. Speed estimation method? Answer N (normal) or K (Kalman).

   There are two methods implemented in the program for estimating speeds. The so-called normal method uses the formula that is used by WSDOT in their WebFlow web-based system to estimate speeds. The Kalman method was developed by D. Dailey at the University of Washington. A preliminary version of the Kalman code has been implemented; however, associated constants have not yet been calibrated. Therefore, users are advised to use the normal method in Version 1.1 of CDR Analyst.

9. Output file name.

   The resulting statistics are sent to an output file of the user's choice. The output file is assumed not to exist yet; if it does exist, the user will be alerted and given the option to overwrite the existing file. The output file is placed in the same directory with the input data and CDR Analyst.

10. Use utilities to create graphical output. This is described later in this section.

<u>To produce average corridor profiles:</u>

<u>(NOTE: THIS FUNCTION IS NOT FULLY OPERATIONAL IN VERSION 1.11. See the CDR-Analyst readme file for additional details.)</u>

1. Double-click the program icon to start. The most recent version is Rev 1.2 (as of May 1999).

2. Do you want to extract weekday data? Answer Y (yes) or N (no).

   All data files include both weekday and weekend data. If you want to produce statistics based only on weekdays, indicate your choice. Otherwise, all seven days of the week will be used to produce the desired statistic.

3. Do you want to run a batch job? Answer Y (yes) or N (no).

The batch job option is used to process multiple locations for corridor profiles and travel time profiles. It is used for a corridor profile, so answer Y (yes).

4. D o you want travel time information? Answer Y (yes) or N (no).

   Answer N (no) to this question to get corridor profiles only.

5. Input the batch file name.

   At this point, type the name of the batch file (a text file) that contains the data file names for the sites that you are interested in (file naming convention is shown below). There is no restriction on the batch file name; however, use the CDR Auto file naming convention for all files listed in the batch file. The list of files should be in ascending or descending order of milepost. The following is the file naming convention:

   CCCCLDYY.dat

   where

   CCCC = cabinet number (i.e., the measurement site); leading zeroes are used if < 4 digits

   L = lane type (G = GP, H = HOV)

   D = direction of travel (N=northbound, S=southbound, E=eastbound, W=westbound)

   YY = year (e.g., 97 = 1997)

   NOTE: CDR Analyst assumes that all files associated with the analysis (the CDR Analyst program, the input data files, batch files if any) are in the same directory.

6. Accept suspect data? Answer Y (yes) or N (no).

   Each 5-minute data point has a data quality indicator flag associated with it. This data flag indicates whether the data point is considered "good", "bad", "suspect", or "disabled" (i.e., the equipment is off-line). CDR Analyst uses all available good data, and attempts to replace bad, suspect, or disabled data points with nearby good data. (The replacement method is described later in this section; see "CDR Analyst Core Algorithm.") However, the user has the option to accept suspect data as good data. This option is available because the "suspect" label is based on a conservative threshold. In many cases, suspect data is very much consistent with the good data that surrounds it. However, if there is some question about this, you can specify that suspect data be considered the same as bad data and therefore subject to replacement by nearby good data.

7. Speed estimation method? Answer N (normal) or K (Kalman).

   There are two methods implemented in the program for estimating speeds. The so-called normal method uses the formula that is used by WSDOT in their WebFlow web-based system to estimate speeds. The Kalman method was developed by D. Dailey at the University of Washington. A preliminary version of the Kalman code has been implemented; however, associated constants have not yet been calibrated. Therefore, users are advised to use the normal method in Version 1.1 of CDR Analyst.

8. Output file name.

The resulting data are sent to an output file of the user's choice. The output file is assumed not to exist yet; if it does exist, the user will be alerted and given the option to overwrite the existing file. The output file is placed in the same directory with the input data and CDR Analyst.

9. Use utilities to create graphical output. This is described later in this section.

<u>To produce average travel time profiles:</u>

1. Double-click the program icon to start. The most recent version is Rev 1.06 (as of March 1999).

2. Do you want to extract weekday data? Answer Y (yes) or N (no).

   All data files include both weekday and weekend data. If you want to produce statistics based only on weekdays, indicate your choice. Otherwise, all seven days of the week will be used to produce the desired statistic.

3. Do you want to run a batch job? Answer Y (yes) or N (no).

   The batch job option is used to process multiple locations for corridor profiles and travel time profiles. It is used for a travel time profile, so answer Y (yes).

4. D o you want travel time information? Answer Y (yes) or N (no).

   Answer Y (yes) to this question to get travel time profiles.

5. D o you want daily travel time information? Answer Y (yes) or N (no).

   This question refers to the option to either calculate travel times for each individual day and compute statistics based on this collection of times, or to compute an overall 24-hour profile of traffic data first, then compute travel time based on this aggregate profile. Answer Y (yes) to this question to get all three travel time profiles (average, 90th percentile, travel time reliability). If you answer N (no), you will only get an average travel time that is based on the aggregate 24-hour profile of all days processed. Usually, you will answer Y.

6. Input the batch file name.

   At this point, type the name of the batch file (a text file) that contains the data file names for the sites that you are interested in (file naming convention is shown below). There is no restriction on the batch file name; however, use the CDR Auto file naming convention for all files listed in the batch file. The list of files should be in ascending or descending order of milepost. The following is the file naming convention:

   CCCCLDYY.dat

   where

   CCCC = cabinet number (i.e., the measurement site); leading zeroes are used if < 4 digits

   L = lane type (G = GP, H = HOV)

   D = direction of travel (N=northbound, S=southbound, E=eastbound, W=westbound)

YY = year (e.g., 97 = 1997)

NOTE: CDR Analyst assumes that all files associated with the analysis (the CDR Analyst program, the input data files, batch files if any) are in the same directory.

7. Accept suspect data? Answer Y (yes) or N (no).

   Each 5-minute data point has a data quality indicator flag associated with it. This data flag indicates whether the data point is considered “good”, “bad”, “suspect”, or “disabled” (i.e., the equipment is off-line). CDR Analyst uses all available good data, and attempts to replace bad, suspect, or disabled data points with nearby good data. (The replacement method is described later in this section; see “CDR Analyst Core Algorithm.”) However, the user has the option to accept suspect data as good data. This option is available because the ”suspect” label is based on a conservative threshold. In many cases, suspect data is very much consistent with the good data that surrounds it. However, if there is some question about this, you can specify that suspect data be considered the same as bad data and therefore subject to replacement by nearby good data.

8. Speed estimation method? Answer N (normal) or K (Kalman).

   There are two methods implemented in the program for estimating speeds. The so-called normal method uses the formula that is used by WSDOT in their WebFlow web-based system to estimate speeds. The Kalman method was developed by D. Dailey at the University of Washington. A preliminary version of the Kalman code has been implemented; however, associated constants have not yet been calibrated. Therefore, users are advised to use the normal method in Version 1.1 of CDR Analyst.

9. Output file name.

   The resulting data are sent to an output file of the user’s choice. The output file is assumed not to exist yet; if it does exist, the user will be alerted and given the option to overwrite the existing file. The output file is placed in the same directory with the input data and the CDR Analyst program.

10. Use utilities to create graphical output. This is described later in this section.

(See the CDR-Analyst readme file for additional details on other travel time output types.)

Other Analysis Options

It is important to note that CDR Analyst produces measures based on the data given to it. While it is common to wish to analyze one calendar year, and thus process one-year data files, any other time period can also be analyzed by CDR Analyst. For example, if a construction project occurs during the year, it may be more appropriate to analyze only the data corresponding to the time period after construction was completed. It is also common to analyze all lanes of a particular type (GP or HOV) at a given site. However, individual lanes can also be analyzed. Examples of such studies include outside vs. inside HOV lane comparisons, or passing lane studies. As indicated previously, CDR should be used to produce input data files for time periods other than one calendar year.

CDR Analyst Algorithms

The following are descriptions of the primary algorithms used in CDR Analyst. The discussion begins with the core algorithm that processes 24-hour site profiles. This is followed by descriptions of the speed estimation method, the congestion frequency histogram method, the peak hour/peak period/daily volume estimation method, the contour computation method, the average travel time profile method, and the 90th percentile travel time profile method, and the travel time reliability histogram method.

CDR Analyst Core Algorithm

CDR Analyst's principal process is the conversion of multiple days of multi-lane traffic data into a single average 24-hour traffic profile. The process uses all available "good" data to produce the 24-hour average profile.

To process multiple days into a single average day, the following steps are taken:

- For each 5-minute interval, do the following:

  1. For each day, do the following two steps:

  a. For each lane of traffic to be processed, do the following:

  Look at the data flag of a given 5-minute data value in a given lane in a given day. If the data point is labeled "good", or if it is labeled "suspect" but the user specifies that suspect data is assumed to be good, then use the value as is. If the data is labeled "suspect" and the user requests a data replacement, or if it is labeled "bad" or "disabled", the program searches for a good value in the (temporal) vicinity (and within the same lane) by moving back 5 minutes, then forward 5 minutes, then back 10 minutes, then forward 10 minutes, to a maximum of ±15 minutes. If a good data point is located within that window, it is used as a replacement value. If no such value is found, the data point is not included in subsequent calculations.

  b. Average the resulting values of each lane to get a per-lane average for that 5-minute interval for that day.

  2. Average the resulting per-lane averages across all days to get an overall average for that 5-minute interval.

- Repeat the process for each 5-minute interval throughout a 24-hour day. Use this process for both volume information and lane occupancy data.

This process is used to produce an average 24-hour profile at one site. The process can then be used at a series of sites along a corridor, in batch mode, with the result used to produce corridor profiles and travel time profiles.

CDR Analyst Speed Algorithm

In the "normal" algorithm, the average speed of vehicles at a site, across all lanes, for a given 5-minute interval, is determined by using the same core process described above, then using the resulting average per-lane volume and occupancy to estimate speed. The formula used is

$$v = q/(o*g),$$

where

v = estimated speed (miles per hour)

q = estimated per-lane vehicle volume (vehicles per lane per hour)

o = estimated per-lane lane occupancy (percentage)

g = a constant that incorporates site characteristics such as average vehicle length and loop detector field length.

Using five-minute volume and occupancy data and a constant of g=2.4 (the value used by WSDOT), the formula becomes

Estimated speed, in mph = [(average 5-min. per-lane volume)*12] / (average per-lane occupancy percentage * 2.4)

Because the formula is considered less accurate at low and high speeds, the following thresholds are used: if the resulting speed is greater than 60 mph or less that 10 mph, the speed is modified to 60 mph or 10 mph, respectively; also, if the lane occupancy is less than 12 percent, the speed is set to 60 mph. These thresholds are the same as those used by WSDOT.

Caveats: The accuracy of this approach is of course dependent on the input data. Other studies have suggested that the accuracy of this formula is especially dependent on the constant g, and that there is a possible dependence of g on not only vehicle and site characteristics, but lane occupancy as well, suggesting that different g values should be used in low, medium, and high occupancy domains. For now, however, analyses of the FLOW system use g = 2.4, which is the value used for FLOW map online displays of speed (as of March 1996). This value can be changed by the user.

The Kalman method was developed by D. Dailey at the University of Washington[1]. A preliminary version of the code has been implemented; however, associated constants have not yet been calibrated. Therefore, users are advised to use the normal method in Version 1.1 of CDR Analyst.

---

[1] D. J. Dailey, *A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements*, Transportation research. Part B: Methodological, 1999.

CDR Analyst Congestion Frequency Algorithm

The congestion frequency percentage that is computed for a specific site is based on the frequency with which the average lane occupancy percentage exceeds a user-specified level. This congestion measure is computed using the same core process described above, with the following differences (noted in underlined text):

- For each 5-minute interval, do the following:

  1. For each day, do the following two steps:

  a. For each lane of traffic to be processed, do the following:

  Look at the data flag of a given 5-minute data value in a given lane in a given day, and perform data replacement if the user requests it and it is feasible (same as described in the core algorithm).

  b. If a sufficient number of lanes of valid values result from the data replacement process for that 5-minute interval, average the resulting values of each lane to get a per-lane average for that 5-minute interval for that day. Then, increment an "eligible day" counter, and determine if the resulting average lane occupancy exceeds the user-specified threshold for determining congestion. If it does, increment a separate "congestion frequency" counter for this 5-minute interval.

  2. Average the resulting per-lane averages across all days to get an overall average for that 5-minute interval. Divide the congestion frequency counter value by the eligible day counter, then multiply by 100 to get a percentage.

- Repeat the process for each 5-minute interval throughout a 24-hour day.

CDR Analyst Peak Hour, Peak Period, and Daily Volume Algorithms

CDR Analyst computes peak hour, peak period, and average annual daily traffic volumes using the AASHTO method, a process that uses minimum thresholds of data availability to produce an estimate of average annual daily traffic that samples across days of the week and months of the year. The process is similar to that used by CDR to produce analogous statistics, with the following exceptions: 1) CDR Analyst computes statistics for multi-lane sites (and therefore has additional data quality thresholds for multi-lane averaging), while CDR processes only one lane at a time; 2) CDR Analyst uses a different algorithm for user-specified data replacement of suspect values. The AASHTO procedure as implemented in CDR Analyst is as follows:

1. For each day, do the following steps:

- For each 5-minute interval, do the following steps:

  a. For each lane of traffic to be processed, do the following:

  Look at the data flag of a given 5-minute data volumes in a given lane in a given day, and perform data replacement if the user requests it and it is feasible (same as described in the core algorithm). (This is different from the CDR data replacement method, which only looks backward in time, and will go as far back as the previous midnight to get a valid replacement value.)

b. Average the resulting volumes of each lane to get a per-lane average for that 5-minute interval for that day. If less than 50 percent of the lanes have a good data value, flag that 5-minute average volume as "invalid."

- If at least 90 percent of the average 5-minute per-lane volumes of that day are considered "valid", sum up the 5-minute volumes for that day. Otherwise, flag the entire day as "invalid."

2. For each month, sum up and average the daily volumes by day of the week, e.g., for January, produce a Monday average, Tuesday average, etc. Skip any "invalid" days. If there is more than one "invalid" day of the week (e.g., an average day-of-week value cannot be calculated for the Mondays in that month), flag that month's volume for that day of the week as "invalid."

3. For each day of the week, average the 12 monthly day-of-week averages to produce a yearly day-of-week average. If there are more than two "invalid" monthly day-of-week values, flag that yearly day-of-week average as "invalid."

4. Average the five (if weekday statistic) or seven (if seven-day statistic) yearly day-of-week averages to produce the average annual daily traffic value. If there is more than one invalid yearly day-of-week average, an annual daily traffic volume will not be computed.

CDR Analyst also computes a value based on a direct average of all days with at least 90 percent good data, which can be used if an AASHTO-based value cannot be computed because of a threshold. Note that the direct average method may introduce some bias in the resulting annual estimate if the days used to compute the average are skewed toward weekdays or weekends, or a particularly time of the year (e.g., winter months).

AM and PM peak period values are computed the same way daily values are, except that they use two fixed subsets of data (6 to 9 AM and 3 to 7 PM, respectively). AM and PM peak hour values are determined by computing the one-hour period with the highest volume in their respective 12-hour periods (midnight to noon, noon to midnight) by using a moving one-hour window that increments at 5-minute intervals (i.e., 12:00 midnight to 1:00 AM, then 12:05 AM to 1:05 AM, etc.). This is based on the 24-hour average produced by the core process.

CDR Analyst Contour Algorithm

The 24-hour traffic profile that is computed for one site by CDR Analyst's core process (described earlier) can also be computed at a number of sites along a corridor. The resulting series of traffic site profile "slices" along the corridor can then be graphically depicted simultaneously in the form of a topographic-style contour map that shows average traffic conditions as a function of both time of day and corridor location. To create the process, CDR Analyst performs the following steps:

1. CDR Analyst starts with a user-specified batch file that contains a list of data file names. Each data file listed includes traffic data about one site on a corridor. The data file names are listed sequentially in the batch file by milepost location. CDR Analyst uses this batch file to produce a series of 24-hour profiles of average lane occupancy percentage, one at each site listed in the batch file.

2. CDR Analyst then takes the resulting profiles and prepares them to be put into an output file in a matrix format. Because the resulting matrix is then used in Microsoft Excel to create the contour map, and Excel's contour graphics option requires that the data points be equidistant (i.e., equal spacing along the corridor), it is first necessary for CDR Analyst to take the (usually) irregularly spaced profile slices and convert them to equal spacings. To do this, it

performs a linear interpolation between profiles to produce regularly spaced (at 0.5 mile increments) interpolated values. Each interpolated value is based on the profile data from sites that are closest to the location at which interpolation is being computed, using the formula

valueINT = valueA+[(milepostINT - milepostA)/(milepostB - milepostA)] * (valueB - valueA)

where

location A = location of closest data value on one side of the interpolated location

location B = location of closest data value on the other side of the interpolated location

valueA = data value (average lane occupancy percentage) at location A

valueB = data value (average lane occupancy percentage) at location B

valueINT = desired data value (average lane occupancy percentage) at interpolation location (0.5 mile spacing)

milepostA = milepost at location A

milepostB = milepost at location B

milepostINT= milepost at location of interpolation

For example, suppose that average lane occupancy profiles have been computed at a series of sites along a corridor at the following irregularly spaced mileposts:

Profiles at mileposts: 0.2 1.1 1.5 2.2 2.4 2.7 3.7

(irregularly spaced, actual data)

CDR Analyst computes linearly interpolated values at regular 0.5 mile spacings within this range from 0.2 miles to 3.7 miles, i.e., at 0.5, 1.0, ..., 3.5 miles. For each evenly spaced location, the closest actual data on each side is used to perform the linear interpolation. For example, to compute the interpolated value at milepost 0.5, the values at 0.2 and 1.1 are used (corresponds to points A and B in the formula above). The same two values are used to get the interpolated value at milepost 1.0, since 0.2 and 1.1 are still the closest data values. The results would be:

Profiles at mileposts: 0.2 1.1 1.5 2.2 2.4 2.7 3.7

(irregularly spaced, actual data)

Interpolation occurs at: 0.5 1.0 1.5 2.0 2.5 3.0 3.5

(regularly spaced, interpolated)

3. The resulting evenly spaced interpolated (or extrapolated) average lane occupancy profiles are sent to an output file in a matrix form (time vs. milepost location), where they can be prepared as a contour graph. Note that all site locations are based on the milepost of the associated field data collection cabinet (the equipment installed at a freeway site to collect

and process data from nearby sensors). Individual sensors do not have separate milepost values.

4. The resulting contour map graphic is color-coded by lane occupancy value. The range of values used for each color corresponds to different levels of traffic congestion, based on the level of service concept described in the Highway Capacity Manual for freeways with a freeflow speed of 65 mph. The following fixed ranges are used:

Color	Lane Occupancy %	LOS	General Traffic Description
green	0 to 10 percent	A, B, C	uninterrupted travel at the speed limit
yellow	10 to 13 percent	D	moderate traffic at or near the speed limit, with restricted movement (e.g., when changing lanes)
red	13 to 19 percent	E	traffic moving at or somewhat below the speed limit, with restricted movement
purple	above 19 percent	F	congested traffic with restricted movement

CDR Analyst Average Travel Time Profile Algorithm

A process similar to that used to develop contour maps can also be used to estimate average travel times from one point to another on a specific corridor using the following process:

1. For each day of the year (or whatever time period is being analyzed), average volumes and lane occupancy percentages are computed as a function of time of day and location along the corridor segment of interest and the travel direction of interest, using the core algorithm. Multiple lanes are combined into an average per-lane statistic during this process. Data replacement uses a ± 15 minute window to attempt to replace any data point that needs to be replaced (either because it is bad/disabled, or because it is suspect data and the user wishes to replace it.)

2. These values are used to develop speed estimates, also as a function of time of day and location, for each day of interest. The speed algorithm (described earlier) is used to develop these estimates.

3. At this point in the process, we have estimated speeds as a function of time of day at a series of measurement locations along the corridor for each day, as well as known distances between those measurement locations (based on the milepost of the associated data collection cabinet). These speeds and distances are combined to estimate travel times as a function of the trip start time, for each day. For example, suppose we want to know the travel time from location A to location E, and there are three additional measurement locations between those two locations (B, C, and D). We therefore have the following information:

Measurement Locations:	A (origin)		B		C		D		E (destination)
Estimated Speeds:	sA		sB		sC		sD		sE
Known Distances:		AB		BC		CD		DE	

Using this information, a travel time estimate can be developed for the trip by dividing the trip into segments by using each pair of adjacent measurement locations to define segment endpoints (A to B, B to C, etc.), computing the travel time for each segment, then adding the segment times together to get an overall trip time. The resulting travel time formula (developed by D. Dailey of the University of Washington[2]) assumes that speeds vary linearly from one measurement location to an adjacent measurement location; the approximate formula is:

Total Trip Travel Time = sum of individual segment travel times, where each segment is defined by successive sensor locations

Each segment time ≈ $2 * (\Delta x) * \{ S + [(\Delta s)^2/3 * S^3] + [\Delta s^4/5 * S^5] \}$

where the segment goes from measurement point i to point (i +1), and

s_i = speed at location i (one end of the segment)

s_{i+1} = speed at location i+1 (the other end of the segment)

Δx = segment length

$\Delta s = s_{i+1} - s_i$

$S = (s_{i+1} + s_i)^{-1}$

If a value is not available because it is determined to be invalid, then the program performs a linear interpolation across the missing data point, based on the two closest speed estimates on each side of the missing value, with the assumption that speeds vary linearly between the two points. In the previous example, if the speed at point C is missing, speeds at points B and D are used, using the assumption that speed between B and D varies linearly. If good data points are not available on both sides of the missing value (e.g., if an endpoint of the trip has a missing value), linear extrapolation is performed using the two closest values. If fewer than two valid data points are available along the route, then interpolation or extrapolation is not possible; therefore no travel time is computed for that trip start time on that day.

4. Because the speed estimates are computed as a function of time of day, travel times can thus be computed as a function of trip start time. For each trip start time (at 5 minute intervals throughout a 24 hour day), a corresponding travel time is thus computed.

5. Note that as the trip time is built up segment by segment, the most current speed estimate is used at each step along the way. For example, suppose that in the previous example (going from A to E), the trip starts at location A at 7 AM. The process of computing the travel time begins by estimating the travel time from A to B (the first segment). This is done by using the formula in step 3 along with the known distance from A to B and speed estimates for 7 AM at points A and B. If that estimated segment travel time exceeds five minutes (say, seven minutes), the segment travel time from B to C is computed by using the appropriate speed estimates at B and C not for 7 AM, but for 7:05 AM (since that is the most current estimate at location B, since it took 7 minutes to travel from A to B). This process continues, with each

[2] D. J. Dailey, *Travel Time Estimates Using a Series of Single Loop Volume and Occupancy Measurements*, Paper 970378, Transportation Research Board, 76th Annual Meeting, January 12-16, 1997, Washington, D.C.

segment time computed by using the most appropriate speed estimates, based on the cumulative travel time up to that point.

6. For each trip start time (at 5 minute intervals throughout a 24 hour day), the corresponding travel times for all the days in the analysis are averaged.

The result is a 24-hour travel time profile of the average travel time for a specific trip as a function of trip start time. Note that this process assumes that the trip occurs on only one corridor (i.e., the entire trip uses the same milepost numbering sequence, and there are no interchanges or ramps used during the trip).

CDR Analyst 90th Percentile Travel Time Profile Algorithm

The process to develop average travel times is extended to compute the so-called 90th percentile travel time as a function of trip start time. After computing the travel times for individual days, the number of travel time estimates corresponding to each 5-minute trip start time are counted, and the travel times are sorted in order of increasing duration. For each 5-minute time interval, starting from the shortest time and going up the list to the travel time that is 90 percent of the way to the longest trip time (i.e., 90 percent of the way to the end of the list), the corresponding travel time in the sorted list is the 90th percentile travel time. This is the time for which 90 percent of the trips with a given start time have a shorter trip travel time.

This value is computed for each 5-minute trip start time interval in a 24 hour time period. The result is a 24-hour travel time profile of 90th percentile travel time as a function of trip start time.

CDR Analyst Travel Time Reliability Algorithm

During the process of computing average travel times, a travel time is computed for each day; from this information, an overall average trip speed is also determined. These speeds are tabulated as a function of time of day to determine the percentage of trips that have average speeds less than 45 mph, as follows:

For each 5-minute interval, an "eligible day" counter is incremented once for each day in which a travel time can be computed for a trip starting at that 5-minute interval. For each day's trip whose average trip speed is less than 45 mph for a trip starting at that 5-minute interval, a separate "slow trip" counter is incremented once for this 5-minute interval. After all daily trips are processed, the "slow trip" counter value is divided by the "eligible day" counter, then multiplied by 100 to get a percentage of trips that are less than 45 mph based on the time period that was studied (alternatively, this can be thought of as the likelihood that a trip on a particular route starting at a particular time will have an average speed of less than 45 mph).

This percentage is recorded for each trip start time, and used for the superimposed histogram in the resulting travel time chart. The result is a 24-hour profile of travel time reliability as a function of trip start time.

CDR Auto

CDR Auto is a program developed by TRAC to re-format 5-minute traffic data stored on CD into files on a hard drive that can then be used by CDR Analyst. CDR Auto extracts information about mainline GP and HOV lanes for every day on each data CD.

Operating Instructions

This program is a pre-processor for CDR Analyst, and needs to be run only once for each year of data. The resulting output files (which can take up to 2 Gb/year on a hard drive) are then used in all subsequent runs of CDR Analyst. To process one year of data (2 to 4 CDs), do the following:

1. Double-click the program icon to start. The most recent version is Rev 1.0 (as of May 1999).

2. Enter the year of the data set that is being processed.

3. Enter the number range of cabinets to be processed. (See the CDR User's Guide in the Appendix of this report for more discussion of the cabinet numbers.) The entire CD will be processed for the cabinets selected.

4. Repeat this process for each CD in the year of interest. Data files will be appended. (This means that each CD does not generate brand new files; instead, data from subsequent CDs is appended to the file created when the first CD of the year was processed.)

Please read the CDR-Auto readme file for updated information on this utility.

CDR Analyst Utilities

CDR Analyst output is often processed to produce graphical output. The following utilities and templates are available to produce graphics.

Site Graphics

An Excel macro allows individual output files (.out extension) from the CDR Analyst site profile process to be displayed as a volume line graph and histogram. To access the file, open the macro file within Excel, then go to the Tools menu, select "Macro", and select and open the "histobat" macro.

The macro first asks for the name of a user-written command file (a text file) that lists the CDR Analyst output files to be processed into graphical form. The command file should have a ".bat" extension in its filename. The files listed in the command file should leave off the extension, which is assumed to be ".out." (The easiest way to produce the command file is to create a file in Excel, and enter the file names in column A.)

The macro then asks if histograms are to be produced. If the answer is yes, the graphs are produced. The macro color-codes the line graph based on the speed profile, using fixed color/speed ranges (red < 45 mph, yellow = 45 to 55 mph, green > 55 mph). The resulting output files have the same name as the input file, except with an ".hst" extension. Each file will contain the original input file in one sheet, and the resulting graph in another sheet.

The user can then specify whether graphs are to be printed out. If so, the graphs are printed out to the default printer. The user can also specify whether the headers of the output data files (summary information) are to be printed out as well. Note: If you answer No to the question about producing graphs, but Yes to the question about printing, the ".hst" version of each file in the command file will be printed out.

Corridor Graphics

An Excel template allows output from the CDR Analyst contour process to be displayed as a contour (topographic style) map. To use the template, copy the entire output data file from CDR Analyst into the “data” sheet of the template ; the “graph” sheet shows the result. In some cases, the milepost range will need to be modified. If significant changes are needed, it may be easier to use the Excel Wizard to create a contour map. Note that the final version of the corridor graphic with both directions of travel and corridor map art must be created manually, using a drawing program.

Travel Time Graphics

An Excel template allows output from the CDR Analyst travel time process to be displayed as a combination line graph/histogram figure. To use the template, copy the entire output data file from CDR Analyst into the “data” sheet of the template; the “graph” sheet shows the result. In some cases, the titles will need to be modified manually.

CD Data Quality Mapping Utility

CD-based traffic data can be pre-processed to evaluate the level of data quality at each site. An Excel macro accesses the traffic data files and tabulates the data quality on a cabinet by cabinet basis. This matrix of counts can then be analyzed to determine if sufficient “good” data exist. The macro processes a user-input command file that lists the data files to be processed.

The macro allows individual data files (.dat extension) created by CDR Auto to be analyzed. To access the macro, open the macro file within Excel, then go to the Tools menu, select “Macro”, and select open the “flag0123” macro. The macro asks for the name of a user-written command file (a text file) that lists the CDR Analyst output files to be processed into graphical form. The command file should have a “.txt” extension in its filename. The files listed in the command file should include the extension. (The easiest way to produce the command file is to create a file in Excel, and enter the file names in column A.)

The macro will then process each file, extracting and tabulating the number of data points for each cabinet that are in each of the four data quality categories: 0, 1, 2, and 3 (bad, good, suspect, and disabled, respectively). The result will be a matrix of such values, which can be reviewed to determine if sites (cabinets) of interest have sufficient valid data. The resulting output file uses the first eight characters of the last input file name in the command file, with an “.flg” extension. The data files should be in the same folder as the mapping macro.

Note that looking at the amount of good data alone may not be a sufficient measure of valid data, since suspect data can often be good data. If you plan to use data replacement of suspect data, you should of course consider the suspect data count as well. Note also that if data is organized by quarters, the output will be similarly organized.

4. Uses of the FLOW Evaluation Analysis Process

Introduction

In this section, three examples of the use of the evaluation process and tools described in this document are discussed. They include the production of a freeway usage and performance report that summarizes levels of use and performance for the central Puget Sound freeway network; a summary of traffic volumes at selected locations; and freeway system operations diagnostics. The following are descriptions of each example and the process by which the evaluation tool set described in section 3 can be used to produce the desired data.

Example A. Central Puget Sound Freeway Usage and Performance Report

In early 1999, the evaluation process and tool set described in this report was used to create the "Central Puget Sound Freeway Usage and Performance Interim Report" for WSDOT. This report summarized 1997 freeway usage at selected locations in the central Puget Sound freeway network, and also provided performance measures that included volume, speed, and congestion frequency at selected sites, as well as corridor-wide congestion patterns and travel time estimates. Variations in usage and performance as a function of year, lane type (GP or HOV), and weekdays vs. weekends were analyzed.

The following is an outline of that report, with comments about how the analysis was performed, which sections of this document are relevant to those processes, and implementation notes. (The reader is encouraged to refer to the usage and performance report while reading the following comments.)

System Usage

Analysis Types. Two measures of system usage were computed: average annual weekday vehicle volume, and average weekday peak period and peak hour vehicle volumes. Estimates of these measures were made at 13 selected locations in the central Puget Sound freeway network, on the major freeway corridors (see Table 1 for information about measurement locations). Locations were selected based on their traffic significance and the availability of usable data. The values used in this section of the interim report were computed using the CDR Analyst core algorithm described earlier. This was done by using CDR Analyst, with input data corresponding to the locations of interest for the traffic direction, lane type (GP/HOV) and time period (in this case 1997) of interest. The resulting output includes system usage statistics, as well as additional information used elsewhere in the interim report.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for computing these measures.

Note 1. The presence of reversible lanes at selected I-5 and I-90 locations used in the interim report requires particular care. Because the same sensors collect data 24 hours a day, it is necessary to separate the data corresponding to each direction of operation of the reversible lane(s). For example, even if the reversible lanes operate westbound from 1 AM to noon, and eastbound from 1 PM to midnight, the system usage statistics shown in the output from CDR Analyst will reflect the entire 24

Table 1. Reference Set of Measurement Locations

Corridor	Location	Traffic Considerations
I-5 (6 sites)		
GP/HOV	**S. 272nd**	Measures traffic between Seattle and South King and Pierce Co.
GP/HOV	**S. 170th**	Measures traffic between Seattle and South King and Pierce Co.
GP/HOV	**Albro**	Measures traffic from south end to Seattle CBD
GP	**NE 63rd**	Measures traffic between north end and U-District/Seattle CBD
GP/HOV	**NE 137th**	Measures traffic in the north end and Snohomish Co.
GP/HOV	**128th St. SW**	Measures traffic in the north end and Snohomish Co.
I-90 (2 site)		
GP/HOV	**Midspan**	Measures bridge use
GP/HOV	**161st Ave. SE**	Measures traffic east of I-405 interchange (near Eastgate)
SR520 (3 sites)		
GP	**76th Ave NE**	Measures bridge use
WB HOV	**84th Ave NE**	Measures bridge use
GP/HOV	**NE 60th**	Measures traffic east of I-405 interchange (near Redmond)
I-405 (4 sites)		
GP/HOV	**SE 52nd**	Measures traffic between Bellevue CBD and South King Co.
GP/HOV	**NE 14th St.**	Measures traffic in vicinity of Bellevue CBD
GP/HOV	**NE 85th**	Measures traffic between Bellevue CBD and North King Co.
GP/HOV	**Damson Rd.**	Measures traffic in Bothell/Woodinville area
SR167 (1 site)		
GP	**TBD**	Measures non-I-5 traffic between South King Co. and South Seattle/Eastside

hour period, i.e., both directions combined. It is therefore necessary to look at the detailed 24 hour profile information (which is included with the output), and separate the volumes into two categories corresponding to the two directions of operation, in order to get appropriate daily volumes by direction of operation.

Note 2. The daily system usage statistics are based on the AASHTO method (described in section 3). However, in many cases there may not be sufficient valid data to meet strict data quality standards. In those cases, daily usage based on a direct average of all days with sufficient "good" data is displayed.

Note 3. It is important to study the data to verify that there is sufficient "good" or valid data to develop a meaningful statistic at a given location. One way to do this is to look at the 24 hour profile data included with the statistics and in particular look at the number of good days that each lane contributed to volumes at each 5-minute interval. If the numbers are low relative to the total number of days analyzed (e.g., 261 weekdays in a year), it could suggest that the resulting statistic might not be accurate. A comparison of "good" day counts in each lane could also indicate whether an individual lane's sensor is having equipment difficulties. Ideally, you will want to verify that the site has sufficient good data prior to analysis. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the site of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

System Performance I: Freeway Corridors (Contours)

Analysis Types. General purpose freeway corridor performance was summarized in this section of the interim report, using average traffic congestion levels by time of day and location (contour maps). Estimates of these measures were made on I-5, I-405, SR 520, and I-90 in the Seattle area. The corridor analyses performed in this section of the interim report were computed using the corridor algorithms described earlier, namely the contour map option. This is done by using CDR Analyst, with input data corresponding to a series of locations along each corridor for each traffic direction, lane type (GP only in this case) and time period (in this case 1997) of interest. The resulting output includes a matrix of estimated congestion level as a function of time of day and location along the corridor, which is converted into a topographic-style contour map.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for corridor profiles, describes the process for computing these measures.

Note 1. It is important to study the distribution of the input data to verify that there is a sufficient distribution of valid data along a corridor to develop a meaningful contour map. Because the contour graph option automatically performs interpolation to fill in areas of missing data, the gaps in the data will not be immediately obvious from the output. It is therefore important to look at the geographical distribution of the initial data points being used along the corridor to identify locations with significant data gaps, and qualify the resulting contour map accordingly. For example, in the 1997 interim report, the south part of I-5 from Tukwila to Boeing Field was left blank in the contour map because of the absence of any valid data. The interpolated results in the north and south ends of I-405 were kept in the report maps, but the sparseness of data in those areas was noted in the text.

Note 2. It is important to study the input data to verify that there is sufficient "good" or valid data at each site used to produce the corridor contour. Insufficient valid data will not be immediately obvious from the output. It is therefore important to look at the data at each location to verify its validity. One way to do this is to look at the 24 hour profile data included with the statistics and in particular look at the number of good days that each lane contributed to volumes at each 5-minute interval. If the numbers are low relative to the total number of days analyzed, it could suggest that the resulting statistic might not be accurate. A comparison of "good" day counts at each lane could also indicate whether an individual lane's sensor is having equipment difficulties. Ideally, you will want to pre-filter the batch file used to create the contour map by verifying that each site listed in the batch file has sufficient good data. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the site of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

Note 3. Unlike some of the other graphics in the interim report, the contour map graphics were put together manually using several graphics from different sources. The two contours on each graph were produced separately using Excel templates; screen dumps of each contour were exported to a drawing program along with a site map. The resulting pieces were then merged together. Note that in some instances, corridor maps had to be "straightened out" or distorted to better match the linear milepost axis. In any case, the corridor map should be considered only a general guide to locations, and not a precisely calibrated scale.

System Performance I: Freeway Corridors (Travel Times)

Analysis Types. The same data used to produce contour maps can be used to estimate travel times. Three measures of trip-oriented general purpose freeway corridor performance were computed in this section of the interim report: average corridor travel times, 90th percentile corridor travel times, and average travel time reliability. Estimates of these measures were made on I-5, I-405, SR 520, and I-90 in the Seattle area. The travel time analyses performed in this section of the interim report were computed using the three travel time algorithms described earlier (average, 90th percentile, travel time reliability). This is done by using CDR Analyst, with input data corresponding to a series of locations along each corridor for each traffic direction, lane type (GP only in this case) and time period (in this case 1997) of interest. The resulting output is a function of trip start time.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for travel time profiles, describes the process for computing these measures.

Note 1. It is important to study the data to verify that there is sufficient "good" or valid data to develop a meaningful travel time along a corridor. Because the travel time option automatically performs interpolation or extrapolation to fill in areas of missing data, the gaps in the data will not be immediately obvious from the output. It is therefore important to look at the geographical distribution of the data points being used along the corridor to identify any areas with significant gaps, and qualify the resulting travel time profile accordingly. Ideally, you will want to pre-filter the batch file used to create the travel time profile by verifying that each site listed in the batch file has sufficient good data. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the site of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

Note 2. All travel time graphics in the interim report include a small map inset that was created in a separate drawing program, then pasted into the Excel graph.

Note 3. It is important that travel time measurements be based on routes that are significant from a traffic perspective. Routes should consider 1) major origin-destination patterns in the region, 2) major corridors or highways in the region, and 3) data collection locations (especially vehicle sensor locations).

Note 4. In the interim report, travel times were estimated for trips spanning the entire length of each of the major freeway corridors. In future evaluations, plans call for the use of specific trips. For example, the trips shown in Table 2 are a tentative list of such trips (subject to data availability).

Table 2. Potential Routes for Travel Time Estimation and Monitoring

Route #	Origin to Destination	Route Type	Primary Freeway Corridor(s)	Freeway Portion of Route: Start, End	Traffic Considerations/ Comments
Route 1	Lynnwood to Seattle CBD	Suburb to Seattle	I-5	164th SW Mercer St.	North-end traffic heading to Seattle on I-5
Route 1A	North Seattle to U-District	Suburb to Seattle	I-5	NE 117th NE 44th	North-end traffic heading to U-District on I-5; historical data available from previous evaluations
Route 2	Federal Way to Seattle CBD	Suburb To Seattle	I-5	S. 272th* Mercer St.	South-end traffic heading to Seattle on I-5
Route 3	Mtlk. Terrace to Bellevue CBD	Suburb to Suburb	I-405	I-5 interchg. NE 8th	North-end traffic heading to Bellevue CBD on I-405
Route 4	Tukwila to Bellevue CBD	Suburb to Suburb	I-405	I-5 interchg. NE 8th	South-end traffic heading to Bellevue CBD on I-405
Route 5	Bellevue to Seattle	Suburb to Suburb	SR520/I-5	NE 60th Mercer St.	Eastside-based traffic heading to Seattle on SR520
Route 6	Issaquah to Seattle	Suburb To Seattle	I-90/I-5	Front St. Mercer St.	Eastside-based traffic heading to Seattle on I-90
Route 7	Auburn to Renton	Suburb to Suburb	SR167	SR18** I-405 interchg.	South-end traffic staying in South end on SR167
Route 8A	Central Bellevue to Seattle CBD	Suburb to Seattle	I-405/ SR520/I-5	NE 8th NB Mercer St.	Central Bellevue resident commuting to Seattle; the SR520 option
Route 8B	Central Bellevue to Seattle CBD	Suburb to Seattle	I-405/ I-90/I-5	NE 8th SB Mercer St.	Central Bellevue resident commuting to Seattle; the I-90 option
Route 9	Redmond to Bellevue CBD	Suburb to Suburb	SR520/I-405	NE 60th NE 8th	Eastside (Redmond) traffic to Bellevue CBD
Route 10	Issaquah to Bellevue CBD	Suburb to Suburb	I-90/I-405	Front St. NE 8th	Eastside (Issaquah) traffic to Bellevue CBD

* The southernmost I-5 loop data that is available on-line is at S. 170th. Data from location(s) closer to S. 272nd will be used as they become available.

** The proposed starting location will be at or near the SR18 interchange. The exact starting location will be determined after evaluating recently-activated SR167 loop locations.

System Performance II: Selected Freeway Sites

Analysis Types. Three measures of freeway site performance were computed in this section of the interim report: average traffic volume profile at a site, by time of day; average speed profile at a site, by time of day; and average travel reliability at a site, by time of day. Estimates of these measures were made at a core set of 4 central freeway measurement locations in the Seattle area "rectangle" bounded by I-5, I-405, SR 520, and I-90 in the Seattle area, with one measurement location on each corridor. This was done by using CDR Analyst, with input data corresponding to each location for each traffic direction, lane type (GP, HOV, reversible) and time period (in this case 1997) of interest. The resulting output is a 24-hour traffic performance profile of each site.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for computing these performance measures.

Note 1. The presence of reversible lanes at the I-5 and I-90 locations used in the interim report require particular care. Because the vehicle sensors collect data 24 hours a day, it is necessary to separate the data corresponding to each direction of operation of the reversible lane(s). For example, if the reversible lanes operate westbound from 1 AM to noon, and eastbound from 1 PM to midnight, the resulting profile graph should note this fact. It is therefore necessary to look at the detailed 24 hour profile and separate the volume profiles into two sections corresponding to the two directions of operation, in order to get an appropriate traffic profile graph for each traffic direction.

Note 2. It is important to study the data to verify that there is sufficient "good" or valid data to develop a meaningful statistic at a given location. One way to do this is to look at the 24 hour profile data included with the statistics and in particular look at the number of good days that each lane contributed to volumes at each 5-minute interval. If the numbers are low relative to the total number of days analyzed (e.g., 261 weekdays in a year), it could suggest that the resulting statistic might not be accurate. A comparison of "good" day counts at each lane could also indicate whether an individual lane's sensor is having equipment difficulties. Ideally, you will want to verify that the sites have sufficient good data prior to analysis. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the sites of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

System Performance III: Performance Variations

Analysis Types. The following measures of freeway site performance variations were computed in this section of the interim report: 1995 vs. 1997 daily vehicle volumes; 1995 and 1997 weekday vs. weekend daily volumes; and 1995 and 1997 weekday vs. weekend average traffic volume profiles, by time of day and lane type (GP, HOV). Estimates of these measures were made at a core set of 4 central freeway measurement locations in the Seattle area "rectangle" bounded by I-5, I-405, SR 520, and I-90 in the Seattle area, with one measurement location on each corridor. This was done by using CDR Analyst, with input data corresponding to each location for each traffic direction, lane type (GP, HOV) and time period (1995 and 1997) of interest. The resulting output is a 24-hour traffic performance profile of each site, as well as daily volume statistics for each combination of year, lane type, and traffic direction.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for computing these performance measures.

Note 1. The presence of reversible lanes at the I-5 and I-90 locations used in the interim report require particular care. It is necessary to recognize each direction of operation of the reversible lane(s). For example, if the reversible lanes operate westbound from 1 AM to noon, and eastbound from 1 PM to midnight, the resulting profile graph should note this fact. It is therefore necessary to look at the detailed

24 hour profile and separate the volume profiles into two sections corresponding to the two directions of operation, in order to get an appropriate traffic profile graph for each traffic direction.

Also, because the sensors collect data 24 hours a day, it is necessary to separate the data corresponding to each direction of operation of the reversible lane(s) before computing summary statistics. For example, even if the reversible lanes operate westbound from 1 AM to noon, and eastbound from 1 PM to midnight, the daily system usage statistics shown in the output from CDR Analyst will reflect the entire 24 hour period, i.e., both direction combined. It is therefore necessary to look at the detailed 24 hour profile information (which is included with the output), and separate the volumes into two categories corresponding to the two directions of operation, in order to get appropriate daily volumes by direction of travel.

Note 2. It is important to study the data to verify that there is sufficient "good" or valid data to develop a meaningful statistic at a given location. One way to do this is to look at the 24 hour profile data included with the statistics and in particular look at the number of good days that each lane contributed to volumes at each 5-minute interval. If the numbers are low relative to the total number of days analyzed (e.g., 261 weekdays in a year), it could suggest that the resulting statistic might not be accurate. A comparison of "good" day counts at each lane could also indicate whether an individual lane's sensor is having equipment difficulties. Ideally, you will want to verify that the sites have sufficient good data prior to analysis. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the sites of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

Note 3. The daily system usage statistics are based on the AASHTO method (described in section 3). However, in some cases there may not be sufficient valid data to meet strict data quality standards. In those cases, daily usage based on a direct average of all days with sufficient "good" data is also displayed.

Note 4. The graph of 1997 vs. 1995 change in travel reliability was computed by taking each year's congestion frequency data and subtracting one from the other in a separate spreadsheet.

HOV Lane Network

Analysis Types. The number of vehicles traveling on GP and HOV lanes at selected locations was combined with available data about the average number of persons per vehicle (vehicle occupancy) to compare the number of people using GP and HOV lanes at selected sites on major corridors during the peak period. Three vehicle categories were analyzed: passenger cars, vans, and transit buses. For the cars and vans, the number of vehicles of each type at the site of interest was determined by multiplying the overall vehicle volumes (from CDR Analyst) by the corresponding mode split data from the WSDOT HOV Lane Evaluation project (i.e., the percentage of all vehicles traveling by a site that are of a particular vehicle type). The number of persons traveling past the site in a car or van was then computed by multiplying the number of vehicles of that type by the average number of passengers per vehicle, as determined from research project or transit agency sources (see Note 3). Average bus ridership is obtained from transit agencies and does not need to be computed; however, the bus ridership is adjusted to reflect the actual percentage of buses traveling on that type of lane (GP or HOV). The three resulting person volumes by vehicle type are then added together to produce an overall person volume estimate.

Comparisons of GP and HOV volumes at selected locations on an example corridor were also presented. Average traffic volume profiles were computed at selected sites on I-405, by time of day and lane type (GP, HOV) during November 1998 (in September 1998 the HOV lane was moved from the outside to inside lane; the November time period was chosen to reflect the new HOV lane placement). The traffic volume profiles were computed by using CDR Analyst, with input data corresponding to each location for each traffic direction and lane type (GP, HOV) of interest. The resulting output is a 24-hour traffic performance profile of each site, as well as daily volume statistics for each combination of lane type and traffic direction.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for computing the vehicle volumes. The peak period person volume formula is as follows:

Total persons carried = (# of persons in cars) + (# of transit riders) + (# of van riders)

= (Total Veh. Vol.) * (% Car)*(ACO) + (Bus Ridership) * (HOV/GP Bus Dist.)

+ (Total Veh. Vol.) * (% Van) * (VanOcc)

where

Total Veh. Vol.	Total number of vehicles traveling at that site (for a given direction of travel and lane type)
%Car, %Van	Percentage of vehicles that are cars or vans traveling at that site (for a given direction of travel and lane type)
Bus Ridership	Total number of peak period riders at a site traveling in a given direction on a given lane type.
ACO	Average car occupancy at that site (weighted average occupancy of 1, 2, 3 and 4+ person cars)
HOV/GP Bus Dist.	Percentage of buses that travel in a given lane type at that site (GP or HOV)
VanOcc	Average van occupancy at that site (including driver)

The data sources are:

Total Vehicle Volume	CDR Data (1997 yearly peak period average)
% Car, % Van	ACO Data from HOV Lane Evaluation Phase IV Report (Q2 and Q3 1997 data)
Average Car Occupancy	HOV Lane Evaluation Phase IV Report (Q2 and Q3 1997 data)
Bus Ridership	Transit Agencies (1995 peak period data)
Bus Dist	ACO Data from Phase IV Report (Q2 and Q3 1997 data)
Average Van Occupancy	King County Metro Rideshare Operations Performance (9.74 for 1996 including driver, 9.58 for 1997, 9.27 for 1998)

Note 1. In the interim report, person volumes were estimated only for selected locations in the central Puget Sound region. In the future, person volumes are tentatively scheduled to be estimated for as many of the usage sites listed in Table 1 as possible, subject to data availability. Table 3 is an example of potential data source information that was collected in 1996 to estimate person volumes. These data

Table 3. Example of Supporting Data Location Table for Reference Measurement Sites (1997)

	Volume, Lane Occupancy		Average Car Occupancy (ACO)**		Average Bus Occupancy	Comments
Corridor	Location	Cabinet #	Location	Site #	Location	
I-5						
GP/HOV	**S. 272nd**	TBD	**S. 216/70th E.**	34 or 92	**TBD**	no loops
GP/HOV	**S. 170th**	059D	**S. 216th**	34	**S. 170th**	no SB HOV
GP/HOV	**Albro**	086D	**Albro**	25	**Albro**	
GP	**NE 63rd**	143D	**NE Northgate**	16	**NE 63rd**	
GP/HOV	**NE 137th**	165D	**NE 145th**	14	**NE 137th**	
GP/HOV	**128th St. SW**	213D	**NE 145th**	14	**TBD**	
I-90						
EB GP	**Midspan**	858D	**Island Crest**	54	**Midspan**	
WB GP	**Midspan**	857D	**Island Crest**	54	**Midspan**	
HOV	**Midspan**	857D	**Lk Wa. Blvd.**	52	**Midspan**	
GP/HOV	**161st Ave SE**	910D	**Newport Wy.**	57	**TBD**	
SR520						
EB GP	**76th Ave NE**	514D	**92nd Ave NE**	42	**Midspan**	
WB GP	**76th Ave NE**	514D	**92nd Ave NE**	42	**Midspan**	
WB HOV	**84th Ave NE**	516R	**92nd Ave NE**	42	**Midspan**	
GP/HOV	**NE 60th**	544D	**148th Ave NE**	45	**TBD**	
I-405						
GP/HOV	**SE 52nd**	662D	**112th Ave. SE**	65	**SE 52nd**	
GP/HOV	**NE 14th St.**	696D	**NE 4th St.**	73b	**NE 14th St.**	
GP/HOV	**NE 85th**	716/717	**NE 85th**	81	**NE 85th**	
GP/HOV	**Damson Rd.**	761D	**NE 85th**	81	**TBD**	
SR167						
GP	**TBD**	**TBD**	**S. 208th**	98	**S. 208th**	loops TBD

Data sources in this list were based on information collected during 1996. This entire list should be updated prior to future evaluations.

locations are based on measurement data being collected as of 1996, and are likely to change by the time future evaluations take place. Nevertheless, this table illustrates the importance of evaluating data availability as well as the proximity of data collection locations associated with a particular site (i.e., are volume and vehicle occupancy measurements for a given site collected near enough to one another).

Note 2. The daily system usage statistics are based on the AASHTO method (described in section 3). However, in some cases there may not be sufficient valid data to meet strict data quality standards. In those cases, daily usage based on a direct average of all days with sufficient "good" data is also displayed.

Note 3. The passenger vehicle occupancy data were obtained from the WSDOT HOV Lane Evaluation project's field measurements, which also include mode split data for major vehicle types (what percentage of all vehicles at a site are cars, buses, vans, etc.). The HOV Lane Evaluation project's web site is <**http://www.wsdot.wa.gov/eesc/atb/atb/hov/Titlepg.html**>. Transit and van per-vehicle ridership data were provided by local transit agencies.

Note 4. It is important to study the data to verify that there is sufficient "good" or valid data to develop a meaningful statistic at a given location. One way to do this is to look at the 24 hour profile data included with the statistics and in particular look at the number of good days that each lane contributed to volumes at each 5-minute interval. If the numbers are low relative to the total number of days analyzed (e.g., 261 weekdays in a year), it could suggest that the resulting statistic might not be accurate. A comparison of "good" day counts at each lane could also indicate whether an individual lane's sensor is having equipment difficulties. Ideally, you will want to verify that the sites have sufficient good volume data prior to analysis. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the sites of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

Example B. WSDOT Northwest Region Ramp and Roadway Traffic Volume Report

The Traffic Systems Management Center (TSMC) of WSDOT Northwest Region produces a Ramp and Roadway Traffic Volume Report approximately every two years. This report includes average weekday volumes as well as AM and PM peak hour volumes at selected ramp and mainline (GP, HOV, and reversible) locations on freeways throughout the Northwest Region. The traffic counts are taken during several months of the year using a combination of electronic sensors (inductance loops) and portable tube counters. Weekday volumes are based on Tuesday through Thursday counts. Peak hour volumes represent the highest one hour volume during the AM or PM. The report also lists the one-hour peak hour time period if it is outside the fixed peak periods of 6 AM to 9 AM and 2:30 PM to 6 PM.

The evaluation process and tool set described in this report can be used to produce some of the data for this report. The following describes the portions of the ramp and roadway traffic volume report analysis that could be performed using the tool set. (The reader is encouraged to refer to the ramp and roadway report while reading the following comments.)

<u>Daily and Peak Hour Weekday Volumes</u>

Analysis Types. Two measures of system usage are summarized in the ramp and roadway report: average weekday vehicle volume, and average peak hour vehicle volume. Estimates of these two measures were made at locations throughout the freeway network included in the Northwest Region. Highways included in the report were I-5, SR 18, SR 167, I-405, SR 518, SR 520, and I-90. These two usage measures can be computed by the FLOW evaluation tool set, provided that electronic sensor data is available. This is done by using CDR Analyst, with input data corresponding to the locations of interest for the traffic direction, lane type (GP/HOV/reversible) and time period of interest. The resulting output includes summary usage statistics, as well as a more detailed 24-hour volume profile at 5-minute intervals throughout an average day.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for computing these usage statistics.

Note 1. The tool set described in this document operates using archived electronic sensor data only. Therefore, it can only provide statistics for sites with sensor installations. Locations in the report that do not have corresponding or nearby sensors will not be able to use this method.

Note 2. The sensor sites might not always correspond precisely with locations used in past ramp and roadway traffic volume reports.

Note 3. The presence of reversible lanes at selected locations requires particular care. Because the sensors collect data 24 hours a day, it is necessary to separate the data corresponding to each direction of operation of the reversible lane(s). For example, even if the reversible lanes operate westbound from 1 AM to noon, and eastbound from 1 PM to midnight, the system usage statistics shown in the output from CDR Analyst will reflect the entire 24 hour period, i.e., both direction combined. It is therefore necessary to look at the detailed 24 hour profile information (which is included with the summary statistic output), and separate the volumes into two categories corresponding to the two directions of operation, in order to get appropriate daily or peak hour volumes by direction of travel.

Note 4. In most cases, the normal hours of each direction of reversible lane operation correspond well with the midnight-to-noon, noon-to-midnight time periods used for AM and PM peak hour statistics respectively, so the standard CDR Analyst peak hour statistics output will still be valid. Peak hour volumes reported by CDR Analyst should be reviewed to verify that they take the previous caveat about reversible lane operation into account, however.

Note 5. The statistics used in the report are based on Tuesday through Thursday for selected months (e.g., March, April, May). In contrast, CDR Analyst offers the option to choose any time period up to the entire year, if data is available. While the use of 12 months of data might be considered a more appropriate measure in some situations, it might also be worthwhile to consider using the same data collection periods of previous ramp and roadway reports to maintain year-to-year consistency of results over time. Reconstruction of a previous set of results using CDR Analyst, then comparing the original and Analyst-based sets of values to determine seasonal variations, might be useful in this regard. The user has the option to create data files using CDR that contain only Tuesday through Thursday data (see section 3, Overview: Selecting data to be processed by CDR Analyst).

Note 6. Ideally, you will want to verify that the sites of interest have sufficient good volume data, prior to analysis. One way to do this is to use the data quality mapping option to get an overview of the quality of available data at the sites of interest (the mapping option is described later in this section as Example C, System Operations Diagnostics).

Example C. System Operations Diagnostics

In addition to estimating freeway usage and performance, the tool set described in this report can also provide diagnostic information about the operation of traffic systems management and data collection field installations, and can contribute to a better understanding of the validity of summary statistics. In the process of developing the usage and performance measures for the central Puget Sound Freeway Usage and Performance report, this diagnostic capability was used in three different situations. The following is a description of each situation, and the methods used to diagnose the problem in each case.

Data Quality Mapping

Analysis Types. The analysis performed by CDR Analyst is of course highly dependent on the quality of the original sensor data that is being processed. The ability to evaluate the level of quality prior to an analysis can save time and enhance the efficiency of the analytical process. For example, in the process of developing site and corridor analyses in the freeway usage and performance report, it became clear that in some situations, there were significant quantities of suspect or invalid data that affected the process of computing measures of performance. In some cases, data quality problems could be anticipated by referring to available information about construction projects that were likely to disrupt data collection devices during the time period of interest. In other cases, however, there was not always a prior indication that sufficient valid data might not be available. In those situations, time was spent setting up and running analyses, only to discover that the results were either of questionable quality because of insufficient "good" data, or were clearly unusable. Initially, this problem was dealt with by performing random "spot checks" of the data set prior to analysis. This was tedious and often inconclusive. To better address this issue, a standalone spreadsheet-based prototype tool was developed to analyze the traffic data files (output files from the CDR Auto program) and compute the percentages of "good", "suspect", "bad", and "disabled" data at each measurement site, on a quarter by quarter basis throughout the year. This information was then summarized in graphical form. This summary was useful in determining which locations were likely to be problematic from a statistics computation point of view, as well as determining which sites should be skipped altogether when developing contour maps or computing travel times. Figure 7 shows an example of the mapping output.

Relevant Sections of this Document. Section 3, CDR Data Quality Mapping Utility operating instructions, describes the process for running this program.

Note 1. The tool set described in this document summarizes the amount of data that is tagged as "good." However, the user must still determine the minimum amount of "good" data that is considered acceptable for the type of analysis being performed. This minimum threshold should take into account not only the overall amount of valid data, but its temporal distribution. For example, a minimum data quality standard of at least 50% good data for the entire year does not take into account the fact that such a standard could be met by having 50 percent good data uniformly throughout the year, or having 100 percent good data in two quarters of the year and 0 percent good data in the other two quarters; in the latter situation, the resulting average yearly statistics might not fully reflect seasonal variations. The precise acceptance criteria is a user decision.

Note 2. The data replacement option in CDR and CDR Analyst could also affect the resulting analysis. For example, if the user chooses to use data replacement, transient suspect or bad data points that are surrounded by good values will be replaced by nearby good data. Extended periods of suspect or bad data, however, might not be replaced if the data replacement window is small (e.g., default = ± 15 minutes).

Note 3. It might be appropriate to consider data quality categories besides "good" data. For example, the standard for data tagged as "suspect" is conservative. As noted earlier in this report, a given 5-minute data value is tagged as suspect if even one of the 15 20-second values that make up that 5-minute value is considered to be unrealistic. In a number of situations during the data analysis process for the usage and performance report, a value labeled as "suspect" was consistent with the "good" values that preceded and followed it, suggesting that the value was likely to be acceptable. This possibility is the reason why CDR and CDR Analyst both offer the option of accepting suspect values as valid data.

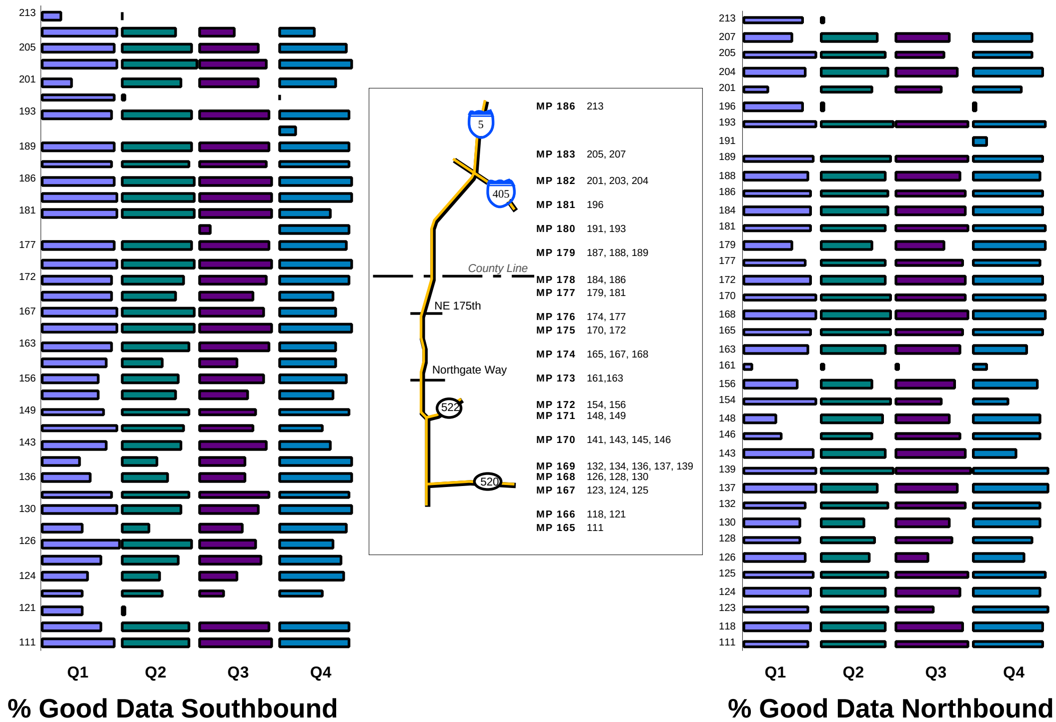


Figure 7. Sensor Data Quality By Location and Time of Year

I-5 North of Seattle (1977)

This should be taken into account when determining the user's minimum threshold for valid data. For example, rather than have a threshold based on the percentage of good data alone, it may be useful to consider a combination of good and suspect data in the threshold computations. NOTE: A user's decision to assume that suspect data is valid when determining the data quality of a site should be linked with the user's response to the CDR/CDR Analyst option to accept suspect data as valid. If the user chooses to accept suspect data as valid, it would then be appropriate that the combined amount of good and suspect data be considered by the user when evaluating data quality thresholds.

Note 4. The data quality tool provides statistics for the time intervals that are reflected in each data file. If there is one quarter of data per file, the data quality tool will compute quarter by quarter data quality information. If one file contains a year's worth of data, yearly data quality totals are provided.

Electronic Sensors I

Analysis Types. The ability to see detailed time-of-day volume patterns from individual lane sensors using CDR Analyst can be useful in spotting potential equipment problems. For example, in the process of developing GP vs. HOV analyses in the freeway usage and performance report, the data at one location on I-405 suggested that HOV volumes were about the same as those of the GP lanes (see Figure 8). Although HOV volumes on this corridor are significant, it seemed unlikely that the per-lane vehicle volume of an HOV lane would equal the average per-lane vehicle volume of the two GP lanes. After studying the data being collected by each lane individually in a one-day sample at that site, and comparing the 24-hour profiles to those of nearby sites, it appeared that one of the GP lanes had been inadvertently wired as if it was an HOV lane, and the HOV lane sensor had been wired as a GP lane (see Figure 9). As a result, the HOV lane statistics were actually reporting GP lane performance, while the GP per-lane performance was being reduced because it was averaging in one lane of HOV lane volumes, thus accounting for the HOV lane's apparent high volumes relative to the average GP lane at that site. In this case, CDR Analyst was able to diagnose a potential equipment installation issue.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for obtaining 24-hour profiles.

Note 1. CDR Auto creates input files for CDR Analyst that include all the lanes of a given direction and type at each site. To obtain individual lane data, CDR should be run for each lane separately to produce the input data files for CDR Analyst.

Electronic Sensors II

Analysis Types. CDR Analyst can also help determine the validity of summary statistics. For example, in the process of developing peak hour and daily volumes in the freeway usage and performance report, data at the midspan on I-90 suggested that reversible lane volumes were significantly lower at the midspan than they were at other nearby reversible lane locations. This transient drop in volume did not appear in the volume estimates reported in the ramp and roadway volume report, suggesting that a small study should be performed to determine possible explanations for the difference in values. Several hypotheses were considered:

- Because the CDR Analyst results were for 1997, but the ramp and roadway results used for comparison were for 1996 and 1998 (1997 results were not available), there was a possibility that the 1997 drop in reversible lane midspan volume was real. While it seemed unlikely that there would be a transient dip in 1997 midspan volumes given the general upward trend in other years, this possibility was checked by re-running the CDR Analyst data for 1998, and comparing the results to the same year's ramp and roadway values. Furthermore, because the CDR Analyst results were based on a year of weekday data, while the ramp and roadway values were based on selected measurements from a four-month period (March through June

1998, Tuesdays through Thursdays) there was a possibility that seasonal variations might account for at least some of the difference. So, CDR Analyst data was collected for a comparable time period (March through June 1998, Wednesday values) to reduce this possibility.

Figure 8. Estimated Weekday Volume Profile: GP and HOV Lanes

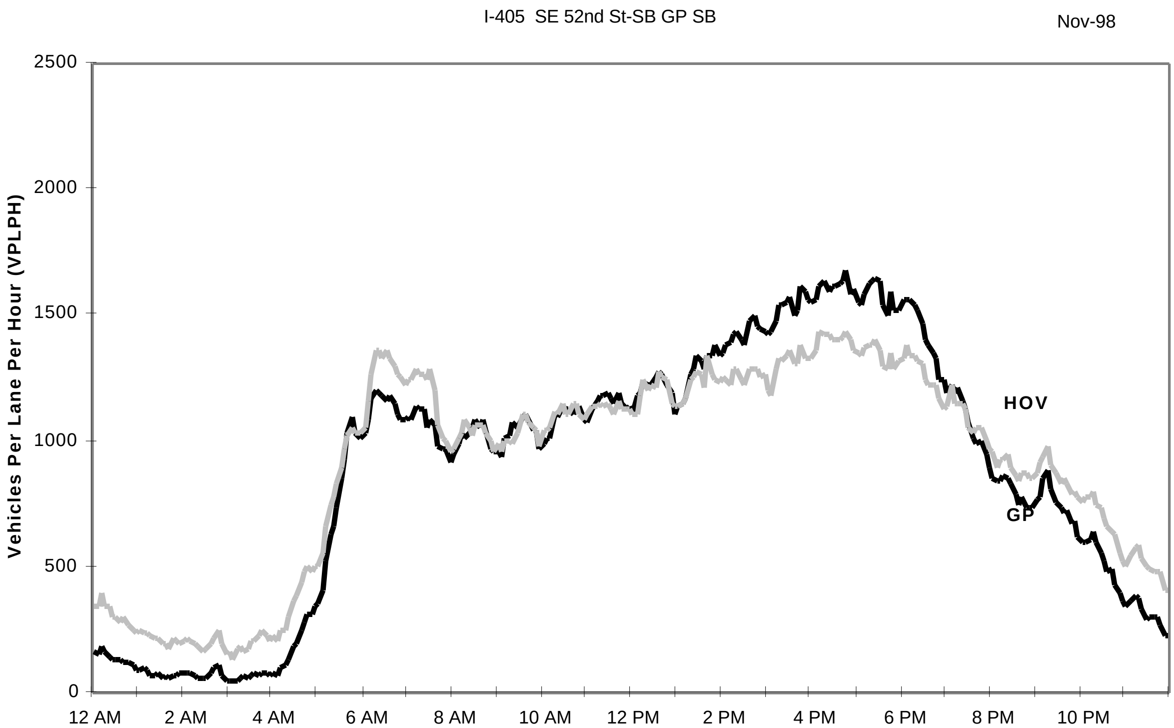
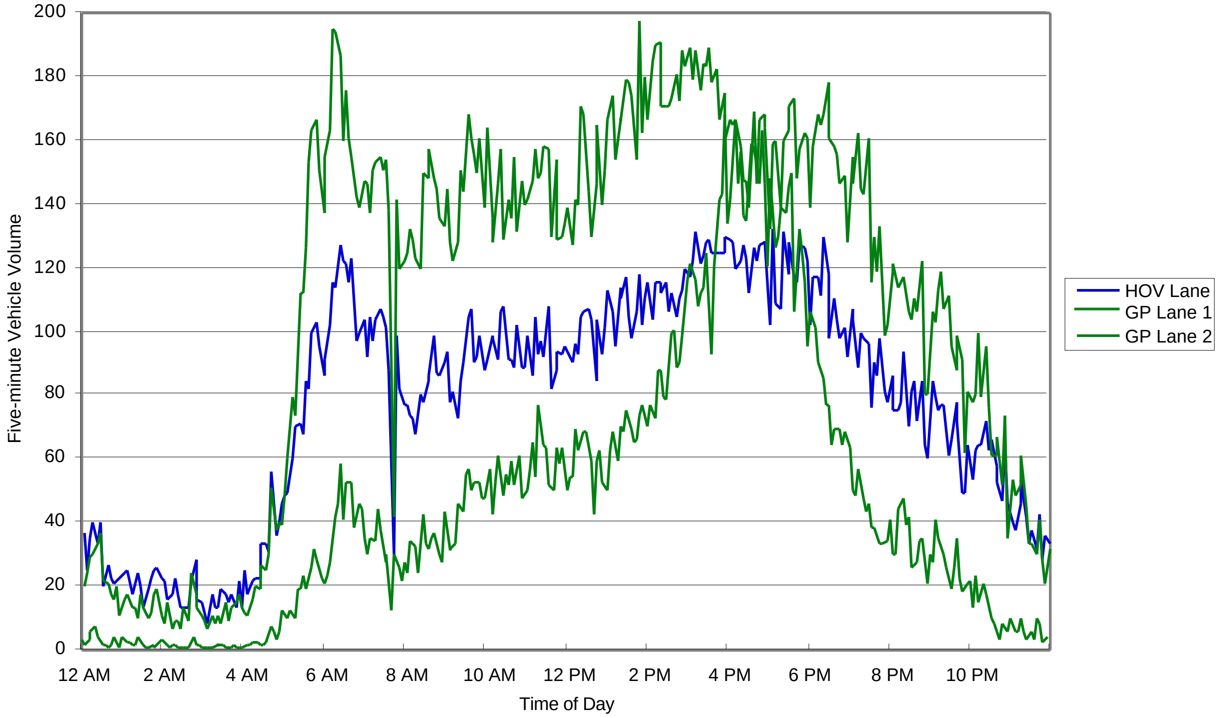


Figure 9. Per-Lane Volume Profile: GP and HOV Lanes

I-405 SE 52nd St Southbound (November 11, 1998)



- The CDR Analyst process might be systematically underestimating volumes, or the ramp and roadway volumes might be systematically overestimating volumes. Although measurements at other locations did not suggest that either was the case, the possibility was tested by using CDR Analyst not just at the midspan of I-90 but along a series of other locations along the I-90 corridor from Seattle east toward Bellevue as well, and comparing the results to ramp and roadway volumes.

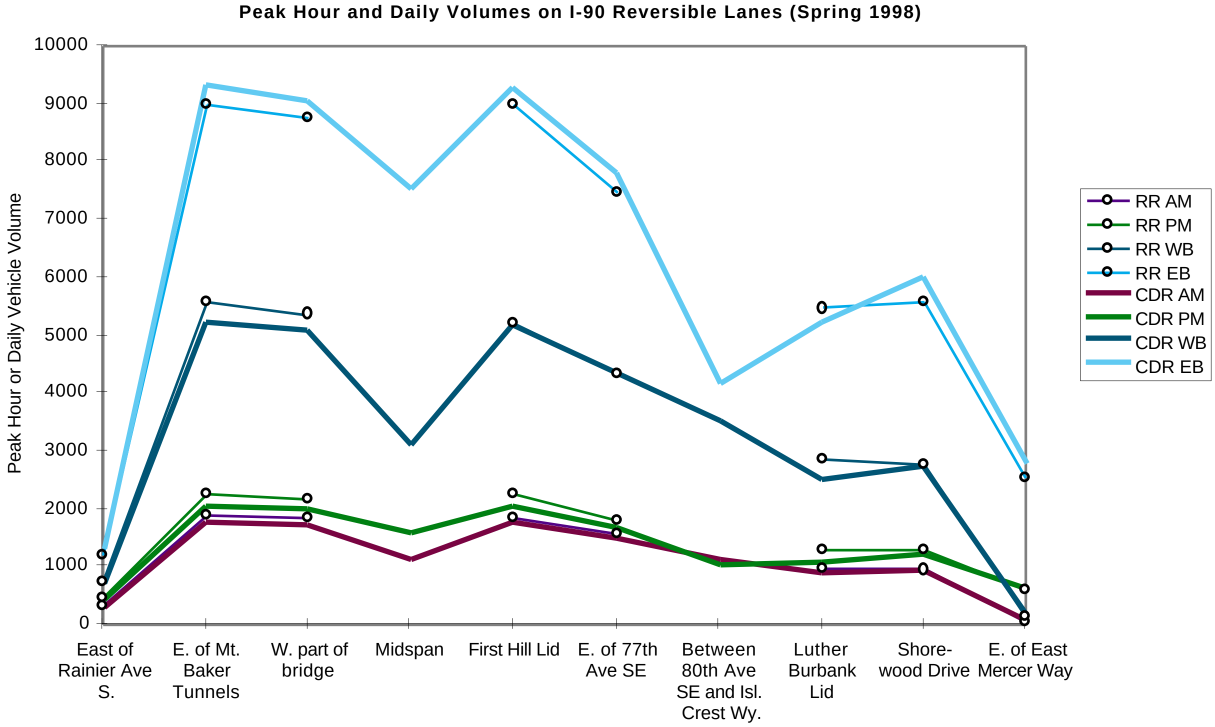
    - There was a possibility of an equipment problem at the midspan that was producing an underestimate at that one location. To study this issue, a data quality check was performed to study the quantity of valid data (see earlier example, "Data Quality Mapping").

The results of a series of comparisons along the I-90 corridor show that with the exception of the midspan location, other measurement sites show good agreement in the two sets of volumes, thus suggesting that the difference at the midspan is not due to seasonal or year-to-year fluctuations (see Figure 10). Furthermore, the agreement between the two sets of values did not support the idea that either process was universally underestimating or overestimating volumes (although since both used loop data, this was inconclusive). Looking specifically at the midspan values, the noticeable drop in volume at the midspan measurement location does not seem likely given that the measurement site just to the west, near the west highrise of the bridge, has a much higher volume even though there are apparently no on- or off-ramps between the two measurement sites. Furthermore, the overall trend of volumes at successive locations along the corridor also supports the notion that this sudden drop in volume is not correct. A data quality study of data cabinets along the corridor showed that although the results from the cabinet (857) associated with the midspan measurements had a high percentage of "good" data, adjacent cabinets 852, 854, and 855, all had moderate to significant quantities of "suspect" or bad data. Another clue is that cabinet 854, which has a smaller percentage of good data, is located at the same milepost as midspan cabinet 857. Finally, the ramp and roadway report indicates that midspan values are "unavailable" for 1998, suggesting that there might have been a data collection problem at that location. Taken together, these results, while not conclusive, are all clues that suggest that the CDR Analyst results should not be taken at face value at this site, and that a more detailed equipment study might be useful.

Relevant Sections of this Document. Section 3, CDR Analyst operating instructions for site profiles, describes the process for obtaining 24-hour profiles.

Note 1. As noted above, CDR Analyst and the ramp and roadway report both use loop data; agreement of their respective results does not necessarily mean that both sets of results are correct. However, the use of cumulative data from a number of sources (ramp and roadway data, data quality maps, detailed CDR Analyst results), combined with user experience, can help focus diagnostic activities and suggest likely possibilities.

Figure 10. Comparison of CDR Analyst and Ramp and Roadway Report Estimates



5. Analytical Assumptions and Limitations

Overview

In order to develop an evaluation process that can synthesize a large data set in a timely manner and also produce performance measures that are succinct and understandable, some assumptions were made regarding the algorithms that are used, and limits were set on the capabilities and intended purpose of the evaluation method and the accompanying tool set. This section presents a summary checklist of considerations that the prospective user should keep in mind when producing, modifying, or interpreting analytical results based on the FLOW evaluation method and tool set.

The user is strongly encouraged to review the discussion in this section as well as comments in the preceding two sections of this document regarding the assumptions or limitations of specific aspects of the evaluation process or the analytical tools, prior to using the FLOW evaluation tools or interpreting its analytical results.

Input Data Considerations

The FLOW evaluation process starts with and is dependent upon the 5-minute CD-based traffic data. The following considerations should be kept in mind when designing an evaluation effort that uses this data set:

1. **Sensor availability at the location(s) <u>and</u> time periods of interest.**

   The traffic data is recorded at approximately 0.5-mile intervals along the major freeway facilities. If the analysis must take place at a precise location, it is important to note whether an available sensor location can be placed precisely enough for the desired analysis.

   Note also that a measurement site's milepost value is based on the location of the associated data cabinet. Furthermore, a single field data cabinet collects information from multiple sensors. Therefore, the milepost value is only approximate; it is not necessarily the exact location of an individual vehicle sensor.

2. **Data availability at the location(s) <u>and</u> time periods of interest.**

   Even when a sensor is available at a desired location, data availability can be disrupted by equipment problems or construction activity. It is important to evaluate the data that would be used in an analysis to determine whether sufficient valid data is available during the time periods of interest.

3. **Data quality at the location(s) <u>and</u> time periods of interest.**

   Each archived 5-minute data value has an associated data quality flag that summarizes the data quality of its constituent 15 20-second values. The 5-minute data validity codes include "good" (all 15 constituent values are considered valid), "bad" (all 15 values are considered invalid), "disabled" (all 15 values were collected when the data collection equipment at that sensor site was disabled or not operational), or "suspect" (all other combinations of 15 data point conditions). Data validity for each 20-second value is determined by checking the equipment status as well as the nature of the values themselves. If a 20-second data point indicates that the loop is hung (i.e., sensor is stuck in the "on" position for a longer than reasonable time period), or consists of a volume and lane occupancy combination that is outside of an envelope of "reasonable" values (based on earlier research), the data point is considered bad or invalid. (The thinking behind these thresholds is that a prolonged "on" condition by the sensor or a highly unusual volume/occupancy value combination is symptomatic of equipment difficulties.) If the operator has disabled the operation of the sensor, the 20-second data point is labeled disabled. In all other cases, the data point is labeled good.

   The 5-minute data flags are indicators that should be used to evaluate the overall quality of the prospective data set. The user then has the option to use the data set as is, to reject a data set, or to use the data set but attempt to "fix" data that is labeled as other than good (see next item).

4. **Data quality threshold for "good" and "suspect" data.**

   The user has the option to specify the minimum amount of "good" data that is required to complete a computation. There is no absolute guideline for the minimum amount of good data that is required in order to develop a performance measure that one can feel confident about. However, a data quality map of the prospective data set will indicate whether a significant quantity of good data is available. (See section 4, Example III: Data Quality Mapping for additional details.) The user also needs to specify if "suspect" data is assumed to be good. The user's choice for this option can also affect the level of comfort that the user has in the analysis results.

   As an example, a minimum threshold of 90 percent good data was used for the central Puget Sound Freeway Usage and Performance report, while suspect data was always replaced by CDR Analyst whenever feasible during the analysis for that report, with the data replacement window set to ± 15 minutes.

5. **Data resolution for the analysis to be performed.**

   The data is archived at 5-minute intervals. This is sufficient for many types of analyses; however, for detailed studies that focus on short-term or highly variable traffic measurements, 5-minute data might not be appropriate. For example, some types of ramp analyses might fit this category.

6. **Time period for the analysis to be performed.**

   It is important to select an appropriate time period for the analysis. This includes the day-of-week selections (weekday vs. weekend vs. seven day) as well as the overall time period (week, month, year).

7. **Other scheduling considerations.**

   As noted above, other activities such as construction can have a significant impact on data availability. In addition, the nature of the construction project or other disruption can affect the selection of the time period of the analysis. For example, an HOV study should take into consideration a construction project that adds or disrupts an HOV lane (and possibly affects associated vehicle sensors in the pavement), or moves the HOV lane from an outside to inside lane.

Summary Statistic Considerations

The computation of summary statistics such as average daily volumes and peak hour or peak period volumes is based on an AASHTO-based process that requires user specification of minimum data quality thresholds.

1. **Data quality thresholds for "good" data at multiple-lane sites.**

   One of the differences between CDR and CDR Analyst is that CDR only looks at one lane at a time, while CDR Analyst is based on the idea of a site that can include multiple lanes. The latter capability is important because performance measures typically look at overall characteristics at a site (e.g., the GP volume) rather than individual lane performance (e.g., the Lane 1 volume). However, if the location of interest has more than one lane of interest (e.g., multiple GP lanes), individual lane characteristics must be combined. For example, to get a total volume at a site for a given 5-minute interval, the individual lane volumes must be added together. The same process needs to take place for the lane occupancy percentage, except that the values are generally averaged.

   Because of this, the user must determine an appropriate minimum number of good lanes that is required to compute a meaningful multilane measure. Once a user specifies a threshold of good lanes (as a percentage of total number of lanes), for each 5-minute interval the data flags of each individual lane value are checked to verify that the user-specified threshold of minimum "good" lanes is met. If it is not, the combined multi-lane statistic for that 5-minute interval is labeled "invalid."

   There are no absolute guidelines for setting the threshold. However, if the threshold is too relaxed, an atypical lane could produce a non-representative multilane summary value. If the threshold is too strict, measures might not be computed at all.

   As an example, a minimum threshold of 50 percent good lanes was used for the central Puget Sound Freeway Usage and Performance report.

2. **Data quality thresholds for "valid" data points at each step in the summary statistics process.**

   The AASHTO process starts with the archived 5-minute data and aggregates the values up to daily, then monthly, then yearly statistics. At each stage, user-specified thresholds of data quality must be met. For example, if an average weekday daily volume for a year is desired, the process works as follows: First, for each weekday of the year a total daily volume is summed up from the archived 5-minute data points (after summing across all lanes). During that process, the corresponding data quality flags of each 5-minute value are tabulated; the results are checked to verify that a minimum user-specified threshold of good data is available (e.g., at least 90 percent of all 5-minute values for the day are good values). Each day that meets that threshold can then contribute to the next level of aggregation; all other days are tagged as "invalid" and are not included in subsequent aggregations. Next, for each month a day-of-week average is computed (i.e., average Monday daily volume, average Tuesday daily volume, etc. for each month). For each of these day-of-week average values, a minimum threshold of valid days must be met (e.g., no more than one invalid day per day-of-week average value). If the threshold is met, that day-of-week average can contribute to the next level of aggregation; otherwise, that day-of-week average for that month is tagged as "invalid" and is not included in subsequent aggregations.

Next, for the entire year a day-of-week average is computed (e.g., average 1997 Monday daily volume, average 1997 Tuesday daily volume, etc.). For each day of the week, a minimum threshold of valid days must be met, e.g., at least ten day-of-week averages must be valid (out of the 12 average values corresponding to each of the 12 months). If that threshold is met, that yearly day-of-week average can contribute to the next level of aggregation; otherwise, that yearly day-of-week average is tagged as "invalid" and is not included in subsequent aggregations. Finally, if the resulting yearly day-of-week averages (Monday through Friday) meet a yearly threshold (e.g., at least 4 of 5 yearly day-of-week averages must be valid) the values are averaged together. The resulting value is then considered the average daily volume for the year.

Users must specify minimum thresholds of valid or good data points at each step of this process. There is no absolute guideline for the selection of those thresholds. However, because the results can be sensitive to those thresholds, it is important to use the same threshold values when doing a series of analyses whose results will then be compared to one another.

As an example, the minimum thresholds used for the central Puget Sound Freeway Usage and Performance report were:

Minimum number of good data points per day:	90 percent
Minimum number of good day-of-week values per month:	≤ 1 invalid value per day-of-week
Minimum number of good day-of-week values per year:	≥ 10 out of 12
Minimum number of good yearly day-of-week values:	≥ 4 of 5 (weekday)
	2 of 2 (weekend)
	≥ 6 of 7 (seven day)

These values can be set by the user in the CDR preferences file.

24-hour Site Profile Considerations

The time-of-day profile option in CDR Analyst produces per-lane volume, approximate speed, and approximate congestion frequency as a function of time of day for an average day. Interpretation of these measures should take into account the following:

1. **Speed formula accuracy.**

The speed estimation formula used in the tool set is a relatively simple function of volume and lane occupancy percentage. There are several assumptions built into this formula. First, previous studies have indicated that the formula is less accurate at the low and high speed ranges. Therefore the formula truncates the speeds at 10 mph and 60 mph to reduce the effect of this error. Second, the formula includes a coefficient that reflects the combined effect of the vehicle length and the sensor detection range on the resulting speed estimate. Vehicle length and sensor detection characteristics can both vary from site to site. However, CDR Analyst uses a single average value for all sites; this could introduce some site-to-site errors. Given these factors, it is important to consider the resulting speed estimates as general estimates that may be more useful as comparative rather than absolute values.

2. Definition of "congestion."

The congestion frequency histogram uses a user-defined threshold of congestion, using the lane occupancy percentage as a surrogate measure for congestion. As an example, the threshold of congestion used for the central Puget Sound Freeway Usage and Performance report is a lane occupancy percentage that corresponds approximately to the transition from traffic condition Level of Service E to F, based on a freeway with 65 mph freeflow speed characteristics as described in the Highway Capacity Manual. This assumes, however, a fixed average vehicle length; this assumption may introduce some error depending on the site. As with speed estimates, it may be more useful to consider the resulting congestion frequency estimate as comparative and qualitative rather than an absolute value.

Corridor Profile Considerations

The corridor profile option in CDR Analyst produces approximate congestion patterns as a function of time of day and corridor location for an average day. Interpretation of these measures should take into account the following:

1. Definition of "level of service."

Similar to the congestion frequency histogram, the contour map uses the concept of level of service to define different traffic conditions. In the case of contours, four traffic condition levels are used (green = Level of Service A, B and C; yellow = LOS D; red = LOS E; purple = LOS F). Each traffic condition level is color-coded on the map based on a corresponding range of lane occupancy percentage values. The levels of service are based on a freeway with 65 mph freeflow speed characteristics as described in the Highway Capacity Manual. As noted earlier, this assumes a fixed average vehicle length; this assumption may introduce some error. The resulting values are therefore most useful as comparative and qualitative measures.

2. Interpolation between measurement locations.

The contour maps are produced by sampling 24-hour profiles at several discrete locations along the corridor, then interpolating between those measured locations to produce smoothly varying traffic condition contours. The interpolation is linear. This assumption's validity will vary with the local conditions, and especially with the distance between measurement locations. This is an issue if there is a gap in measurements due to construction or equipment problems; the user must decide if a linear interpolation of traffic conditions is a reasonable representation of traffic patterns if a significant distance between measurement locations is being spanned. As with most of the user-specified values noted in this document, there are no absolute guidelines for determining the maximum gap size that can reasonably be spanned by interpolation; it will depend on such factors as traffic variability in that segment or the presence of nearby features such as major interchanges or significant sources/destinations of traffic (e.g., employment centers). Ultimately, this is a user judgment based on professional knowledge, experience, and common sense.

3. **Measurement location milepost values.**

   As noted earlier, a measurement site's milepost value is based on the location of the associated data cabinet, and a single field data cabinet collects information from multiple sensors. Because the milepost value is therefore an approximation, the specified locations of data values used in the contour map interpolation are approximations as well. As a result, specific traffic patterns seen on the contour map cannot be precisely located based solely on the map display. For example, if the contour map suggests a sharp change in level of service at milepost 170, it should not be assumed that the transition occurs precisely at that corridor location, but rather in the vicinity. (Furthermore, the corridor map that accompanies the contour maps should only be used as an approximate reference. As noted earlier, the corridor map can only be approximately positioned relative to the milepost axes of the contours. This is especially true if the corridor map must be distorted or "straightened out" to better match a curving corridor with the linear milepost axis.)

Travel Time Profile Considerations

The travel time profile option in CDR Analyst produces approximate average and 90th percentile travel times for an average day as a function of the start time of the trip and the trip route. Interpretation of these measures should take into account the following:

1. **Method of interpolation/extrapolation between measurement locations.**

   Travel times are computed by estimating spot speeds at a series of locations along the trip route. A travel time is estimated for each segment of the trip (where a segment's endpoints are defined by adjacent measurement locations), using the speed information and the length of the segment. These segment times are then added up to produce an overall trip time. The trip time can also be a function of the start time of the trip, by using spot speeds based on data from different times of the day. As with the contour interpolation, the reasonableness of the resulting travel time depends in part on the frequency of sampled measurements, and the resulting length of the interpolated/extrapolated segments.

2. **Speed formula accuracy.**

   As noted earlier, there are assumptions and simplifications built into the speed formula. This will of course affect the travel times. Given this factor, it is important to consider the resulting travel times as general estimates that may be more useful as comparative rather than absolute values.

3. **Speed variation between measurement locations.**

   As noted above, travel times are computed by estimating a travel time for each segment of the trip (where a segment is defined by successive measurement locations), by combining the speed information and the length of the segment. These segment times are then added up to produce an overall trip time. In this process, it is assumed that speeds vary linearly from one measurement site to the next. The reasonableness of this assumption depends in part on the frequency of sampled measurements, and the resulting length of those segments, as well as traffic characteristics on that corridor.

4. Measurement location milepost values.

As with the contour maps, a measurement site's milepost value is based on the location of the associated data cabinet, and a single field data cabinet collects information from multiple sensors. Because the milepost value is therefore an approximation, the specified locations of data values used in the travel time computation are approximations as well. The resulting travel times are best used as general estimates and may be more useful as comparative rather than absolute values.

Person Volume Considerations

Person volumes are based on a combination of CDR Analyst results (vehicle volumes) and mode split and vehicle occupancy values from other research projects and agencies.

1. Consistent measurement locations.

Because the data sources used for person volume computations vary, it is likely that measurement locations are not going to be coincident. In the interim report (section 4, Example I), sensor measurements, transit measurements, and vehicle occupancy measurements were not always at the same locations. In those cases, the user must decide if a site's measurement locations are "close enough" to one another, based on nearby traffic patterns.

2. Consistent measurement times.

Because the data sources used for person volume computations vary, it is likely that measurement times are not going to be coincident. In the interim report (section 4, Example A), sensor measurements, transit measurements, and vehicle occupancy measurements were not always collected at the same times. In those cases, the user must decide if the times are "close enough", based on seasonal variations or the timing of any construction projects that might affect the results. In some cases, data will be used because it is the only available data.

Appendix: CDR User's Guide

(available in separate PDF file)