CDR User's Guide

March 1998

John M. Ishimaru
Senior Research Engineer
Washington State Transportation Center (TRAC)
University of Washington
Seattle, WA

Version 2.52

Table of Contents

Introduction 1
Information Sources and Acknowledgments 1
CDR Software Installation 2
FLOW Data CD Archives 3
How to run CDR 4
Notes about the Output Options 13
How Summary Statistics are Computed 19
Notes about Data Validity 20
Notes about Loop Names 27
CDR User's Guide Version History 30

Figures

1. Main CDR Screen 8
2. CDR Options 9
3. Date Selection Screen 10
4. Location Selection Screen 11
5. Excel Import Wizard Window 12
6a. Example of 5-minute Output (Volume and Occupancy) 14
6b. Example of 5-minute Output (Speed and Vehicle Length) 15
7a. Example of 15-minute Summary Output 16
7b. Example of Daily Summary Output 17
8. Example of Multi-Day Output 18
9. Data Acceptance Threshold Window 25
10. Data Acceptance Threshold Window, continued 26

Introduction

This guide describes the procedures for using **CDR**, the **C**ompact disc **D**ata **R**etrieval program. CDR is a PC-based software tool developed by the Washington State Department of Transportation (WSDOT) to extract and process data from freeway traffic sensors in the central Puget Sound region. CDR uses a CD-based archive of traffic volume data and other information that are collected using a network of sensors embedded in the roadway pavement. CDR offers the user the ability to save and analyze a desired subset of the data based on specific days, locations, and analysis options of the user's choice.

This guide describes CDR's user options, the underlying algorithms for the computation of summary statistics, and caveats associated with CDR's output. The following topics are discussed:

- Installation instructions
- Data set description
- Operating instructions
- Output options
- Summary statistics
- Data validity considerations
- Loop naming conventions

Information Sources and Acknowledgments

The information in this guide was based on documentation provided with CDR and its DEC VAX- and modem-based predecessor, VDR; the author's experiences with the program; and email communications with WSDOT staff. The author wishes to acknowledge the technical assistance of WSDOT Northwest Region staff, with particular thanks going to Les Jacobson, Mark Morse, Dongho Chang, Lanping Xu, Mark Leth, Mike Forbis, Dave Berg, and CDR's programmer Alan Shen for their assistance with the use of CDR as well as the preparation of this document. This report was produced as part of documentation activities in support of the WSDOT FLOW Evaluation Design project.

If you have any comments about or corrections to this document, please contact John Ishimaru at the Washington State Transportation Center (TRAC) or via email at <jmi@u.washington.edu>.

CDR Software Installation

CDR requires a PC-compatible computer (preferably a 100 Mhz Pentium or faster) running Windows 95 or Windows NT, and a CD-ROM drive. The following instructions assume CDR use on a Win 95 machine.

(If the program has already been installed, you can skip this section.)

- **Install the software files.**

There are three software components of the **CDR** program:

CDR.EXE	The main program. **Version 2.52** is the latest official version as of March 1998. This version reads CD-based data from mid-1993 to the present.
CDR.INI	A file used by the main program to store user preference values. **Do not edit.**
RMDC.LST	A file used by the main program to match cross street locations to the cabinet list. This file should be included on the CD with the data.

To install the program on your computer, simply copy the first two files (CDR.EXE and CDR.INI) into a folder on your PC.

Note: For some earlier data sets, the RMDC.LST file is provided separately from the data CDs. In such cases, the RMDC.LST file should be copied into the same folder as the CDR program file and preference file.

FLOW Data CD Archives

CDR uses data stored on the FLOW data compact discs. These CDs hold traffic information collected on 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. The traffic data are stored in 5-minute intervals, which are aggregated from 20-second data collected by a network of WSDOT-operated data collection devices known as inductance loop sensors. Inductance loops are embedded in the pavement within a particular freeway lane at approximately 0.5-mile intervals in the central Puget Sound region, as well as at key locations such as interchanges. Vehicles that pass over a loop are sensed and their presence is then recorded by field data collection cabinets. These field collection devices then transmit data to the WSDOT's Transportation Systems Management Center at 20-second intervals, where it is processed and archived. In most cases, the loops collect two types of data: a) traffic volumes, i.e., the number of vehicles that pass over a loop in a given time interval, and b) average lane occupancy, the percentage of time that a loop detects a vehicle passing over it. Lane occupancy is useful as a general measure of congestion, and can also be used to estimate speeds. The CDs also store data validity information. (See **Notes about Data Validity** for more information.)

FLOW data CDs are available starting with 1993 data. It takes from 2 to 4 CDs to store the data for all locations for an entire year. CD-based data is available since approximately mid-1993 (the field devices were added to the computer-based archives between January and May of that year). Earlier data is available, though in less accessible form: volume data since 1981 is available on either tape or microfiche.

In addition to volume and lane occupancy, some loops are equipped to record other information. In particular, so-called **speed traps** (installations of two consecutive closely-spaced loops whose actions are coordinated) directly measure the speed of a passing vehicle and estimate its length, by noting differences in the time that each loop detects the same vehicle. (See **Notes about Loop Names: Numbering Rules #4** for more information.) Speeds are computed for 5-minute intervals based on an aggregation of 20-second speeds. The vehicle length information includes the estimated average vehicle length as well as the approximate distribution of estimated lengths of the vehicles passing by during that time period.

How to run CDR

1. Open up the Windows Explorer and navigate to the subdirectory with the CDR program in it. Double-click the CDR.EXE file to start the program.

CDR will display the main CDR screen which is divided into three columns, one for each of the three primary analysis options. **See Figure 1**.

2. Set up initial CDR parameters (if needed).

Next, you will need to specify the location of the input data (the traffic data archive) and the desired location(s) of the output data produced by CDR. **See Figure 2.**

Specify the CD drive location.

CDR accesses traffic data that is stored on compact discs. To use the CDs, you will need to provide CDR with the drive letter (e.g., E is typical) that represents the CD drive on your computer. Enter this letter designation by selecting the **Options** menu, choosing the **File Locations** option, and then choosing the **Data File** suboption. Enter the CD drive letter designation in the appropriate space, and click OK. If you're not sure what drive letter to use, look for the drive's letter designation in Windows Explorer.

Note: CDR also offers the option of installing the data files on a network hard drive. If you want to access the files that way, you will need to copy the data directory (with the included data files) from the CD to a network directory with the same name. (CDR searches for the directory designation.) Then, follow the procedures above to tell CDR the location, using the "Network Drive" option rather than the CD option.

Specify the output data file location.

Next, you need to tell CDR where to place output files that it creates. To specify the subdirectory where you want output files to be saved, select the **Options** menu, choose the **File Locations** option, and then choose the **Output File** suboption. You will then be given the opportunity to specify the subdirectory path for each of three types of analysis output that CDR can produce (raw data, daily summary, and multiday summary). Enter the full pathnames (e.g., C:\directory\subdirectory\) for each of the three output types. Unless you have a specific plan for organizing your output files, it's often easiest to specify the same pathname for all three options, which will result in all output files going into the same subdirectory; you can move the files around later. (The output file name is specified in the main CDR screen.)

This step only needs to be performed as needed; it does not need to be performed every time you run the program. The CD drive location and output file locations are saved after you enter them, and are used automatically for subsequent CDR runs until you change them.

3. Place a data CD of interest in the CD drive. If you want to produce yearly statistics, insert one of the CDs for that year in the drive.

4. Select the type of output you want.

The main CDR screen has three columns, representing three options for output:

1. Raw Data
2. Single Day Summaries
3. Multiple day Summaries

Set the output switch to "ON" for the output option(s) that you want. (See **Notes about the Output Options** later in this guide for more information about these analysis options.)

5. Specify the output data file name.

For the output option that you picked, enter the desired output data file name in the space provided. Data produced by CDR will be stored using this file name. The default file extensions are .DAT (raw data), .SDS (single day summary), and .MDS (multiple day summary), although you can specify any file extension you want. You are limited to the standard DOS 8.3 file name format (up to 8 characters, followed by a period, followed by up to 3 characters in the extension). You also need to specify the file format. In most situations, the spreadsheet option is preferred, since it generates a tab-delimited file for use by your favorite spreadsheet application.

Note: Be sure to use unique file names each time you access and save data. CDR will overwrite the contents of an existing file **without notification** if you specify one as an output data file.

6. Specify the data that you want.

You will next need to describe the specific data that you want. This is done by indicating which days and locations you are interested in.

Specify Desired Dates

To indicate which days you're interested in, click on the "Change" button under the "Dates" label. The window that appears will offer you several ways of selecting the days that you are interested in. **See Figure 3.** Moving from left to right on the screen, you can first select the specific months or days that you're interested in, by selecting (with one click) the month(s) that you want or the range of calendar day(s) that you want. Moving further to the right, the next filter allows you to further narrow down the dates of interest by specifying particular days of the week that you are interested in. (Drag the mouse or shift-click to select multiple months or days of the week.) Finally, the list on the right will show the resulting dates you've selected based on your filter choices. After you've selected the days you want, click OK to exit the "Change" window. If you need to, you can later verify the choices you've made by clicking on the "View" button above the

"Change" button. If you need to modify your choices after leaving the "Change" area, you will need to re-select "Change" and begin the process from the beginning; CDR does not save your previous selections once you re-enter the "Change" window. (Note that if you chose the multiple day summary option, you will only be asked for the year that you are interested in, since this option produces yearly statistics.)

Specify Desired Locations (Elements)

To indicate which locations you're interested in, click on the "Change" button under the "Element" label. CDR allows you to select specific elements (loops, stations, or speed traps), where each element is typically installed in a particular freeway lane. **See Figure 4.** (See **Notes about Loop Names** for the loop naming system.) You select the loops (lanes) of interest by first specifying the corridor that is of interest (left side of the screen), then selecting a specific location on that corridor (middle of the screen), then selecting which loop(s) at that location are of interest. To select a loop, double-click the loop name. Repeat this process as often as needed, keeping in mind that if you select multiple loops, all the data will be stored in a single output file. The list on the right will show the resulting loops you've selected. If you want to remove a loop you've selected, double-click the name on the selection list. After you've selected the loops you want, click OK to exit the "Change" window. If you need to, you can later verify the choices you've made by clicking on the "View" button above the "Change" button. If you need to modify your choices after leaving the "Change" area, you will need to re-select "Change" and begin the process from the beginning; CDR does not save your previous selections once you re-enter the "Change" window.

CDR only shows loops that exist on the start and end dates of the range of dates you specify. In some cases, you may be prompted for other CD(s); this is done to ensure that the loops that you select actually exist for the entire time period that you want to analyze.

Additional Choices

For the single-day summary and the multiple-day summary, you must choose one of the options listed on the main CDR screen. The single-day summary options are a) 15-minute and b) daily summaries, while the multiple-day summary options are a) monthly and b) yearly day-of-week averages as well as c) yearly statistics (AADT and AWDT).

CDR places vehicle volume and lane occupancy data in the user's output file. (See **FLOW Data CD Archives.)** However, depending on which of the three analysis options you choose, and which location (loop) you select, you can also obtain other data, including speed, vehicle length data, and data validity flags. These other data options are presented on the main CDR screen as part of the appropriate analysis options.

7. Get the desired data.

When you have finished choosing your options, select the "GO!" button in the lower left part of the main CDR screen. The output will be prepared and saved to a file within 5-10 seconds if all the data is on the CD you're using and the request is simple (a few loops and locations). In the case of summary statistics, you might be prompted for other CDs, and the process will take longer. Status messages will be displayed at the bottom left of the main screen during this process. These messages will indicate when output files are being created, when data is missing, or if the output file is already open.

8. Run other cases.

Return to step 3 to create other output files.

9. Look at the output data.

Each output data file will be stored in the subdirectory that you specified earlier. The file can be read directly by a spreadsheet application such as Microsoft Excel if you previously specified the spreadsheet output format (as recommended above). To open an output data file using Excel, begin by starting up Excel, then open the file using the **Open** option from Excel's **File** menu, and accept the default data types that are offered by Excel's Text Import Wizard (which appears as the file is initially opened and translated) by clicking the Wizard's "Finish" button. **See Figure 5.**

CDR also offers a built-in read-only utility to quickly review output data. Under the **File** menu, select **Open Output File**, then choose the file you want. Note that this utility can only view small files; larger files will require the use of an external reader (e.g., a spreadsheet application).

10. End the program.

When you are done using CDR, select **Exit** from the **File** menu.

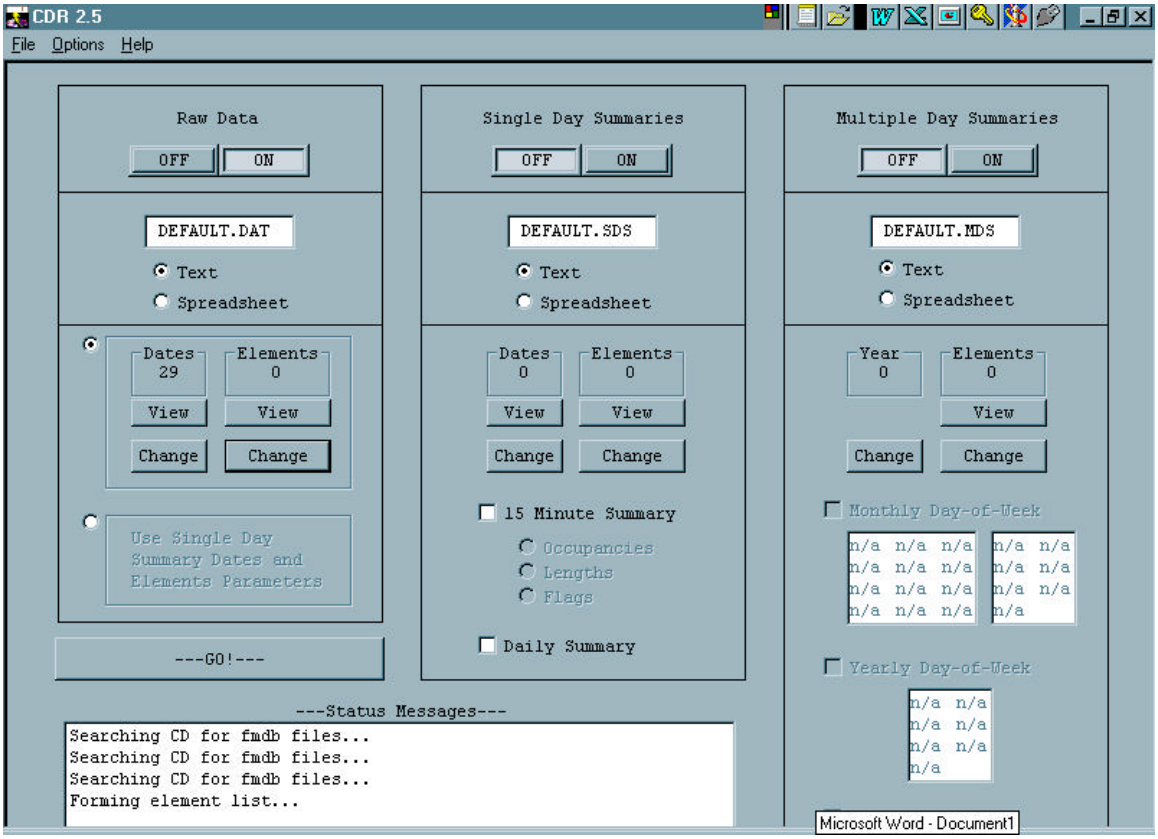
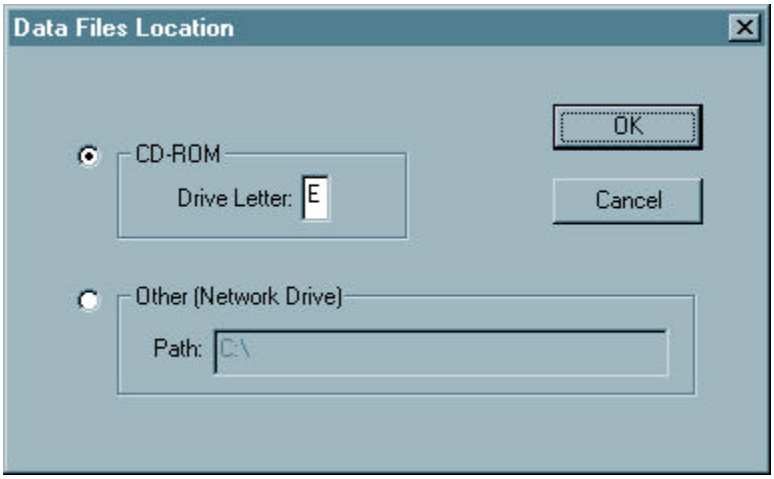


Figure 1. Main CDR Screen



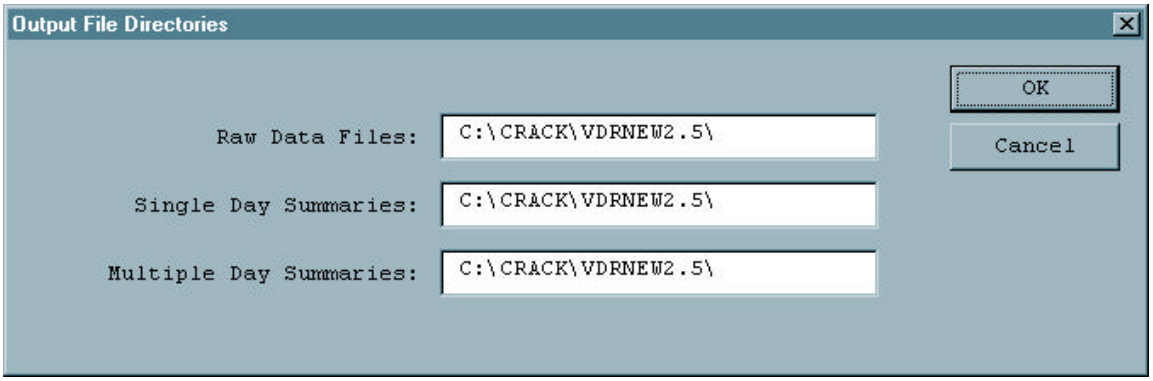


Figure 2. CDR Options

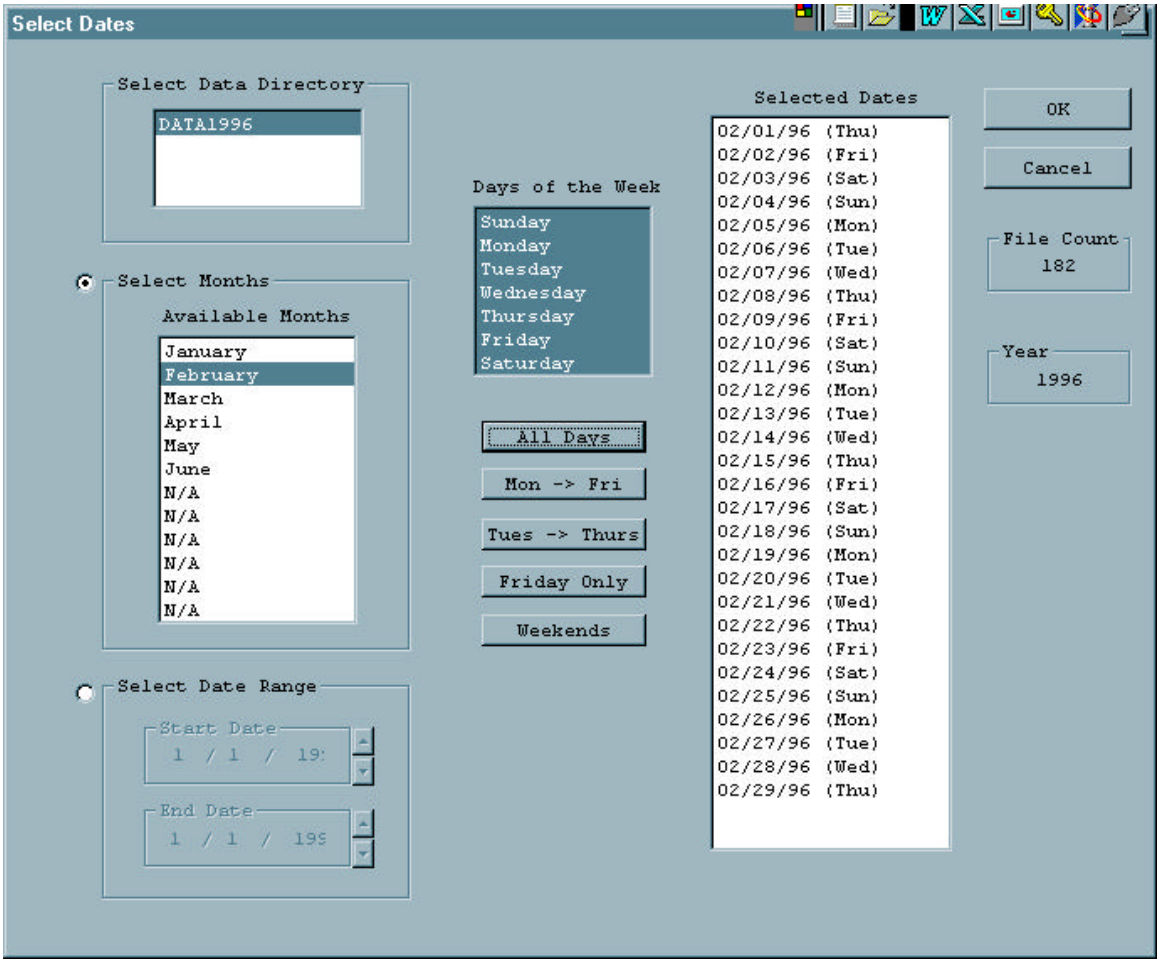


Figure 3. Date Selection Screen

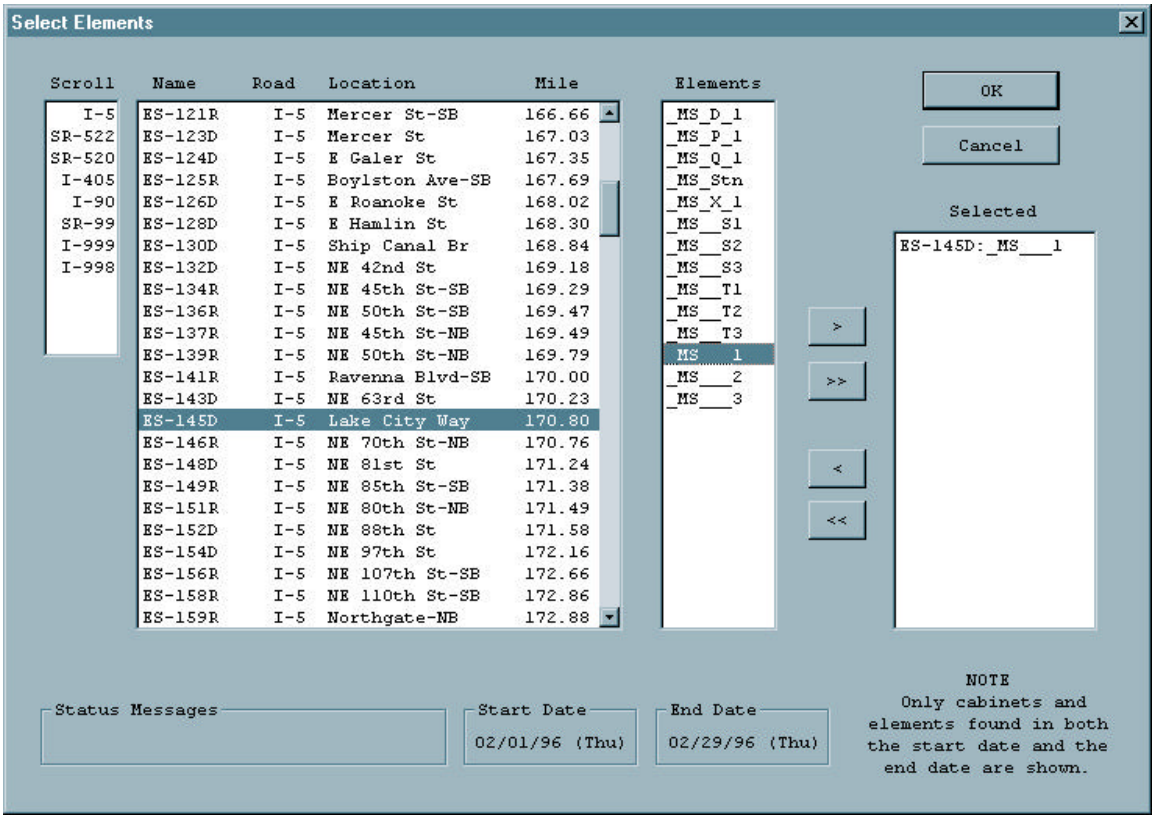


Figure 4. Location Selection Screen

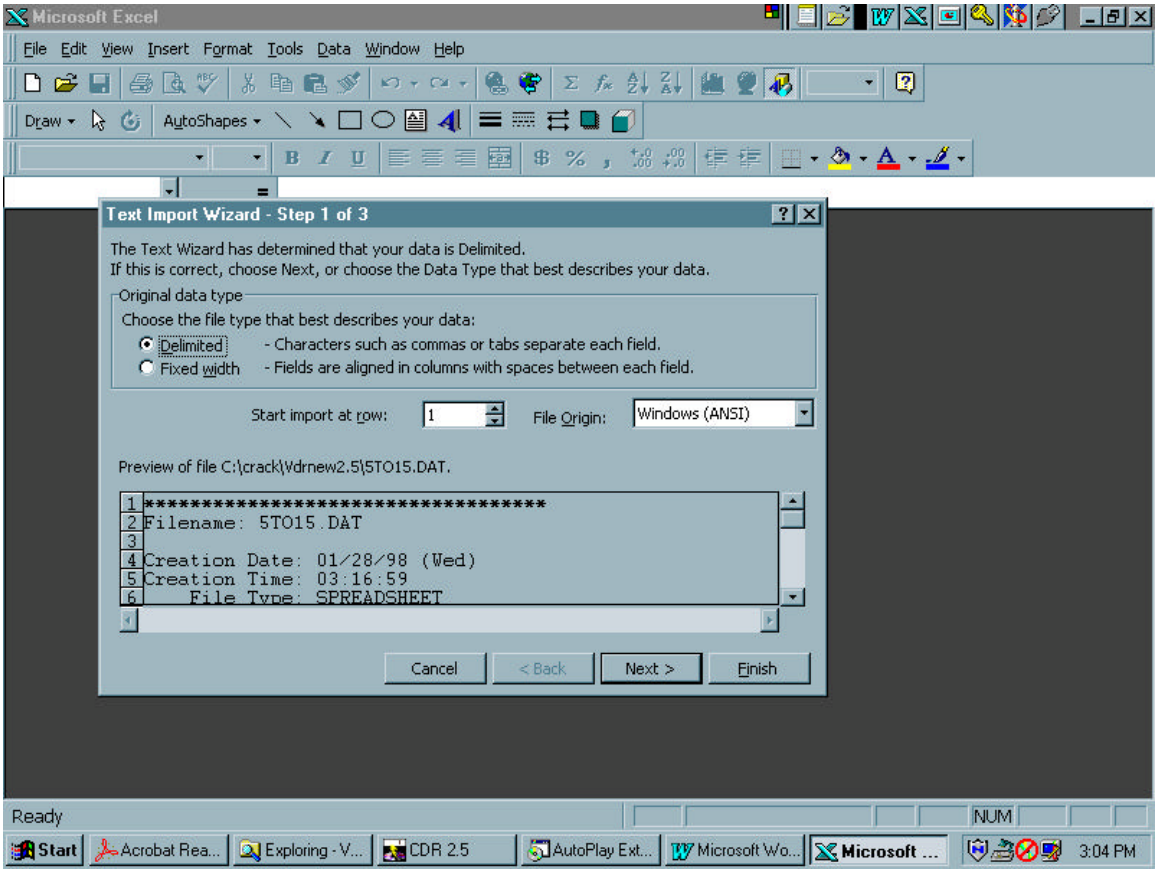


Figure 5. Excel Import Wizard Window

Notes about the Output Options

As noted earlier, there are three types of CDR output data. The following is a more detailed description of the options:

Raw Data (.DAT file extension): This option produces 5-minute counts of traffic volumes and average lane occupancy for the specified lanes and times. These are the actual data counts that are stored on the CD. The raw data output also shows a validity flag for each data point, as well as the number of 20-second data points that make up the 5-minute count; normally this would be a count of 15 (15 x 20 seconds = 5 minutes). Anything less than a count of 15 would indicate one or more missing 20-second data values.

The output includes vehicle speed and length estimates if the selected loop supports such computations (i.e., if it is part of a speed trap).[1] When average speed information is available, it is reported in miles per hour for each 5-minute interval, while vehicle length is reported in feet. The vehicles are further classified in the form of four "bins" or vehicle length ranges. **Bin1** is the volume count of all vehicles 26 feet long or less. **Bin2** is the count of vehicles from 26 to 39 feet long. **Bin3** is the count of vehicles from 39 to 65 feet long. **Bin4** is the count of vehicles that are greater than 65 feet in length.

Note: Speeds are not measured for every vehicle crossing a speed trap, especially during periods of heavy congestion. The four bin counts only include those vehicles for which speed and length values were successfully measured, and therefore the sum of the four bin counts will in general not equal the total vehicle volume for that lane and time interval. However, the distribution of vehicle lengths in the four bins should be representative of vehicles on the freeway at that time.

Daily Summary (.SDS): This option produces 15-minute and hourly counts of traffic volumes and average lane occupancy for the specified lanes and times. These are raw data counts aggregated from the 5-minute data on the CD. This option can also produce daily, peak hour, and peak period summaries. It also provides vehicle speed and length summaries if the loop supports such computations (i.e., if it is a speed loop).

Multi-Day Statistics (.MDS): This option produces monthly and yearly average daily traffic volumes and average lane occupancies for the specified lane, as well as peak hour and peak period averages. You can also specify weekday averages or averages by day-of-week. As noted earlier, when specifying the lane locations for MDS analysis you will be prompted for the other CD(s) that make up the specified year's data set. This is done in order to verify that the loops that you select exist for the entire time period that you want to analyze.

[1] The speed aggregation algorithm was modified starting in July 1996. Prior to that, a 20-second interval without a speed value (due to zero vehicle volume) contributed a 20-second speed of zero to the 5-minute speed average, thus pulling down the average. This was especially apparent late at night and to a lesser extent at other off-peak hours. These zero-volume values are no longer included in the 5-minute speed average starting July 1, 1996.

Figures 6a, 6b, 7a, 7b, and 8 show examples of the 5-minute, 15-minute, Daily, and Monthly output types.

****
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 6a. Example of 5-minute Output (Volume and Occupancy)

****
Filename: FIG6B.DAT

Creation Date: 03/17/98 (Tue)
Creation Time: 09:30:36
File Type: SPREADSHEET
****

ES-146R:_MN__T1 I-5 NE 70th St-NB 170.76
09/01/97 (Mon)

---Raw Speed speed.data Listing---

Time	Spd	Len	Bin1	Bin2	Bin3	Bin4	Flg	nPds
0:00	59.1	15.7	52	0	0	1	1	15
0:05	59.2	15.7	47	0	2	0	1	15
0:10	60.2	15.4	45	0	0	1	1	15
0:15	60.9	15.2	54	0	1	0	1	15
0:20	60.4	17.2	38	0	2	1	2	15
0:25	61.1	15.1	49	0	1	0	1	15
0:30	60.8	14.1	50	0	0	0	1	15
0:35	59.3	14.7	38	0	0	0	1	15
0:40	61.2	14.5	40	0	0	0	1	15
0:45	60.9	15.5	36	0	1	0	1	15
0:50	58.7	20.2	27	0	1	2	1	15
0:55	60	16.4	30	1	0	1	1	15
1:00	60	14.5	28	0	0	0	1	15
1:05	59.4	17.1	41	1	1	1	1	15
1:10	61.7	15.7	30	0	1	0	1	15
1:15	60.7	15.3	35	0	0	1	1	15
1:20	58.4	18.5	18	2	0	1	1	15
1:25	63.8	15	28	0	1	0	1	15
1:30	60.2	14.3	31	0	0	0	1	15
1:35	58.9	18.8	33	0	1	1	2	15
1:40	61.8	14.1	26	1	0	0	1	15
1:45	62.5	14	24	0	0	0	1	15

Figure 6b. Example of 5-minute Output (Speed and Vehicle Length)

ES-145D:_MS___1 I-5 Lake City Way 170.80
09/01/97 (Mon)

---15 Minute Loop Summary Report---

Time	:00	:15	:30	:45	Hour	Time	:00	:15	:30	:45	Hour	G	S	B	D
0:00	124	126	96	88	434	1:00	62	68	56	61	247	24	0	0	0
	3.40%	3.50%	2.80%	2.40%	3.00%		1.80%	1.90%	1.60%	1.80%	1.80%				
2:00	73	62	41	48	224	3:00	31	38	32	34	135	24	0	0	0
	2.00%	1.60%	1.10%	1.20%	1.50%		0.80%	1.10%	1.00%	0.90%	0.90%				
4:00	35	45	51	60	191	5:00	49	73	93	90	305	24	0	0	0
	1.20%	1.40%	1.60%	1.70%	1.40%		1.20%	1.90%	2.60%	2.50%	2.00%				
6:00	96	129	160	161	546	7:00	132	133	154	167	586	24	0	0	0
	2.70%	3.40%	4.30%	4.40%	3.70%		3.50%	3.80%	4.10%	4.60%	4.00%				
8:00	123	189	186	201	699	9:00	192	222	262	299	975	24	0	0	0
	3.40%	5.10%	5.20%	5.60%	4.80%		5.30%	6.10%	7.30%	8.00%	6.70%				
10:00	300	336	318	371	1325	11:00	386	406	436	449	1677	23	1	0	0
	8.10%	9.30%	8.50%	10.00%	9.00%		10.70%	11.10%	12.20%	12.50%	11.60%				

Figure 7a. Example of 15-minute Summary Output

****
Filename: 5TO15.SDS

Creation Date: 02/2/98 (Wed)
Creation Time: 03:19:11
File Type: SPREADSHEET
****

ES-145D:_MS___1 I-5 Lake City Way 170.80
09/01/97 (Mon)

---Daily Loop Summary Report---

Summary	Valid	Vol	Occ	G	S	B	D		
Daily	INV	0	0.00%	286	2	0	0		
AM Peak	VAL	1831	4.20%	36	0	0	0		
PM Peak	VAL	6226	10.50%	47	1	0	0		
AM Pk Hr	VAL	1599	11.00%	11	1	0	0	10:45	11:45
PM Pk Hr	VAL	1910	13.80%	12	0	0	0	12:45	13:45

Figure 7b. Example of Daily Summary Output

****
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 8. Example of Multi-Day Output

How Summary Statistics are Computed

There are four categories of summary statistics: 15-minute/hourly, daily, monthly, and yearly. Each is built up from the 5-minute raw data on the CD(s).

15-minute summary: The 15-minute traffic summary is computed by adding up the traffic volumes for the day's 5-minute segments over 15-minute increments. Occupancy is averaged over the 15-minute increment. Hourly totals are also computed. The validity of the 5-minute constituent data points is not considered in the 15-minute sums.

Daily summary: The daily traffic summary is computed by adding up the traffic volumes for the day's 5-minute segments to get the total daily volume. Peak hour and peak period volumes are computed analogously. The fixed peak periods used are 6-9 AM and 3-7 PM. Occupancy is averaged over each time period. The peak hour volumes are computed by using a moving one-hour window that moves at 15-minute steps. In contrast, peak hour speed data are computed at fixed one hour periods starting at 7 AM and 5 PM. The daily, peak hour, and peak period summaries are subject to user-specified data validity thresholds. (See **Notes about Data Validity** for more information.)

Monthly summary: The monthly traffic summary is computed by adding up the traffic volumes for each daily summary volume for that month. Peak hour and peak period volumes are computed analogously. The peak periods used are 6-9 AM and 3-7 PM. Note that peak hour volume averages over multiple days are computed based on the peak values for each day, even if they are at different times. Occupancy is averaged for each time period. The monthly summaries of daily, peak hour, and peak period statistics are subject to user-specified data validity thresholds. (See **Notes about Data Validity** for more information.)

Yearly summary: The annual traffic summary is computed by progressively summing up and averaging data from the daily level up to the monthly level up to the yearly level. First, total volume of each day in a given month is computed (see daily summary above). Then, an average volume is computed for each day of the week for that month. This process is repeated for each month. Then, an average volume is computed for each day of the week for the entire year, by averaging all 12 Monday averages together, all 12 Tuesday averages, etc. Finally, the resulting seven annual day-of-week averages (or five weekday averages) are averaged together to produce the AADT (average annual daily traffic) or AWDT (average weekday daily traffic) for the year. The yearly summaries of daily, peak hour, and peak period statistics are subject to user-specified data validity thresholds. (See **Notes about Data Validity** for more information.)

Notes about Data Validity

CDR keeps track of the validity of individual data points as well as the validity of statistics based on those data points, where "validity" is based on the operational status of the traffic sensor (loop) as well as the values being produced by the sensor. CDR also allows the user to specify how much "bad" or questionable data can be tolerated in the computation of a summary statistic. The following describes the validity checks used to evaluate data points, as well as the mechanism by which users can specify their tolerance for potentially invalid data.

- **Types of Data Validity Checks**

Two types of validity checks are performed on data collected by the WSDOT loop network.

5-minute data check

The first type of check is performed on the 20-second data that make up each 5-minute raw data value that is stored on CDs. This validity check is performed by the data collection and archiving software prior to data storage on the CD. During this check, each of the 15 20-second data points that make up every 5-minute data value is checked to see if it meets one of three conditions:

Bad Data: If the loop is hung (locked in an "on" state for longer than a prescribed time period) or the loop's data is "outside the envelope" (i.e., the volume-occupancy combination is not reasonable)[2], the corresponding 20-second data value is labeled "**Bad Data**". The assumption is that a locked "on" state or highly atypical combinations of volume and occupancy values are symptomatic of erroneous data collecting conditions.

Disabled: If the loop has been disabled by a system operator, the corresponding 20-second data value is labeled "**Disabled**".

Good Data: If the loop does not meet the criteria for the first two conditions, the corresponding 20-second data value is considered "**Good Data**".

After each of the 15 20-second counts that make up a 5-minute value is evaluated in this way, the associated 5-minute value is given an overall data validity flag, according to the following rules:

[2] The approximate envelopes of acceptable volume-occupancy combinations (based on earlier TRAC research) have been encoded in the loop controller software as the following (for 20-second data):

If occupancy = 1, then volume must be = 2.
If 1 < occupancy = 5, then 0 = volume = 7.
If 5 < occupancy = 10, then 2 = volume = 11.
If 10 < occupancy = 16, then 1 = volume = 17.
If occupancy > 16, then 0 = volume = 17.

If all 15 20-second values are labeled “Disabled”, the associated 5-minute data value is considered “**Disabled**”.

If all 15 20-second values are labeled “Good”, the associated 5-minute data value is considered “**Good**”.

If all 15 20-second values are labeled “Bad”, the associated 5-minute data value is considered “**Bad**”.

For all other combinations of 20-second validity flags, the associated 5-minute data value is considered “**Suspect**”.

The overall validity flags are placed next to each data value in the 5-minute output file, using a number code. (For information on the flag codes, see **Data validity codes**, later in this section.)

Summary statistic check

The second type of check is performed by CDR when computing summary statistics based on the 5-minute data. CDR has the ability to take raw 5-minute data from the original data files and compute statistics for longer time periods such as daily, monthly, and yearly averages. It does this by following a series of data aggregation steps that sum up and average the data. At each step of this process, CDR tries to determine whether the resulting summary statistic is valid, by looking to see how much of the data used to create the statistic is considered good data. Thus, the validity of each summary statistic (daily, monthly, yearly) is evaluated based on the validity of its constituent statistics.

Summary statistic validity checking begins with the initial step of the summing-and-averaging process, which is the computation of daily statistics using 5-minute data (daily sums, peak hours, peak periods). At this first step, CDR looks at the data validity flag associated with each 5-minute data point (the flags stored on the CDs with the volume and occupancy data) to see how much of the data that is being summed up to create the daily summary statistic is “good”, and how much is otherwise questionable. This tabulation is then reported in the summary statistic output file, using letter codes. (For information on the flag codes, see **Data validity codes**, later in this section.) CDR then compares the number of “good” and “questionable” data values with user-specified threshold values of acceptability, to determine whether the resulting daily statistic is valid or not. (See **User-specified data acceptance thresholds** for more information.)

- **Data validity codes**

As noted above, the validity of 5-minute data are determined by the data collection system, then stored on the data CD, while the validity of a summary statistic is evaluated by tracking the

validity of its constituent 5-minute data values, keeping a count of the number of data points that fit different categories of validity, and comparing the counts to user-specified acceptance levels.

Both the 5-minute validity information (determined by the data collection system) and the summary statistic validity tabulations (computed by CDR) can be included in CDR's output, using a code that indicates the nature of the data validity. There are four categories of data validity:

Data is:	5-minute data	Summary statistic
bad value	0	B
good value	1	G
suspect value	2	S
loop is disabled	3	D

Number codes are used for 5-minute data output, while letter codes are used for summary statistics. In the 5-minute output, a number code is provided for each data point, while in the summary statistic output, tabulated validity counts are provided which categorize the total number of B, G, S, and D data points per 2-hour blocks (in the 15-minute output) or per statistic (in the yearly, monthly, daily, peak hour, peak period output). Each summary statistic's tabulation of B, G, S, and D counts is based on the sum of tabulations from its constituent data points. Speed loops show only B, G, and S data.

- **Data replacement to improve bad data**

When a summary statistic is calculated, CDR copes with any invalid data detected in the 5-minute data set used to compute that statistic by replacing the bad data with better data. Specifically, if a 5-minute data point is determined to be suspect, bad, or disabled, it is replaced with the most recent previous "good" data point. (The user also has the option to use suspect data unaltered.) The modified data is then used to compute the statistic. If all data for a day is bad, disabled, or suspect, the resulting summary statistic is set to zero.

In the following example, suspect data (flagged with a "2") is improved by replacing it with the most recent good value (volume = 35, at 10:05):

Time	Volume	Validity Flag	Modified Data
10:00	25	1	25
10:05	35	1	35
10:10	37	2	**35**
10:15	42	2	**35**
10:20	41	1	41

Note that this process is used only to facilitate computation of summary values, and only when there is mostly good data (i.e., user thresholds of data acceptability are met). In other words, data replacement is used to fix occasional bad data in an otherwise good data stream, to facilitate computation of a summary statistic. If an entire day's data are considered invalid because it has too much bad data (i.e., doesn't meet user thresholds), the data are not improved in this manner;

instead, that day's summary is set to zero (as noted earlier), and is not included in any subsequent summary statistic calculation (e.g., monthly, yearly).

- **User-specified data acceptance thresholds**

CDR gives the user the flexibility to set a number of data acceptance thresholds that determine whether or not to accept a summary statistic, based on the validity of the data that is used to compute the statistic. These thresholds are specified by indicating the minimum amount of "good" or valid data that one is willing to accept at each stage in the summary statistic calculation process, below which the resulting statistic is tagged as unreliable and is discarded or otherwise replaced. One can also specify the maximum amount of "questionable" data that one is willing to accept, above which the resulting statistic is tagged as unreliable and is discarded or otherwise replaced.

This flexibility is a double-edged sword. While it gives the user control over the summary statistics that are created, it also requires the user to make more decisions. The data acceptance thresholds provide the user with additional flexibility to cope with missing or questionable data, but there are no specific rules for determining what those threshold values should be. The resulting statistic could be significantly affected by the user-specified thresholds. For example, overly strict thresholds could result in large quantities of data being thrown out, resulting in summary statistics based on a limited subset of the original data that are therefore of questionable validity. It could even result in no statistic at all, if so much data has been thrown out at lower levels (daily statistics) that thresholds at higher levels (monthly or yearly) cannot be met. Caution is therefore recommended when specifying these values. It might be desirable to perform a sensitivity analysis to determine good threshold values, or alternatively, to determine that it is better not to have any thresholds at all.

- **How to set data acceptance thresholds**

Data acceptance thresholds are specified in the **Options** menu, under the **Preferences** option. Thresholds are set up for six categories of summary statistics: Daily, AM Peak, PM Peak, AM Peak Hour, PM Peak Hour, and Multiple Day. (**Note**: These threshold values are retained during the program run, but are **not** stored once you exit CDR.) For the first five statistics categories (Daily, AM Peak, PM Peak, AM Peak Hour, PM Peak Hour), you have the following choices **(see Figure 9)**:

> Perform Running Flag Check
> CDR scans the 5-minute data set (whether it's daily, peak or peak hour) from the start time to the end time, and checks to see if consecutive blocks of data meet the user's thresholds. This is done by setting up a "window" of consecutive data points and moving the window along the data set from the start time to the end time. The user specifies the size of the window, the number of data points that the window moves before checking the next block, and the minimum/maximum acceptance thresholds within the window for good, suspect, bad, and disabled data.

As an example, in the case of a daily running flag check, suppose we have the following settings:

step size	1	maximum suspect	1
width	12	maximum bad	1
minimum good	10	maximum disabled	1

These values mean that the window is 1 hour wide (i.e., width = 12 x 5 minutes per data point = 1 hour), and moves along at 5-minute intervals (step size = 1 data point = 5 minutes). Within the moving window, there can be no more than 1 suspect, bad, or disabled 5-minute data value, and a minimum of 10 good 5-minute data values. Otherwise, the resulting summary statistic is considered invalid.

Perform Total Flag Check
This option works similarly to the running flag check, except that the thresholds are for the entire data set, not a moving subset.

Use Suspect Data
Data that is suspect, bad, or disabled is normally "patched" with the most recent good value. The user can select this option to instead use suspect data unaltered.

For the last statistics category (Multiple Day), you have the following choices **(see Figure 10)**:

Days to Monthly Day-of-Week: The user specifies the maximum allowable number of invalid or missing days while computing valid day-of week averages for a month. Invalid days are days that don't meet validity checks. Missing days are days when no data is available due to data corruption or because of miscellaneous reasons.

Monthly Day-of-Week to Yearly Day-of-Week: The user specifies the maximum allowable number of invalid monthly day-of-week averages while computing valid day-of week averages for a year.

Yearly Day-of-Week to AWDT or AADT: The user specifies the maximum allowable number of invalid yearly day-of week averages while computing the AWDT or AADT.

Note that the thresholds for the first five categories of statistics operate independently of one another, but the thresholds for the last category of statistics are dependent on threshold settings from the other five categories, since multi-day statistics are aggregated from the daily values.

The user thresholds described above are combined with the tabulated counts of valid and invalid data points that make up each summary statistic to determine whether the resulting statistic is valid. For example, an AADT is made up of the combination of 7 (day-of-week) values, each of which is determined to be either valid or invalid based on its constituent data. The user threshold is then used to determine whether the resulting AADT is valid. These results are shown in the

output file, which indicates the count of how many of the 7 values are considered valid or invalid, as well as the end result (i.e., is the statistic valid or not), reported as **VAL** or **INV**.

(**Note**: In CDR version 2.52, speed loop data are not subject to validity checks.)

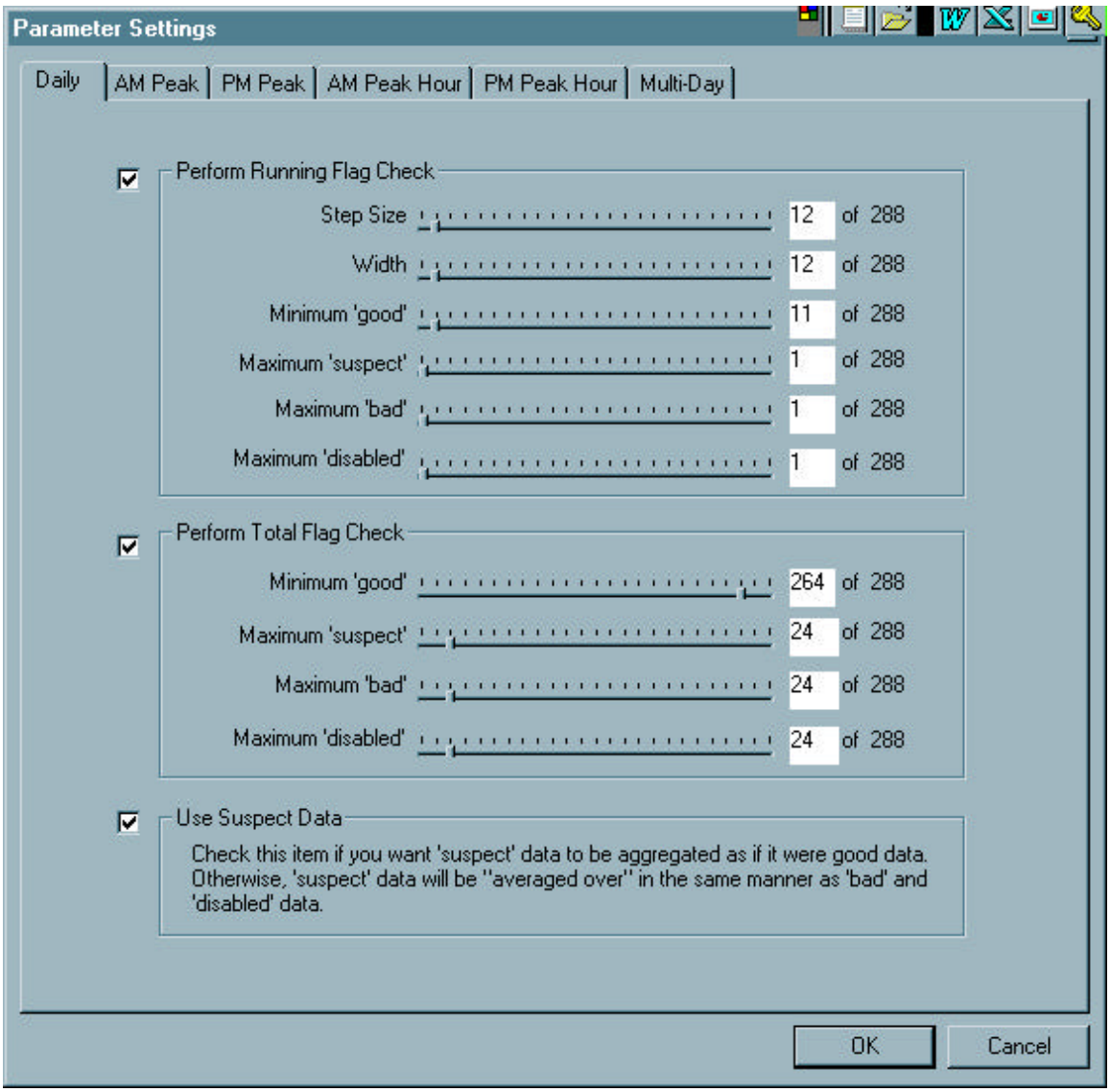


Figure 9. Data Acceptance Threshold Window

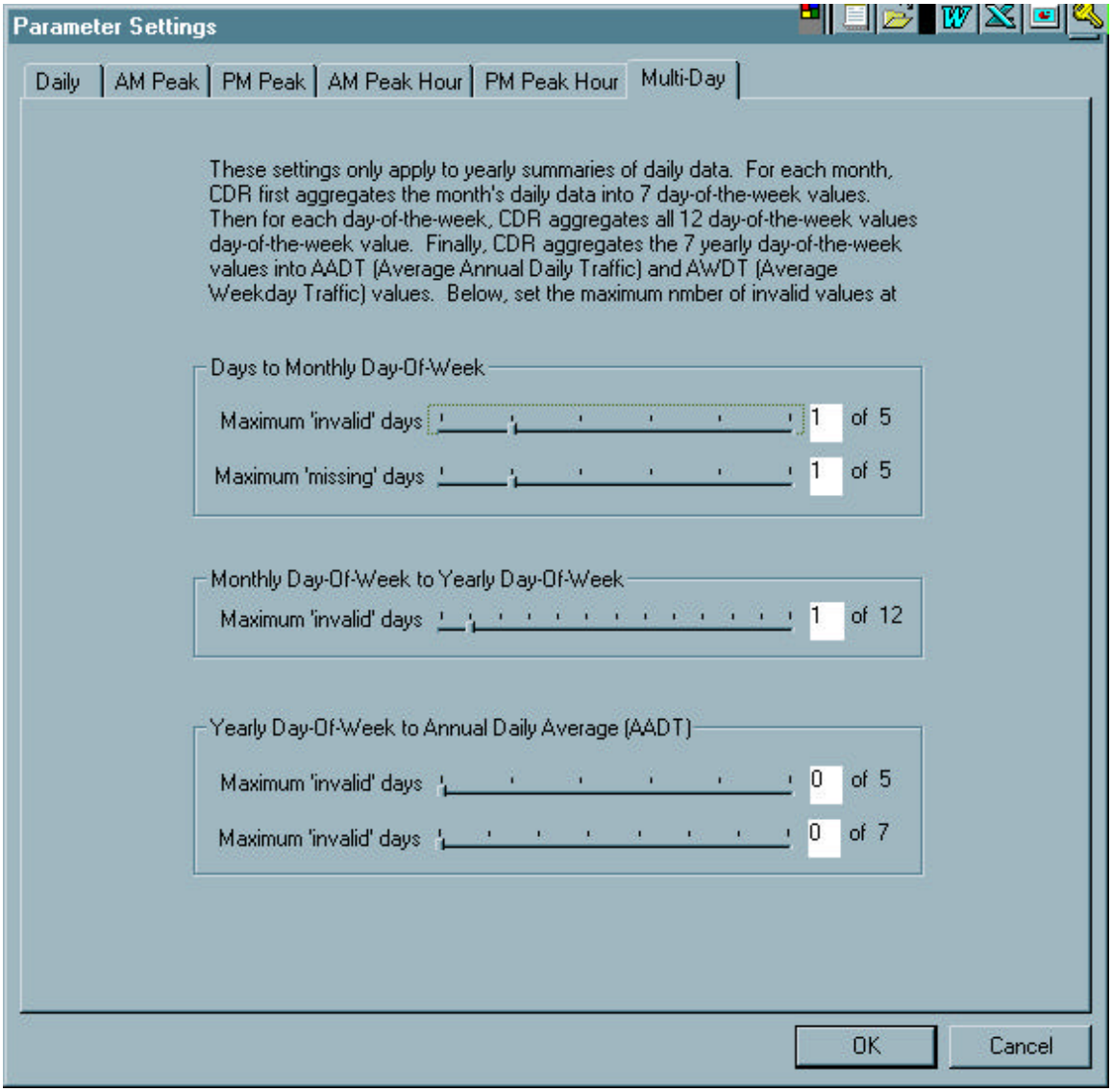


Figure 10. Data Acceptance Threshold Window, continued

Notes about Loop Names

The following information is adapted from the Loop Naming Rules documentation of the Ramp Metering Database.

Each loop name contains exactly 7 characters, and is based on the loop's location and purpose. The code is created by concatenating the following codes from left to right:

2 characters		1 character		2 char.		2 characters	
Roadway Code	Description	Direction Code	Description	Lane Type Code	Description	Lane # Code	Description
_M	mainline	S	southbound	_X	exit	_1	
_C	CD	N	northbound	_O	on ramp	_2	
_R	reversible	E	eastbound	RA	Rt Adv Q	_3	
AM	aux mnln	W	westbound	LA	Lt Adv Q	_4	
AC	aux CD			_Q	queue	_5	
AR	aux rev			_I	inter Q	_6	
MM	metered mnln			_D	demand	_7	
MC	metered CD			_P	passage	_8	
MR	metered rev			HX	HOV exit	_9	
				HO	HOV on	S1...S9	speed loop
				HD	HOV demand	T1...T9	"virtual" speed loop
				HP	HOV passage		
				H_	HOV mainline		
				__	mainline		

The underscores are considered characters and are required.

For example, the seven character names

_MN___1 and **_MNH__1**

represent a mainline northbound general purpose lane and a mainline northbound HOV lane, respectively. Both are on the far right when facing downstream, as indicated by the lane 1 designation.

The following are additional naming rules:

Roadway Naming Rules

1) Each cabinet is assigned a principal roadway. For example, if a cabinet is assigned to principal roadway I-5, but also has loops on I-90, then the I-90 loops are considered auxiliary (A in column 1). In the case of on-ramps and exits, the principal name assignment prevails; in the

previous example, for instance, if that cabinet had a loop whose lane could be considered both an exit from I-90 or an on-ramp to I-5, it would be assigned on-ramp status, since I-5 is the principal roadway designation for that loop's cabinet.

2) Loops that are part of a ramp metering station (i.e., the loop's data are used in the metering algorithm) are designated with an M in column 1. Speed loops and HOV lanes do not use the M in column 1.

3) All loops must have a roadway type. On-ramps and exits take on the roadway type of the type of roadway that the ramp enters or leaves respectively.

Direction Rules

1) All loops must have a direction code: northbound, southbound, eastbound, westbound.

2) Reversible roadways will take on the direction of the increasing milepost. I-5 express lanes are considered Northbound.

Lane Naming Rules

1) All loops will have one lane type.

2) Each metered lane can have two advance queue loops, one left movement (LA) and one right movement (RA). If no movement is associated with the advance queue loop, use RA.

3) Depending on ramp length, each metered ramp lane can have up to 2 queue loops, one intermediate (_I) and one queue (_Q), depending on ramp length. The queue loop and intermediate queue loop will be evenly spaced, 200-500 feet apart, depending on ramp length. The demand (_D) and passage (_P) ramp loops are just before and after the meter signal, respectively.

4) Use _X for exit ramps, and _O for on-ramps that are not metered.

5) HOV lanes are specified by an H in column 4.

6) Use __ (double underline) or H_ for mainline loops.

Numbering Rules

1) Loops of similar type are numbered from upstream to downstream.

2) All multi-lane roadway lanes are numbered from right to left looking downstream (traffic flow direction). If one or more of the lanes are HOV lanes, they are numbered as if they are GP lanes.

3) Each HOV lane bypass will have the same lane number as the ramp lane it bypasses.

4) Speed loops are designated by an S in column 6. Where speed traps are installed, there are three relevant loops. The pair of mainline loops used to compute speeds is a mainline loop with an _ (underline) in column 6, and a corresponding speed loop whose name is identical except with an S in column 6. The computed speeds and vehicle length information are stored in a "virtual" loop that has the same name, except with a T in column 6. The T loop does not actually exist in the field.

5) Ramp meter lanes are numbered right to left, upstream to downstream. All pre-1992 ramp meter lanes are numbered 2.

Station Naming

1) All station names contain roadway and direction type followed by _Stn or Hstn. (A station is typically a grouping of mainline general-purpose lanes that can be used as input data for ramp metering operations, although it can consist of up to 8 loops of any type. Data from a station reflects the sum of all volumes (and the average occupancies) associated with loops at that station.)

2) This naming scheme works only for mainline, CD, reversible, and HOV lanes on these roadway types.

CDR User's Guide Version History

Version	Date	Description
unnumbered	1/30/98	Original version
2.0	2/4/98	More information on bin count, vol-occ envelope, nPds, and _Stn Figures added Minor wording changes **WSDOT review version**
2.52	3/18/98	Incorporate WSDOT review comments Update to reflect CDR version 2.52 feature set Add Figure 6b Add Table of Contents and Table of Figures Correct descriptions of RMDC.LST file, network drive option, vehicle length bin counts, and data validity flags Minor wording and format changes **PDF (Adobe Acrobat) version created and placed on data CDs produced starting 3/18/98**