United States
Environmental Protection
Agency

Air and Radiation

EPA420-R-03-018
August 2003

Roadway-Specific Driving Schedules for Heavy-Duty Vehicles

EPA420-R-03-018
August 2003

Roadway-Specific Driving Schedules for Heavy-Duty Vehicles

Assessment and Standards Division
Office of Transportation and Air Quality
U.S. Environmental Protection Agency

Prepared for EPA by
Eastern Research Group, Inc.
EPA Contract No. 68-C-00-112
Work Assignment No. 3-07

NOTICE

This technical report does not necessarily represent final EPA decisions or positions. It is intended to present technical analysis of issues using data that are currently available. The purpose in the release of such reports is to facilitate the exchange of technical information and to inform the public of technical developments which may form the basis for a final EPA decision, position, or regulatory action.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

EPA is currently beginning development of a new mobile source emissions model that will replace MOBILE6. As part of that development, modules are being written to calculate emission factors from typical driving traces of different kinds of vehicles under different operating conditions. For this work assignment, EPA has asked ERG to use existing heavy-duty vehicle activity data to develop representative speed versus time driving cycles for heavy-duty vehicles. Once these driving cycles are developed, they can be incorporated into the code of the new mobile source model where they can be used by simulation programs to estimate the emissions of vehicles as they might be produced by heavy-duty vehicles actually driving the particular schedules.

In this study, we construct cycles from so called micro-trips in an effort to match the speed, acceleration, and vehicle specific power characteristics of the non-idle driving portions of the dataset. These three particular activity variables are chosen for matching purposes because they largely influence the emissions behavior of a given vehicle. The effects of vehicle weight and road grade are not included in developing these cycles even though they are known to have important effects on emissions because those parameters were not available in the existing dataset. Accordingly, the cycles developed in this study should be regarded as a temporary solution to describing heavy-duty vehicle driving behavior. When data becomes available that has vehicles weights and road grade in addition to speed, acceleration, and vehicle specific power on a second-by-second basis, then improved driving schedules can be developed.

Another reason that these cycles should be regarded as temporary is that the mix of vehicle types and vehicle usage was not planned during data collection. This means that the data represents operation of the vehicles that just happened to be instrumented rather then a representation of the mix of different types of vehicles and usage that occurs in the fleet.

To meet the particular needs of the new mobile source emissions model, EPA requested that a set of separate driving cycles be developed for different combinations of vehicle type/usage, freeway/non-freeway driving, and different average speeds. A separate, although related, analysis of the activity data was performed to identify the speed bins for which an adequate amount of existing data was available.

The cycles developed in this study are not intended to be used to test vehicle emissions on a dynamometer but are solely to be used in the new model. As a result, the duration of the individual schedules did not need to be limited although EPA did want to have the duration of the cycles be reasonable so that a large amount of computer memory would not be required for them. In addition, we were also instructed to consider all of the existing data to be driving under warmed-up, running operating conditions. In other words, we were not to build different "bags" that characterized vehicle operation during cold starts and hot starts, for example.

Section 2 describes the sources of existing data. Section 3 describes the preparation of the raw data for cycle building. It includes a discussion of the numerous quality checks and edits that were made to the millions of second-by-second observations in the datasets. Section 4 describes the analysis of the datasets for the purposes of identifying the different cases for which cycles would be built. Section 5 describes the general methodology and the specific details of building the cycles. The results of an analysis that characterized the heavy-duty operating characteristics of the vehicles are presented in Section 6. A comparison of statistics for the cycles and the datasets on which they were built is presented in Section 7. Finally, in Section 8, we make recommendations for the development of final cycles using these datasets.

2.0 Data Sources

Three sources of data were used to develop the cycles in this study.

Second-by-second driving data on four Texas Department of Transportation dump trucks were provided by TxDOT. The data was collected using dataloggers based on the Cummins QuickCheck that attached to the vehicle's serial data communication port following the SAE J1587/J1708 protocol.

Heavy-duty truck activity data from the Battelle study was collected using data logged from global positioning system (GPS) units installed on 140 vehicles. Data from 120 of those vehicles were used to help develop the cycles in the study. The collection of this data is described in "Heavy-Duty Truck Activity Data," Battelle, Columbus, Ohio, April 30, 1999.

Activity data on heavy-duty trucks was also collected using GPS units by Jack Faucett and Associates. Second-by-second data from 31 trucks were available and data from 30 of the trucks were used to help build driving cycles in this study.

3.0 Preparation of Raw Data for Cycle Building

The raw data taken in the TxDOT, Faucett, and Battelle heavy-duty vehicle studies required varying amounts of preparation before the data could be used to develop heavy-duty cycles. The TxDOT data on the four dump trucks had already been quality checked and corrected by ERG prior to its use in this study. While the Faucett and Battelle data had been used previously in other studies, that data was reviewed for quality in this study to ensure a consistent level of quality throughout the datasets and to identify any specific issues that needed to be addressed during cycle development.

An initial examination of the quality of the Faucett and Battelle second-by-second data, which were collected from GPS units, revealed that a substantial effort beyond the cost limits of the project would be required to detect and repair problems in the millions of observations in those datasets. Accordingly, as far as quality control was concerned, we took the following approach:

1) Faucett and Battelle timestamps were repaired to the degree possible to provide a continuous flow of 1-second time steps when vehicle engines were believed to be running.

2) Faucett and Battelle speed values that we could reliably determine to be suspect were changed to missing values.

3) Faucett and Battelle idle observations were detected as well as could be and the corresponding speed values were set to 0.00 mph.

4) For all three datasets, all trips were divided into micro-trips based on idles assigned to observations. Micro-trips from Faucett and Battelle data were "ragged" because they contained some missing speed values, some remnant GPS dither, some uncertainty about what was idle, and some uncertainty about the location of the beginning and end of micro-trips. Micro-trips from TxDOT had none of these problems; they were "clean."

5) For all three datasets, micro-trips were categorized into cases according to vehicle type/usage, freeway/non-freeway, and average micro-trip speed.

6) For each case, cycles were built using micro-trips to describe all of the operation for the case. Note that both "clean" and "ragged" micro-trips were in the operation dataset and were considered for use in the cycle.

7) After the micro-trips for each cycle were selected, the missing values, remnant dither, remnant high accelerations and decelerations, and any other suspect features were to be repaired in the cycles.

By changing suspect speed values in the dataset to missing values so that they not affect micro-trip selection, and by repairing the micro-trips only as they are selected, we believe that representative cycles can be developed by this approach, without embarking on the huge job of repairing every suspect observation in the entire dataset.

The several subsections that follow describe the major steps in quality checking the Faucett and Battelle data and the steps used to mark special second-by-second observations in the vehicles from all three studies. The first two subsections describe the methods used to correct the timestamps in the raw datasets and to flag second-by-second speed values that were suspect. SAS programs (Battelle/qc_bl.sas and Faucett/qc_bl.sas) were written to automate the timestamp insertion and speed value flag process. This was done so that hand editing of the millions of observations in these datasets could be avoided.

The approach for development of the cycles is based on a process of building up cycles from individual micro-trips selected from the database. For this to work, the trips for all vehicles must be separated into micro-trips. In this study we define the beginning of a micro-trip as the point at which the vehicle speed moves from a non-idle speed to an idle speed. For a dataset such as that collected by TxDOT, which used dataloggers from a speed transducer on the vehicle, the distinction between non-idle and idle speeds is clear-cut. That is, when the speed is 0.00 mph, the vehicle is idling. Unfortunately, the vehicle speeds from the GPS units in the Faucett and Battelle studies were almost never observed to be 0.00 mph – probably because of the effects of dithering. Dithering is the term we use in this study to describe the effects of Selective Availability (SA)[1]. The speeds from the Battelle and Faucett datasets showed low speeds that moved up and down in the vicinity of 0 to 5 mph when the vehicle was not moving. We developed a probabilistic method of detecting idle speeds using SAS programs (stats/fid.sas and stats/vid.sas) for the Faucett and Battelle datasets.

As an aside, there may be a question of whether dithering in the GPS speed values of the Faucett and Battelle data introduces a positive bias in the reported speed values. We believe that the reported speeds above about 10 mph are not biased on the average, but the speeds below 5 mph have a substantial positive bias. At high speeds dither causes the length and direction of the velocity vector to be uncertain. However, since the magnitude of the velocity vector is much less than the magnitude of the velocity vector, fluctuations in the length of the velocity vector average out to be zero over several seconds. On the other hand, at low vehicle velocities, when

[1] According to www.garmin.com, "SA is an intentional degradation of the GPS signal once imposed by the U.S. Department of Defense. SA was intended to prevent military adversaries from using the highly accurate GPS signals. The government turned off SA in May 2000," which was after all of the data in the Faucett and Battelle data used in this study were collected.

the magnitude of the dither is comparable or greater than to the magnitude of the velocity vector, fluctuations in the length of the velocity vector will not average to zero. The extreme case is when the vehicle is not moving. In this situation, the average speed is zero but because of the dither the velocity vector points from second to second in different directions. The length of the vector is not zero and a vector cannot have a negative length. Therefore, the average speed is biased positive.

Once the Battelle and Faucett datasets had timestamp corrections, flags to designate suspect speed values, and a flag variable to mark observations believed to be idles, the three datasets were combined into a master dataset. Then, the idle flags were used to separate trips into individual micro-trips. The entire dataset was then written to a final SAS dataset. This was done using the program stats/prep.sas.

3.1 Timestamp Corrections

The qc_bl.sas programs were used to make timestamp corrections to the raw Faucett and Battelle datasets. In general, each timestamp was compared to the previous and subsequent timestamps to determine if each timestamp was consistent with those around it. Examination of the data indicated that some timestamps were skipped, duplicated, or otherwise incorrect. The SAS programs detected timestamp errors by calculating the size of the time steps, which is the time difference between adjacent observations in the datasets. Except at the beginning of trips, all time steps should be one second. The following discussion describes the general approach for detecting timestamp problems and describes how the timestamps were modified. In most instances, the time step was found to be one second. However, other time step values were also found.

The first step was to look for duplicate adjacent records in the datasets. These were records where the timestamp, speed, latitude, and longitude for a given vehicle were the same on two or more consecutive observations. For these instances the duplicated records were removed from the dataset. Occasionally, the same timestamp was found on consecutive observations but the speed values were different. In these cases, we incremented the timestamp on consecutive seconds to provide a continuous flow of time.

In the Battelle dataset, a large number of timestamps were found to be 01JAN04 which was the value that was logged when the GPS unit had not yet found its satellite signal. In these cases, we looked at the timestamps before and after these and entered a corrected timestamp.

Some time steps were found to be greater than one second during a trip. This indicated that records were missing. In these instances, in the Faucett dataset we inserted observations in the dataset with timestamps that would make the set of timestamps continuous during the trip and assigned missing speed values to those timestamps. In the Battelle dataset, we found time steps that were negative followed several observations later where the time step was positive to bring the following observations back to the correct time. In these instances we hand corrected this "drop out" period to have correct timestamps to which the speed values were assumed to be correctly assigned.

In the Battelle dataset we examined the frequency distribution of the time steps and found that there were a large number of time steps of exactly 30 seconds. This corresponded to the Battelle datalogging system going "dormant" during a perceived idle of the vehicle. During these periods, the Battelle datalogger recorded vehicle speed for one second every 30 seconds until the datalogger found the vehicle to be no longer idling. For these 30-second dormant periods, if the logged speed at the end of the dormant period was 0.00 mph, then we assigned the previous 29 seconds to have a speed of 0.00 mph and inserted the 29 observations with the appropriate timestamps. However, if the speed following the 30-second time step was greater then 0.00 mph we inserted the previous 29 observations with timestamps but assigned them to have missing speeds.

In the Battelle dataset, the distribution of time steps indicated a moderately large number of time steps with durations between 2 and 29 seconds. We believe these instances were from cases where the datalogger awoke from its dormancy during which the vehicle had already started moving. Since we could not tell from the data which seconds of those periods were idles and which were not idles, we had to assign all speed observations in those periods to missing values. We took the same approach for the relatively small number of periods with time steps greater then 30 seconds.

In both the Faucett and Battelle datasets, the insertion of timestamps for time steps that were greater then one second caused a large increase in the total number of observations in the datasets. In the Faucett dataset, the number of observations increased from 2.0 million to 3.6 million and for the Battelle dataset, it increased from 8.0 million to 13.7 million. In the Battelle dataset, the majority of the increased time was caused by the large number of 30 second dormancy periods during idles. However, in the Faucett dataset, the large increase in number of observations was caused by the insertion of a large number of timestamps for a relatively few large time steps. The largest time step was around 90,000 seconds.

As best we can determine, the beginning of trips (key on) was detected by the datalogging system by monitoring increases in noise and small changes in voltage levels of the 12-volt vehicle auxiliary power outlet. If an engine-on or engine-off event was not detected by the datalogging system, then the system could incorrectly assign a trip number to a portion of the data. The result could be a period of missing timestamps for a trip when actually the vehicle was between trips.

Accordingly, we compared the speeds before and after time steps greater than one second. We found two types of behavior. As the duration of time steps increased, the speeds just before and just after the time gap were more likely to be idle speeds. However, for short time steps greater then one second, the speeds before and after the time gap tended to be in good agreement and tended to not be idle speeds. Accordingly, for the Faucett data if the before speed was less than 12 miles per hour and the after speed was less then 15 miles per hour, we removed the observations with missing timestamps. This broke the trip designated by the datalogging system into two separate trips. If the before or after speeds were longer than these limits, then we left the inserted timestamps with missing speeds and the single trip remained as we had already corrected it. We, therefore, assumed that the missing time gap represented a continuing trip that simply had speed values missing. In the case of the Battelle dataset, the criteria were speeds before of less than 5 miles per hour and speeds after of less than 35 miles per hour.

3.2 Speed Value Flags and Vehicle Deletions

The data quality checking programs for the Faucett and Battelle datasets also were written to set suspect speed values to missing values. This was done individually for each of the vehicles in the datasets by examining the speed values on a plot of acceleration versus speed for each individual vehicle.

In the Faucett and Battelle datasets, we made plots of acceleration versus speed for each of the individual 171 vehicles. Usually, the plots showed high accelerations at low speeds and low accelerations at high speeds in the familiar triangular shape seen for light-duty vehicles. We arbitrarily drew upper and lower limit lines on these plots for each vehicle to help designate the points that appeared to be outliers. Observations where the speed and acceleration values were outside of these upper and lower limits were given a flag designation of A for acceleration that was too high or D for a deceleration that was too low.

However, for 20 of the 140 vehicles in the Battelle dataset, the acceleration versus speed plots indicated a large amount of noise in the speed values as shown by many extremely large accelerations and decelerations in the dataset. An examination of the speed versus time plots for portions of these vehicles indicated that the data from these vehicles was not useful for the purpose of generating cycles. Accordingly, these 20 vehicles were eliminated from further consideration in this project. In the case of the Faucett dataset, one vehicle (Vehicle 143) was found to have speed values that were excessively noisy. This vehicle was dropped from the Faucett dataset.

In the Faucett dataset, in addition to the flagging of certain observations with A and D, we noticed that observations adjacent to these flagged observations were sometimes "stuck" at constant speed values. Thus, the Faucett dataset exhibited periods of datalogging failure that were made up of different combinations and orders of high acceleration, high deceleration, and stuck speeds. Therefore, in the Faucett dataset, we also included a flag of Z for accelerations of exactly 0.00 mph/s when the speed was not equal to 0 mph. When we examined the groups of observations that were contiguous in flags of A, D, and Z we found these to be periods in which the speed values were unquestionably erroneous. We set the speed values during these identified periods to missing.

We also found a large number of periods in the Battelle data with "stuck" speeds, but we did not change speeds of the affected observations to be missing since such a large portion of the Battelle data had 0.00 mph/s arising from the lower speed resolution (0.11 mph) of the datalogging system. In the case of the Battelle dataset, any observation that had an A or D flag had its speed changed from its reported value to a missing value, but the periods of stuck speed were left as reported.

After the timestamp insertions and vehicle deletions, the Faucett dataset had 2,125,097 observations and the Battelle dataset had 10,714,023 observations. The TxDOT dataset had 709,581 observations. The combined data from these three datasets were used to build the cycles for this study.

3.3 Idle Designations

Because the cycle development process is based on building up candidate cycles from micro-trips, which are defined by idle periods, it is important to correctly designate what observations are idles. Because of the presence of dithering in the GPS speed values, non-zero speeds are almost always reported in the Battelle and Faucett datasets – even when the vehicles are not moving. Consequently, we had to develop a method for designating when a vehicle was at idle. Our examination of the Faucett and Battelle data on an acceleration versus speed plot showed that a large peak in observations occurred at 0.6 mph and 0.00 mph/s which is near the expected idle values of 0.00 mph and 0.00 mph/s. This finding caused us to develop a probabilistic method of estimating whether an individual observation represented an idle or a non-idle condition for the vehicle.

The method is based on the frequency of observations that occur in speed/acceleration bins for each individual vehicle in the vicinity of idle conditions. The vicinity that we used was for speeds between 0 and 10 mph and for accelerations between –0.5 and +0.5 mph/s. (The 10 mph value was used for all vehicles in the Faucett dataset. Different values were used for different vehicles in the Battelle dataset.) In this region, we counted the number of observations in bins that were 0.2 mph wide and 0.02 mph/s wide. The bin with the maximum frequency was assigned a probability value of 1.00. Other bins in this vicinity were given probabilities in proportion to their frequencies relative to the maximum frequency. Speeds that were greater than 10 mph or accelerations greater than +0.5 mph/s acceleration or less than –0.5 mph/s were assigned probabilities of zero. All observations in the dataset were assigned individual probabilities (variable name: p_i) that corresponded to the probability assigned to the speed acceleration bin into which that observation fell. This resulted in observations having probabilities assigned to them of being an idle observation.

When we examined speed versus time plots for these assigned probabilities, we found that there was a lot of noise in the assigned probabilities from second-to-second because of small changes in accelerations from point to point. Consequently, we calculated a rolling average probability (variable name: p_i7) for each observation by calculating the joint probabilities for the current observation and the three probabilities in the previous three observations and the three probabilities in the following three observations. This provided a smoothing of probabilities. Special code was written for calculating the joint probabilities at the beginning and end of trips since, in those locations, three seconds before and three seconds after the current observation do not always exist.

Our examination of the joint probabilities and speed as a function of time indicated that it was quite possible to make a reasonable separation of observations into non-idle and idle observations. To determine the value of the threshold that should be used to separate the probabilities into idle and non-idle, we made a frequency distribution of the joint probabilities for each vehicle in the two datasets. We found that a minimum in the distribution was observed near a joint probability of 0.45 for all of the vehicles. Accordingly, we used this value to separate the idle from non-idle observations. This separation was manifested as a flag variable called idle_mark that had a value of one when the observation was believed to be an idle condition and zero when the observation was believed to be a non-idle condition.

3.4 Trip and Micro-Trip Designations

The edited data from the TxDOT, Faucett, and Battelle datasets were next combined and prepared for use in cycle development. This preparation was done by the prep.sas program. First, the data from all three sets were put into one dataset. Next, the reported speeds that were believed by the qc_bl.sas program to be idles (idle_mark=1) were changed to have values of 0.00 mph. This was done only for the Faucett and Battelle data since the TxDOT data already had measured speeds of zero for idle periods.

Once the speeds at idle were set to zero, the designation of trips and micro-trips could be done. The beginning of trips occurred:

- At the first observation of the dataset;
- When the vehicle changed; or
- When the time step between sequential time stamps was greater than 1 second or less than –1 second.

Whenever the program detected the start of a new trip using these criteria, the trip number was incremented. Trip numbers for the entire dataset were unique.

The observations were marked for the beginning of a micro-trip:

- If a new trip began;
- If the current speed was zero and the previous second's speed was non-zero.

Whenever a new micro-trip was detected, the micro-trip number was incremented. Micro-trip numbers for the entire dataset were unique.

Finally, the data from all three datasets with the new flags were written to a permanent dataset for use by the cycle development programs. The dataset variables included study, vehicle ID, trip number, trip seconds, trip start flag, micro-trip number, micro-trip seconds, micro-trip start flag, the date/time stamp, the corrected speed to be used for cycle building, the original raw speed from the original datasets, and a flag that gave the reason that the original speed was changed. Table 3-1 provides a list of the different flag values and their meanings.

Table 3-1. Missing Value Flag Definitions

Value of Flag	Action Taken
FIX1	Inserted a missing observation. Provided the appropriate timestamp value. Speed value was set to missing.
FIX10	Corrected the timestamp and left the reported speed as is. Occurred when the raw timestamp value was duplicated, but the raw speed value was not duplicated.
FIX13	For Battelle data only where the time step was 30 seconds. Inserted 29 seconds with appropriate timestamps. If the reported speed on the thirtieth second was 0.00 mph, then the previous 29 1-second inserted speed values were set to 0.00 mph. Otherwise, the inserted speed values were set to missing.
FIX2*	Same explanation as for the FIX1* values above, but the action was taken on the second pass through the dataset.
A	The indicated acceleration was greater than the high acceleration limit for that vehicle. The speed value was set to missing.
D	The indicated acceleration was less then the low acceleration limit for that vehicle. The speed value was set to missing.
P	For Faucett data only. Suspect observations near A and D flag values that were part of a pattern of speed value excursions associated with "stuck" speed values. The speed values were set to missing.
JAN04	For Battelle data only. When the reported timestamp was 01JAN04 (GPS unit lost satellite signal). The timestamp was set to an appropriate value and the speed was set to missing.
SPN	A time step was negative and then several seconds later it was positive by just an amount that put the timestamp back where it should have been. All timestamps during this period were hand edited and speed values were left as reported.

4.0 Selection of Cycles to be Developed

In this study, the heavy-duty vehicle activity data was used to develop individual cycles for the operation of vehicles for different combinations of vehicle type/usage, freeway/non-freeway operation, and average micro-trip speeds. In this study these different combinations are called cases. The following subsections describe the analysis of the edited heavy-duty activity dataset to arrive at descriptions of the different cycles that were developed. The vehicle type/usage designations and freeway/non-freeway trip designations were arrived at based on definitions suggested by EPA. The selection of micro-trip average speed bins was based on an analysis of the heavy-duty vehicle activity dataset.

4.1 Vehicle Type/Usage Designations

As specified by EPA, the 154 heavy-duty vehicles in the activity dataset were divided into three categories:

- **Heavy heavy-duty vehicles –** these vehicles had gross vehicle weight ratings of 33,001 lbs. and greater;
- **Non-parcel medium heavy-duty vehicles –** these vehicles had gross vehicle weight ratings from 19,501 to 33,000 lbs. and the vehicles were those that were not used for postal/parcel service;
- **Parcel medium heavy-duty vehicles –** these vehicles had gross vehicle weight ratings from 19,501 to 33,000 lbs. and were specifically designated as being used for postal/parcel service. The only vehicles that fell into this category were a portion of those in the Battelle dataset.

The vehicle type designations for all 154 vehicles that were used to develop cycles in this study are shown in Appendix A.

4.2 Freeway Micro-Trip Designations

Micro-trips were designated as freeway or non-freeway micro-trips. EPA had decided to designate micro-trips with regard to freeway use based on the distance of the micro-trip. Micro-trips that had a total distance of greater than or equal to 3 miles were designated as freeway micro-trips. The freeway designation is more a designation of whether a micro-trip was involved in stop-and-go driving rather than an actual verification that a given micro-trip occurred on a freeway. While the GPS latitude and longitude information could be used to determine which observations were associated with the presence of a vehicle on a freeway, this approach was not taken in this study.

4.3 Micro-Trip Average Speed Bins

The heavy-duty vehicle activity dataset was analyzed to arrive at several speed bins for the six different combinations of vehicle type/usage and freeway/non-freeway designations. First, all of the micro-trips in a dataset were designated for vehicle type/usage and freeway/non-freeway operation. Then, the average vehicle speed for all of the micro-trips in each of the six combinations of vehicle type/usage and freeway/non-freeway operation were calculated. The micro-trip average vehicle speeds were then binned into the average speed bins that will be used for the new MOBILE model. These speed bins were created by rounding the average speed to the nearest 5 mph. Table 4-1 shows the distributions of the binned average micro-trip speeds for the six different combinations of vehicle type and freeway/non-freeway operation. From these distributions, the final description of cycles to be developed in terms of combinations of vehicle type/usage, freeway/non-freeway operation, and micro-trip average speed are shown in Table 4-2.

The goal in creating the different cases was to combine adjacent average speed bins such that each combination of vehicle type and freeway/non-freeway operation had five or six speed bins associated with it. In addition, each of the cases needed to have a relatively large number of micro-trips so that the typical operation was well defined and so that the cycle development software had a sufficiently large number of micro-trips to choose from to build the cycle.

Table 4-1. Distribution of Binned Average Micro-Trip Speeds

Vehicle Type	HHDV	HHDV	MHDV Non-Parcel	MHDV Non-Parcel	MHDV Parcel	MHDV Parcel	
Operation	Non-Freeway	Freeway	Non-Freeway	Freeway	Non-Freeway	Freeway	
Rounded Average Micro-Trip Speed (mph)							Total
0	16993	0	3926	0	9318	0	30237
5	4561	2	1362	0	4712	0	10637
10	2461	2	887	1	3942	0	7293
15	1610	12	829	2	3370	0	5823
20	1317	29	758	5	2320	2	4431
25	944	66	497	13	1239	9	2768
30	510	129	224	40	586	18	1507
35	167	208	53	41	227	70	766
40	62	291	14	67	112	71	617
45	16	429	3	94	36	88	666
50	3	616	0	87	7	95	808
55	1	534	0	74	1	61	671
60	2	172	0	34	0	39	247
65	0	20	0	2	0	7	29
70	0	0	0	1	0	0	1
Total	28647	2510	8553	461	25870	460	66501

Table 4-2. Final Descriptions of Cases

Case Name	Vehicle Type/Usage	Operation	Micro-Trip Average Speed Bin	Speed (mph) Bin Definition	Number of Micro-Trips
H_0_5	HHDV	Non-Freeway	5	Avg Micro-Trip Speed < 7.5	14911
H_0_10			10	7.5 ≤ Avg Micro-Trip Speed < 12.5	2461
H_0_15			15	12.5 ≤ Avg Micro-Trip Speed < 17.5	1610
H_0_20			20	17.5 ≤ Avg Micro-Trip Speed < 22.5	1317
H_0_25			25	22.5 ≤ Avg Micro-Trip Speed < 27.5	944
H_0_30			30	27.5 ≤ Avg Micro-Trip Speed	761
H_1_30		Freeway	30	Avg Micro-Trip Speed < 35	337
H_1_40			40	35 ≤ Avg Micro-Trip Speed < 45	595
H_1_50			50	45 ≤ Avg Micro-Trip Speed < 55	1157
H_1_60			60	55 ≤ Avg Micro-Trip Speed	421
N_0_5	MHDV Non-Parcel	Non-Freeway	5	Avg Micro-Trip Speed < 7.5	4571
N_0_10			10	7.5 ≤ Avg Micro-Trip Speed < 12.5	887
N_0_15			15	12.5 ≤ Avg Micro-Trip Speed < 17.5	829
N_0_20			20	17.5 ≤ Avg Micro-Trip Speed < 22.5	758
N_0_25			25	22.5 ≤ Avg Micro-Trip Speed < 27.5	497
N_0_30			30	27.5 ≤ Avg Micro-Trip Speed	294
N_1_30		Freeway	30	Avg Micro-Trip Speed < 35	81
N_1_40			40	35 ≤ Avg Micro-Trip Speed < 45	136
N_1_50			50	45 ≤ Avg Micro-Trip Speed < 55	171
N_1_60			60	55 ≤ Avg Micro-Trip Speed	73
P_0_5	MHDV Parcel	Non-Freeway	5	Avg Micro-Trip Speed < 7.5	10127
P_0_10			10	7.5 ≤ Avg Micro-Trip Speed < 12.5	3941
P_0_15			15	12.5 ≤ Avg Micro-Trip Speed < 17.5	3370
P_0_20				17.5 ≤ Avg Micro-Trip Speed < 22.5	2318
P_0_25				22.5 ≤ Avg Micro-Trip Speed < 27.5	1239
P_0_30			30	27.5 ≤ Avg Micro-Trip Speed	968
P_1_30		Freeway	30	Avg Micro-Trip Speed < 35	50
P_1_40			40	35 ≤ Avg Micro-Trip Speed < 45	165
P_1_50			50	45 ≤ Avg Micro-Trip Speed < 55	175
P_1_60			60	55 ≤ Avg Micro-Trip Speed	70
Total					55234

5.0 Cycle Development

The cycles to be built for heavy-duty vehicle operation will ultimately be used in EPA's new MOBILE model of vehicle emissions for heavy-duty vehicles. The idea of a cycle is that it contain the essence of heavy-duty vehicle driving behavior. To make a representative cycle practical, the cycle should be relatively short so that it does not take up a large amount of memory in the model. Since the heavy-duty vehicle activity database contains second-by-second data for 154 vehicles over a considerable amount of time, the key challenge for the cycle builder is to compress the dataset to produce a reasonably short cycle while maintaining the essence of the heavy-duty vehicle driving behavior. Such a short cycle can then be used by the model to calculate emissions of a heavy-duty vehicle.

Representative cycles can be built using different methodologies. The methodology we have chosen for this study's cycles is to use pieces of real driving, called micro-trips, from the heavy-duty activity database, which when connected together can be expected to have similar emissions behavior to heavy-duty vehicles driving on the road. However, the emissions behavior of different vehicles with different technologies – even future technologies – will differ. Accordingly, we cannot create cycles based directly on emissions behavior. Instead, the cycles will be built around parameters of vehicle operation and usage that are known or expected to be important to exhaust emissions of heavy-duty vehicles. By using this approach of matching vehicle operation between measured driving behavior and candidate cycles, it can be inferred that the emissions behavior of vehicles over the cycles will be similar to the emissions behavior of heavy-duty vehicles on the road.

In the creation of these heavy-duty cycles, we have chosen vehicle speed, acceleration, and vehicle specific power (VSP) as the variables that are important to the exhaust emissions of heavy-duty vehicles. All three of these variables together provide a measure of the load on the engine, which is an important variable associated with exhaust emissions. In this study, we are building cycles only for warmed up operation of heavy-duty vehicles. That is, we are not building special cycles for cold starts and warm starts. We assume that all data in the datasets represent warmed-up driving.

5.1 General Methodology

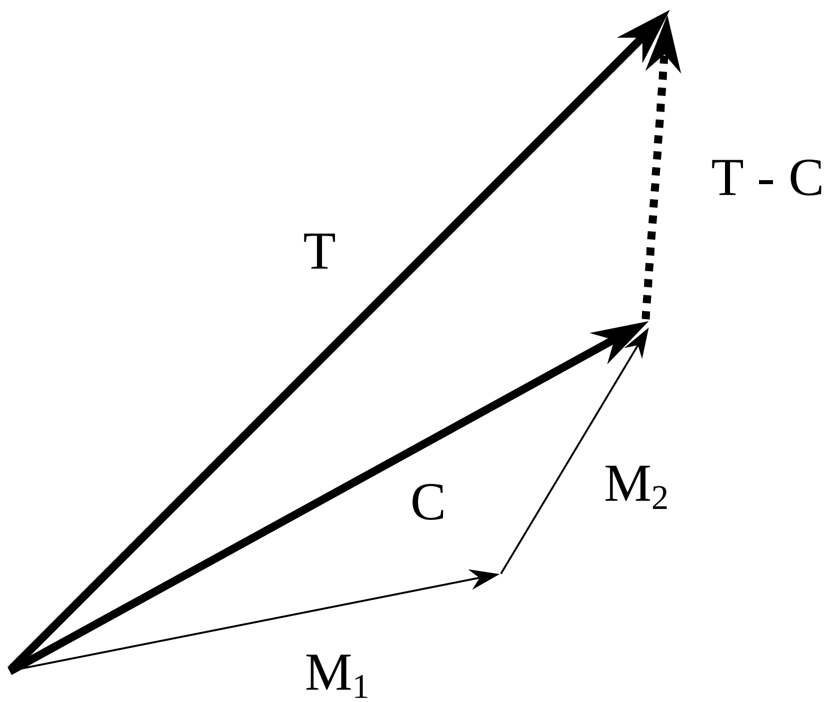
The cycles were created by combining micro-trips of actual driving. Each cycle should be a good representation of the driving behavior in the dataset. The three critical variables (speed, acceleration, vehicle specific power) were used for selection of micro-trips. The speed of each vehicle was measured directly in the TxDOT, Faucett, and Battelle datasets. The acceleration of the vehicle for each second was estimated as the derivative of the speed. The vehicle specific power for each second was estimated from the speed and acceleration as described in Section 5.3.1.

To identify specific segments of vehicle driving for inclusion in the cycle, the entire activity dataset was converted to a set of micro-trips. A micro-trip is defined as a contiguous speed trace of vehicle driving and is made up of an idle followed by all non-idle driving until the next idle begins. A single vehicle trip may be composed of numerous micro-trips.

A strategy based on a minimizing the difference between a cycle vector **C** representing the driving in the candidate cycle and a target vector **T** representing the driving in the activity database for the case was used to select micro-trips from the database for inclusion in the cycle. As micro-trips are used to build-up a candidate cycle, the difference between the two vectors tends to become smaller and smaller. The build-up process ends when the cycle developer decides that the two vectors are substantially the same and the duration of the cycle that has been built up is acceptable. The multi-dimensional space that these vectors are in will be described shortly, but first let us consider how the build-up process works for developing a cycle.

The goal of building the cycle is to select micro-trips such that when their vectors $\mathbf{M_i}$ are added together, the vector **C** of the resulting cycle is as similar as possible to the target vector **T** of the activity database. Figure 5-1 shows the hypothetical situation of the vectors after two micro-trips have been used to create a cycle. In this hypothetical example, the first micro-trip was selected from the activity database for the case as the one whose vector $\mathbf{M_1}$ was closest to the target vector **T** for the database. Then, a second micro-trip is searched for such that when its vector $\mathbf{M_2}$ is added to $\mathbf{M_1}$ to create the resultant vector **C** shown in Figure 6-1, the distance between the tips of **C** and **T** is minimized. This distance is the length of the vector **T-C** as denoted in the figure by the dashed vector. As micro-trips are added to create the built-up cycle represented by **C**, the length of **T-C** is calculated after each additional micro-trip is added to the cycle to follow the progress of the build-up process. It should be noted that the order of the micro-trips in the final cycle is unimportant from the point of view of the selection of the micro-trips. The reason for this is that the resultant **C** is independent of the order in which the micro-trip vectors $\mathbf{M_i}$ are added together.

Figure 5-1. Vector Description of Comparing Target and Cycle Activity



It should also be noted that we are forcing micro-trips to be added to the candidate cycle. This is done even if the addition of the best incremental micro-trip causes the length of **T-C** to increase in some instances. Generally, as the cycle is built up there will be a decrease in the length of **T-C**. After several micro-trips have been added, the length of **T-C** may increase slightly. Later, with the addition of more micro-trips, a "discovery" will be made that will produce a relatively abrupt decrease in the length of **T-C** so that the accumulated cycle will be substantially better than the cycle was much earlier in the build-up process.

All of the vectors used above to describe the build-up process are based on representations of the frequency distributions of observations in cumulative speed, acceleration, vehicle specific power space. This statement requires some explanation. A segment of driving, whether it is a micro-trip, a piece of a driving cycle, or the entire activity database can be described as a frequency distribution. The distribution consists of combinations of the three variables: speed, acceleration, and vehicle specific power. The continuous values for each of these variables were converted into frequency distributions through the use of bins. Each observation in the database was placed in a particular speed/acceleration/VSP bin. The <u>cumulative</u> frequency distribution is made up of the number of observations that fall "below" the current bin for each of the three-binned variables. The binning criteria for each of the three variables are described in Section 5.3.2. To help the reader understand the process, we will present a numerical example in one dimension and another example in two dimensions to demonstrate how the comparison of the vectors **T** and **C** works.

Suppose we wanted to compare a candidate cycle with the database using a single vehicle operation variable that was monitored second-by-second in the collection of data for the activity database. The single variable might be engine load. In this hypothetical example, we have 35,900 one-second observations of engine load in the target activity database and 68 one-second observations in the cycle. The first step in comparing **T** and **C** is to bin the observations of load in the target data and in the cycle data. Table 5-1 shows the binning of the hypothetical data in Columns 2 and 3. Note that the number of observations in the target data in Column 2 is much higher then the number of observations in the cycle data in Column 3. This is a consequence of the activity database containing all of the observations for all micro-trips but the cycle has just one micro-trip. The frequency counts in Columns 2 and 3 are then converted to cumulative frequency counts in Columns 4 and 5. This is done to provide proximity information for the micro-trip searching algorithm. In other words, we wanted to the algorithm to be able to select a micro-trip even if the observations for a given micro-trip were not in exactly the same bins as the target but did have observations at least in a nearby bin. The use of the cumulative distributions helps ensure that proximity information is available.

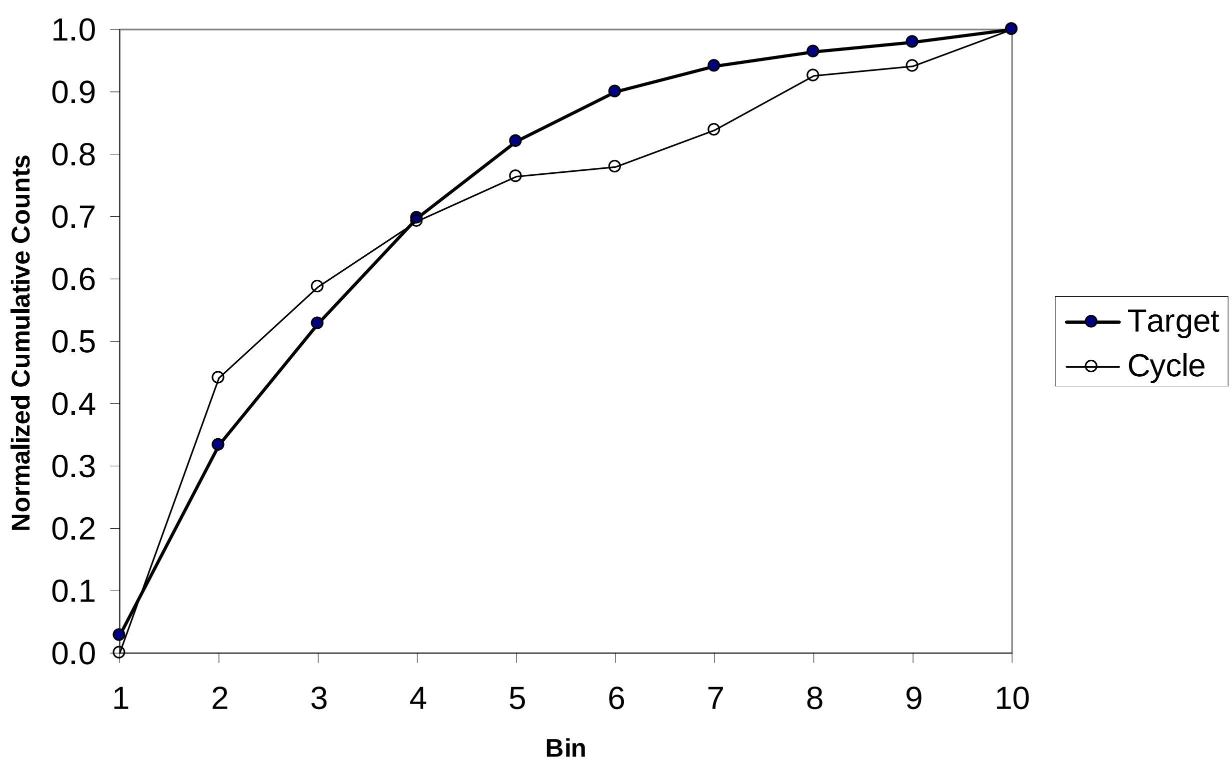
Table 5-1. Comparison of Cycle and Target Vectors for a Hypothetical One-Dimensional Example

	Counts		Cumulative Counts		Vector (Normalized Cumulative Counts)		Vector Length		
Bin	**Target**	**Cycle**	**Target**	**Cycle**	**Target**	**Cycle**	**T**	**C**	**T-C**
1	1000	0	1000	0	0.028	0.000	1.246	1.266	0.138
2	11000	30	12000	30	0.334	0.441			
3	7000	10	19000	40	0.529	0.588			
4	6000	7	25000	47	0.696	0.691			
5	4500	5	29500	52	0.822	0.765			
6	2800	1	32300	53	0.900	0.779			
7	1500	4	33800	57	0.942	0.838			
8	800	6	34600	63	0.964	0.926			
9	600	1	35200	64	0.981	0.941			
10	700	4	35900	68	1.000	1.000			

A comparison of the cumulative counts for the target and cycle information in Columns 4 and 5 shows that if we used these counts to create the **T** and **C** vectors, the lengths of the vectors would be greatly different simply because the target vector, which is made up of the 10 elements in Column 4, would be a much longer vector then the cycle vector, which is made up of the 10 elements in Column 5. Accordingly, we normalize the target and cycle cumulative counts in 4 and 5 to produce the target vector elements and the cycle vector elements as the fractional values between 0 and 1 shown in Columns 6 and 7.

The values in Columns 6 and 7 become the elements of the **T** and **C** vectors, which are in 10-dimensional space. A visualization of the elements of these vectors is provided in Figure 5-2. This figure shows the normalized cumulative counts of the target and cycle from Columns 6 and 7 as a function of the bin number. What we want to do in developing the cycle is select micro-trips so that the curve for the cycle is as close as possible to the curve for the target in this figure. The way we do this is to minimize the sums of the squares of the differences between the value for the corresponding elements of the target and cycle vectors. This corresponds to the square of the length of **T-C**. Table 5-1 shows the calculated length of **T**, **C**, and **T-C**. These lengths can be determined from the values of the elements for **T** and **C** in Columns 6 and 7 using the standard relationship for determining the length of a vector if its elements are known.

Figure 5-2. Visual Comparison of Vector Elements



Extension of the one-dimensional example shown in Table 5-1 and Figure 5-2 to multiple dimensions is demonstrated by the spreadsheet calculations shown in Table 5-2. In this example, 100 matrix elements are used. The table shows 10 rows which might be accelerations and 10 columns which might be speeds. The left side of Table 5-2 shows the calculations for the target matrix and the right side shows the calculations for the cycle matrix. In Tables a) and b), the second-by-second observations of the target and cycle data are binned. The numbers in each bin represent the frequency of observations that meet the criteria for those bins. In Tables c) and d), the counts in the Tables a) and b) are accumulated across each row. Then, in Tables e) and f), the accumulated frequencies in Tables c) and d) are accumulated down each column. This produces a field of frequencies on a cumulative basis that run from a low value in the upper left corner of each matrix to a high number in the lower right corner of each matrix. The value in the lower right hand corner of Tables e) and f) is equal to the total number of observations in the target or cycle matrix. These total observation numbers in the lower right hand corner of e) and f) are used to normalize all of the frequencies in Tables e) and f) to arrive at the normalized cumulative matrices in g) and h). The values in g) and h) are then used to calculate the square of the differences in each corresponding matrix element to produce the values in Table i). The value in Table j) is just the summation of all of the elements of Table i) and represents the square of the length of the **T-C** vector. This is the value that we attempt to minimize when selecting micro-trips for the cycle.

Note that the counts in a) and b) did not need to be in corresponding bins for this comparison process to work. The use of cumulative distributions permitted the two matrices to be compared successfully.

Extension of the technique to the third dimension for vehicle specific power or any number of higher dimensions is made by analogy.

5.2 Generation of Alternative Candidate Cycles

The implementation of the cycle development methodology is provided by three computer programs: makemicro.sas, findcycle.f, and makecycle.sas, which are run sequentially. The important details of what the three programs do are described in the following subsections. However, what they do in general is described here.

Table 5-2. Comparison of Cycle and Target Matrices for a Hypothetical Two-Dimensional Example

Target Activity Matrix

a) Count the second-by-second observations in each bin.

	A	B	C	D	E	F	G	H	I	J
1	2									
2		1								
3		2		5						
4			5		3		2	1		
5		5		9	1			2	9	3
6			2			4	1			
7										
8			6			1				
9		1								
10										

c) Accumulate the above frequencies across each row

1	2	2	2	2	2	2	2	2	2	2
2	0	1	1	1	1	1	1	1	1	1
3	0	2	2	7	7	7	7	7	7	7
4	0	0	5	5	8	8	10	11	11	11
5	0	5	5	14	15	15	15	17	26	29
6	0	0	2	2	2	6	7	7	7	7
7	0	0	0	0	0	0	0	0	0	0
8	0	0	6	6	6	7	7	7	7	7
9	0	1	1	1	1	1	1	1	1	1
10	0	0	0	0	0	0	0	0	0	0

e) Accumulate the above frequencies down each column.

1	2	2	2	2	2	2	2	2	2	2
2	2	3	3	3	3	3	3	3	3	3
3	2	5	5	10	10	10	10	10	10	10
4	2	5	10	15	18	18	20	21	21	21
5	2	10	15	29	33	33	35	38	47	50
6	2	10	17	31	35	39	42	45	54	57
7	2	10	17	31	35	39	42	45	54	57
8	2	10	23	37	41	46	49	52	61	64
9	2	11	24	38	42	47	50	53	62	65
10	2	11	24	38	42	47	50	53	62	65

g) Normalize the elements in the above matrix.

1	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
2	0.031	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046
3	0.031	0.077	0.077	0.154	0.154	0.154	0.154	0.154	0.154	0.154
4	0.031	0.077	0.154	0.231	0.277	0.277	0.308	0.323	0.323	0.323
5	0.031	0.154	0.231	0.446	0.508	0.508	0.538	0.585	0.723	0.769
6	0.031	0.154	0.262	0.477	0.538	0.600	0.646	0.692	0.831	0.877
7	0.031	0.154	0.262	0.477	0.538	0.600	0.646	0.692	0.831	0.877
8	0.031	0.154	0.354	0.569	0.631	0.708	0.754	0.800	0.938	0.985
9	0.031	0.169	0.369	0.585	0.646	0.723	0.769	0.815	0.954	1.000
10	0.031	0.169	0.369	0.585	0.646	0.723	0.769	0.815	0.954	1.000

Cycle Activity Matrix

b) Count the second-by-second observations in each bin.

A	B	C	D	E	F	G	H	I	J
	1								
			4						
	4				3		1		
							4	1	
			8						2
				3					
						1			
1	5								

d) Accumulate the above frequencies across each row

0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	1	1
0	0	0	4	4	4	4	4	4	4
0	4	4	4	4	7	7	8	8	8
0	0	0	0	0	0	0	4	5	5
0	0	0	8	8	8	8	8	8	10
0	0	0	0	3	3	3	3	3	3
0	0	0	0	0	0	1	1	1	1
1	6	6	6	6	6	6	6	6	6
0	0	0	0	0	0	0	0	0	0

f) Accumulate the above frequencies down each column.

0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	1	1
0	1	1	5	5	5	5	5	5	5
0	5	5	9	9	12	12	13	13	13
0	5	5	9	9	12	12	17	18	18
0	5	5	17	17	20	20	25	26	28
0	5	5	17	20	23	23	28	29	31
0	5	5	17	20	23	24	29	30	32
1	11	11	23	26	29	30	35	36	38
1	11	11	23	26	29	30	35	36	38

h) Normalize the elements in the above matrix.

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026
0.000	0.026	0.026	0.132	0.132	0.132	0.132	0.132	0.132	0.132
0.000	0.132	0.132	0.237	0.237	0.316	0.316	0.342	0.342	0.342
0.000	0.132	0.132	0.237	0.237	0.316	0.316	0.447	0.474	0.474
0.000	0.132	0.132	0.447	0.447	0.526	0.526	0.658	0.684	0.737
0.000	0.132	0.132	0.447	0.526	0.605	0.605	0.737	0.763	0.816
0.000	0.132	0.132	0.447	0.526	0.605	0.632	0.763	0.789	0.842
0.026	0.289	0.289	0.605	0.684	0.763	0.789	0.921	0.947	1.000
0.026	0.289	0.289	0.605	0.684	0.763	0.789	0.921	0.947	1.000

i) Calculate the squares of the differences in corresponding elements of the above two matrices.

	A	B	C	D	E	F	G	H	I	J
1	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
2	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.001	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.001	0.003	0.000	0.000	0.002	0.002	0.000	0.000	0.000	0.000
5	0.001	0.000	0.010	0.044	0.073	0.037	0.050	0.019	0.062	0.087
6	0.001	0.000	0.017	0.001	0.008	0.005	0.014	0.001	0.021	0.020
7	0.001	0.000	0.017	0.001	0.000	0.000	0.002	0.002	0.005	0.004
8	0.001	0.000	0.049	0.015	0.011	0.010	0.015	0.001	0.022	0.020
9	0.000	0.014	0.006	0.000	0.001	0.002	0.000	0.011	0.000	0.000
10	0.000	0.014	0.006	0.000	0.001	0.002	0.000	0.011	0.000	0.000

j) Sum the squares of the differences.

0.754

The first program, makemicro.sas, reads in the edited second-by-second data that is to be used to generate cycles. The data that is read in has already had the micro-trips designated by prep.sas. Each micro-trip is assigned to one of the 30 cases for which an individual cycle will be produced. Then, makemicro.sas estimates the vehicle specific power for every observation in the database based on the database values for speed and acceleration and the coefficients provided by EPA to estimate the effects of aerodynamic drag and rolling resistance on the vehicle specific power. Next, the continuous values for speed, acceleration, and VSP are binned through a rounding process. For each of the micro-trips in a case and for all micro-trips in a case taken together, the program counts the number of one-second observations that are spent in each speed/acceleration/VSP bin. At this point, makemicro.sas outputs the following variables for use by findcycle.f:

- Case ID;
- Micro-trip number;
- Speed bin;
- Acceleration bin;
- VSP bin; and
- Count, which is the number of observations in the speed/acceleration/VSP bin.

The next program, findcycle.f, picks up the dataset produced by makemicro.sas which contains the counts of observations in each speed/acceleration/VSP bin for all micro-trips in all of the 30 cases. The job of findcycle.sas is to use the micro-trip information to find those micro-trips which when concatenated best describe the activity for each case. In this study, we match the <u>non-idle</u> portions of the cycles to the <u>non-idle</u> driving in the activity dataset as the process of selecting micro-trips to build cycles. For each case, the overall activity is described by the sum of the counts in the speed/acceleration/VSP bins for <u>all</u> micro-trips that fall in that case. However, the micro-trips that are eligible for being used in the cycle to describe that target activity are selected from a subset of the micro-trips in the case.

The program works like this. Each case is considered separately and no micro-trips for one case are used to provide a cycle for another case. First, the program finds the best micro-trip whose sum of the squares difference between the cumulative normalized elements of the micro-trip with the corresponding elements of the target is the smallest. This corresponds to finding the micro-trip such that the **T-C** vector is the smallest. This becomes the first micro-trip in the cycle. Then, the program looks through all remaining micro-trips to find the best second micro-trip such that when it is added to the first micro-trip the new vector **T-C** has a minimum length.

This process is repeated until the developer wants to stop searching. In this project, we stopped searching after 25 micro-trips were added to the cycle.

The program, findcycle.f, outputs a list of the 25 selected micro-trips in the order in which they were selected for each of the 30 cases. In addition, the program provides a square of the length of **T-C** vector, the number of seconds in each of the micro-trips, and an accumulated total of the number of seconds in the cycle as the micro-trips were selected and added to the candidate cycle being built up.

The third program, makecycle.sas, uses the output of findcycle.f for each case to visualize each cycle using plots of cycle speed versus time, acceleration versus speed for the target and cycle, VSP versus speed for the target and cycle, and various statistics on the individual micro-trips in each cycle. One of these statistics is the number of missing values in each of the selected micro-trips. Micro-trips that have a large number of missing speed, acceleration, or VSP values may be difficult to repair. Accordingly, we arbitrarily decided to accept no micro-trips that had more then 10 to 15 missing speed values. When micro-trips with large numbers of missing values were selected by the program, we edited findcycle.f so that those particular micro-trips would be skipped. Then, findcycle.f and makecycle.sas were run again iteratively until all micro-trips that were selected by makecycle.sas contained less then 10 to 15 missing values in each micro-trip.

5.3 Specific Details of Cycle Generation

There are several areas of generating the cycles that need to be explained in some additional detail. These areas are described in more detail in the following subsections.

5.3.1 Estimation of Vehicle Specific Power

Before all three continuous variables (speed, acceleration, and VSP) can be binned, the vehicle specific power variable needed to be calculated. This calculation was done in makemicro.sas based on input from EPA. EPA provided the following VSP equation to be used with the coefficients in Table 5-4. The first term in the equation is for the rolling resistance, the second term is a correction for rolling friction and rotational inertia at higher speeds, the third term is for aerodynamic drag, the fourth term is for accelerating the mass of the vehicle, and the fifth term is for changing the potential energy of the vehicle as it moves up and down road grades. The road load coefficients given in Table 5-4 are provided for three different vehicle weight ranges. Since all the vehicles in the dataset are medium and heavy heavy-duty vehicles, only the coefficients for the two upper vehicle weight ranges were used. Note that the coefficient for B/M is zero. This causes the second term of the VSP equation to provide no contribution to the calculated VSP values. In addition, all VSP calculations assume that the road grade was level. This causes the last term in the equation to be zero so that it also does not contribute to the calculated VSP values.

$$VSP = \left(A/M\right) \bullet v + \left(B/M\right) \bullet v^2 + \left(C/M\right) \bullet v^3 + a \bullet v + g \bullet v \bullet \sin\theta$$

where:

VSP	=	vehicle specific power (kW/Mg or W/kg)
v	=	vehicle speed (m/s)
a	=	vehicle acceleration (m/s^2)
g	=	acceleration of gravity (9.8 m/s^2)
θ	=	road grade

Table 5-4. Road Load Coefficients for the VSP Equation

	Vehicle Weight Range		
	8500 to 14000 lbs (3.855 to 6.350 tonne)	14000 to 33000 lbs (6.350 to 14.968 tonne)	>33000 lbs (>14.968 tonne)
A(kW*s/m)/M(tonne)	(0.4777/5.1 =) 0.094	(0.7652/10.7=) 0.072	(1.188/15=) 0.08
B(kW*s^2/m^2) /M(tonne)	0	0	0
C(kW*s^3/m^3)/M(tonne)	(2.037 x 10^{-3}/5.1=) 0.40 x 10^{-3}	(3.52 x 10^{-3}/10.7=) 0.33 x 10^{-3}	(4.93 x 10^{-3}/5.1=) 0.97 x 10^{-3}

Since the actual weight of the vehicles as they were loaded during the TxDOT, Faucett, and Battelle studies were unknown, EPA assumed that for the purposes of calculating the road load coefficients A/M and C/M that the average weight of the vehicles during data collection was near the middle of the gross vehicle weight range. This assumption is demonstrated by the parenthetical calculations shown in the cells of Table 5-4.

5.3.2 Binning of Continuous Variables

To use the cycle development approach discussed above, all of the micro-trips in the edited dataset needed to have all of their second-by-second observations binned in terms of speed, acceleration, and vehicle specific power. While the size of the bins is arbitrary, bins in general need to be narrow enough to resolve important emissions effects. In addition, bins need to be sufficiently narrow to distinguish different micro-trips for low speed, low acceleration, and low VSP micro-trips where those variables do not vary over a large range. On the other hand, from a practical perspective, the number of bins needs to be small so that the program that selects micro-trips can run in a reasonable amount of time.

For the cycle development in this project, we used the following binning schemes:

- **Speed –** the continuous speed values in miles per hour were ceilinged up to the next whole number. For example, 5.6 miles per hour was assigned to bin 6, 5.2 miles per hour was assigned to bin 6, 5.001miles per hour was assigned to bin 6, 5.000 miles per hour was assigned to bin 5.
- **Acceleration –** Acceleration values in miles per hour per second were ceilinged just as the speed values were.
- **Vehicle Specific Power –** VSP values in kW/Mg were rounded to the nearest 5 kW/Mg.

When these bin definitions were used, the counts of observations for the entire edited dataset were found to be distributed for speed, acceleration, and vehicle specific power as shown in Tables 5-5, 5-6, and 5-7.

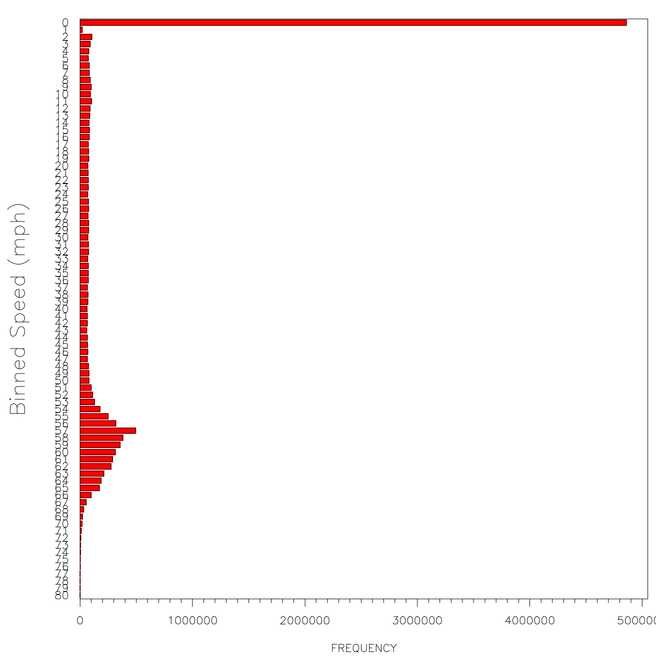
5.3.3 Criteria for Skipping Micro-Trips for a Cycle

In general, the cycle development programs were run for each of the 30 cases to allow selection of micro-trips that best described the operation of vehicles for all operation in the case. However, some types of micro-trips were not considered for inclusion in the candidate cycle.

Some micro-trips were entirely idle operation. These micro-trips were not assigned to any cases since a dedicated idle cycle is not needed.

For the purposes of selecting micro-trips for cycles, observations with extreme acceleration values or extreme VSP values were not considered in the dataset or in the cycle. But they also were not deleted from the dataset or cycles. Specifically, observations with accelerations greater then 14 mph/s or less then –10 mph/s or with VSPs greater than 62.5 kW/Mg or less then –47.5 kW/Mg were not considered.

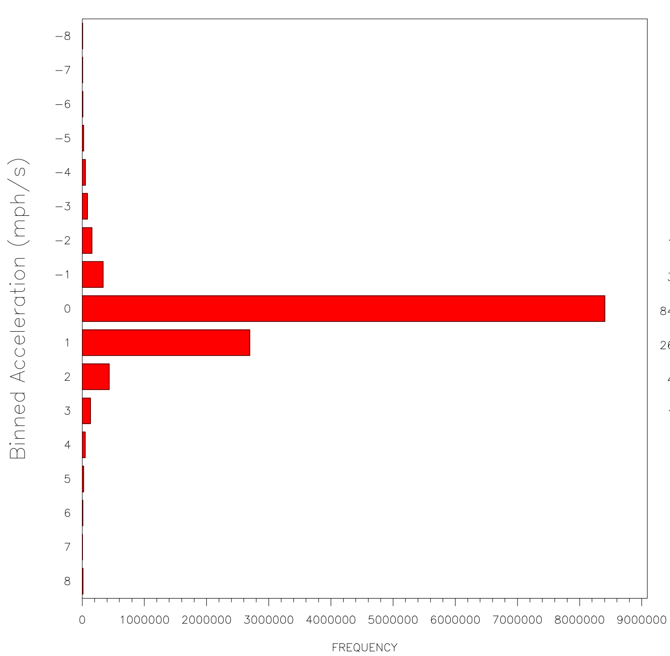
Table 5-5. Distribution of Binned Speeds in the Edited Dataset



Binned Speed (mph)	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
0	4858572	4858572	38.60	38.60
1	18214	4876786	0.14	38.75
2	103487	4980273	0.82	39.57
3	90129	5070402	0.72	40.29
4	75948	5146350	0.60	40.89
5	72887	5219237	0.58	41.47
6	82565	5301802	0.66	42.12
7	80170	5381972	0.64	42.76
8	90374	5472346	0.72	43.48
9	98852	5571198	0.79	44.26
10	92019	5663217	0.73	45.00
11	101946	5765163	0.81	45.81
12	88808	5853971	0.71	46.51
13	86100	5940071	0.68	47.20
14	77666	6017737	0.62	47.81
15	83044	6100781	0.66	48.47
16	81146	6181927	0.64	49.12
17	71899	6253826	0.57	49.69
18	75646	6329472	0.60	50.29
19	76438	6405910	0.61	50.90
20	68352	6474262	0.54	51.44
21	71344	6545606	0.57	52.01
22	71902	6617508	0.57	52.58
23	73059	6690567	0.58	53.16
24	67511	6758078	0.54	53.69
25	74149	6832227	0.59	54.28
26	75636	6907863	0.60	54.88
27	69849	6977712	0.55	55.44
28	74543	7052255	0.59	56.03
29	75609	7127864	0.60	56.63
30	68880	7196744	0.55	57.18
31	74256	7271000	0.59	57.77
32	74786	7345786	0.59	58.36
33	67164	7412950	0.53	58.90
34	72776	7485726	0.58	59.48
35	72194	7557920	0.57	60.05
36	72133	7630053	0.57	60.62
37	64690	7694743	0.51	61.14
38	69965	7764708	0.56	61.69
39	68540	7833248	0.54	62.24
40	61591	7894839	0.49	62.73
41	64988	7959827	0.52	63.24
42	64789	8024616	0.51	63.76
43	59005	8083621	0.47	64.23
44	64227	8147848	0.51	64.74
45	66089	8213937	0.53	65.26
46	68160	8282097	0.54	65.80
47	64531	8346628	0.51	66.32
48	73870	8420498	0.59	66.90
49	78646	8499144	0.62	67.53
50	78943	8578087	0.63	68.16
51	97731	8675818	0.78	68.93
52	111516	8787334	0.89	69.82
53	128941	8916275	1.02	70.84
54	176710	9092985	1.40	72.25
55	250305	9343290	1.99	74.24
56	316816	9660106	2.52	76.75
57	493022	10153128	3.92	80.67
58	380753	10533881	3.03	83.69
59	355817	10889698	2.83	86.52
60	312882	11202580	2.49	89.01
61	289222	11491802	2.30	91.31
62	274003	11765805	2.18	93.48
63	211028	11976833	1.68	95.16
64	186907	12163740	1.49	96.64
65	170943	12334683	1.36	98.00
66	99375	12434058	0.79	98.79
67	54477	12488535	0.43	99.22
68	31824	12520359	0.25	99.48
69	21625	12541984	0.17	99.65
70	17012	12558996	0.14	99.78
71	11054	12570050	0.09	99.87
72	5481	12575531	0.04	99.92
73	3793	12579324	0.03	99.95
74	2450	12581774	0.02	99.97
75	1847	12583621	0.01	99.98
76	918	12584539	0.01	99.99
77	479	12585018	0.00	99.99
78	354	12585372	0.00	99.99
79	236	12585608	0.00	100.00
80	478	12586086	0.00	100.00

/roadhog/EPAHD/stats/utrip_comp.sas 28AUG03 16:40

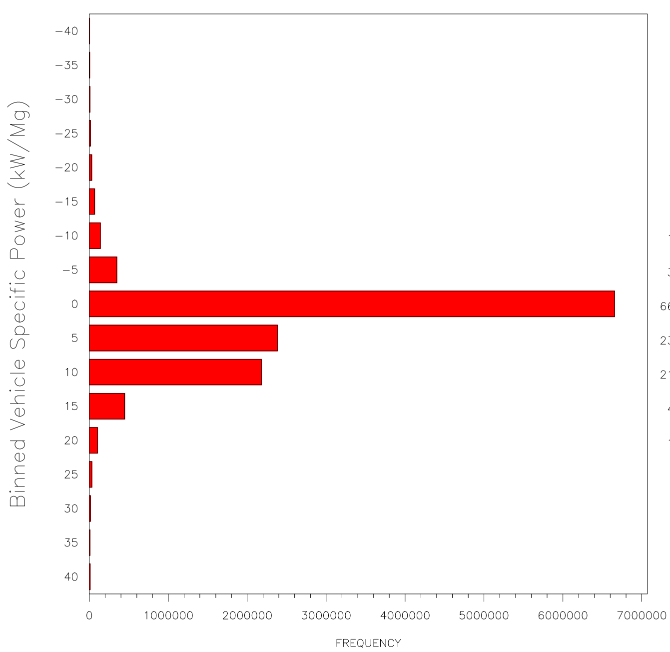
Table 5-6. Distribution of Binned Accelerations in the Edited Dataset



Binned Acceleration (mph/s)	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
−8	5437	5437	0.04	0.04
−7	5159	10596	0.04	0.09
−6	9376	19972	0.08	0.16
−5	22937	42909	0.18	0.35
−4	49018	91927	0.39	0.74
−3	83030	174957	0.67	1.41
−2	156197	331154	1.26	2.66
−1	337299	668453	2.71	5.38
0	8405099	9073552	67.63	73.01
1	2695206	11768758	21.69	94.69
2	435039	12203797	3.50	98.19
3	132814	12336611	1.07	99.26
4	46888	12383499	0.38	99.64
5	20662	12404161	0.17	99.80
6	8719	12412880	0.07	99.87
7	3913	12416793	0.03	99.90
8	11815	12428608	0.10	100.00

/roadhog/EPAHD/stats/utrip_comp.sas 28AUG03 16:40

Table 5-7. Distribution of Binned VSP in the Edited Dataset



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
-40	1661	1661	0.01	0.01
-35	2712	4373	0.02	0.04
-30	6047	10420	0.05	0.08
-25	13432	23852	0.11	0.19
-20	29514	53366	0.24	0.43
-15	65402	118768	0.53	0.96
-10	139898	258666	1.13	2.08
-5	346983	605649	2.79	4.87
0	6653901	7259550	53.54	58.41
5	2379884	9639434	19.15	77.56
10	2179006	11818440	17.53	95.09
15	447719	12266159	3.60	98.69
20	102405	12368564	0.82	99.52
25	31263	12399827	0.25	99.77
30	12996	12412823	0.10	99.87
35	6524	12419347	0.05	99.93
40	9261	12428608	0.07	100.00

/roadhog/EPAHD/stats/utrip_comp.sas 28AUG03 16:40

Any micro-trips less then 20 seconds in duration were not considered for inclusion in a cycle. The reason for not including these is that many short micro-trips can be produced by common, but non-representative, operation of the vehicle. One example is when a vehicle starts moving from a standstill but the engine dies because the clutch is let out too quickly. Another reason that short micro-trips are present in the dataset is because of the algorithm used to divide trips into micro-trips and the presence of dither in the raw data. We believe that the criteria that we used to divide trips into micro-trips erred on the side of making more micro-trips than were actually performed by the vehicles. Accordingly, in some situations very short micro-trips were created that really represent different pieces of dither in the raw data. In any case, we have found in this study as well as in past studies that the micro-trips longer than 20 seconds are adequate to describe the vehicle driving behavior of the entire dataset taken as a whole.

Another reason for deleting some micro-trips after they were selected for a cycle was when the micro-trip contained 10 to 15 or more missing values for speed, acceleration, or VSP. Missing values in micro-trips represent instances where we would be required to manufacture numerical values to produce a complete cycle. Micro-trips that contained too many missing values and especially long strings of consecutive values would be difficult or impossible for us to replace with values that were close to the actual, but unmeasured, speeds that the vehicles drove.

5.3.4 Criteria for Judging Candidate Cycles

For each case the cycle development software built-up a candidate cycle using 25 micro-trips. In each instance, the plot of the square of the length of **T-C** vector as micro-trips were added to the cycle was examined. A sample plot is shown in Figure 5-3. The figure shows that as micro-trips were added, the square of the length of the vector drops substantially at first and then reached a plateau and then dropped again. This drop-following-plateau behavior was commonly seen in many of the cycles generated for the different cases.

Next, we examined a speed versus time plot of the 25 micro-trips that made up the cycle. An example of this is shown in Figure 5-4. The small circles on the plot indicate the beginning of each micro-trip. This candidate cycle plot was used to examine the overall appearance of the cycle and to show the duration of the cycle. At this point, the cycle development analyst decided where the cycle could be ended and still achieve a substantial agreement between the driving behavior in the cycle and the driving behavior in the entire dataset for the case. Typically the cycle was terminated using as few micro-trips as possible but for which the length of the **T-C** vector was quite short and where the addition of more micro-trips caused the length of the vector to increase slightly or be on a plateau.

Finally, we examined scatter plots of acceleration versus speed and VSP versus speed for the candidate cycle and for a random subset of the data in the case under consideration. Examples of these plots are shown in Figures 5-5 and 5-6.

5.3.5 Evaluation of Observations in Micro-Trips after Selection for a Cycle

Once the final set of micro-trips for the cycle of each case was selected, repairs needed by the individual observations in the cycles were identified. A data file for each cycle was created to aid in visual examination by the developer. Each file included the case ID, the micro-trip ID, the edited speed that was used to select the micro-trip, the raw speed that was present in the original dataset, the test vehicle number, the probability that an individual observation was an idle observation, and the flag that indicated the reason for any change of the raw speed to a missing speed value. Plots of edited speed and raw speed versus time for each of the micro-trips selected for a cycle were examined to look for abnormal behavior. Specifically, abnormal behavior would be for portions of speed traces that reflected more the artifacts of the data collection and editing process than the manner in which heavy-duty vehicles are driven.

Figure 5-3. Square of the Length of T_C as Micro-Trips are Added for Case H_1_50

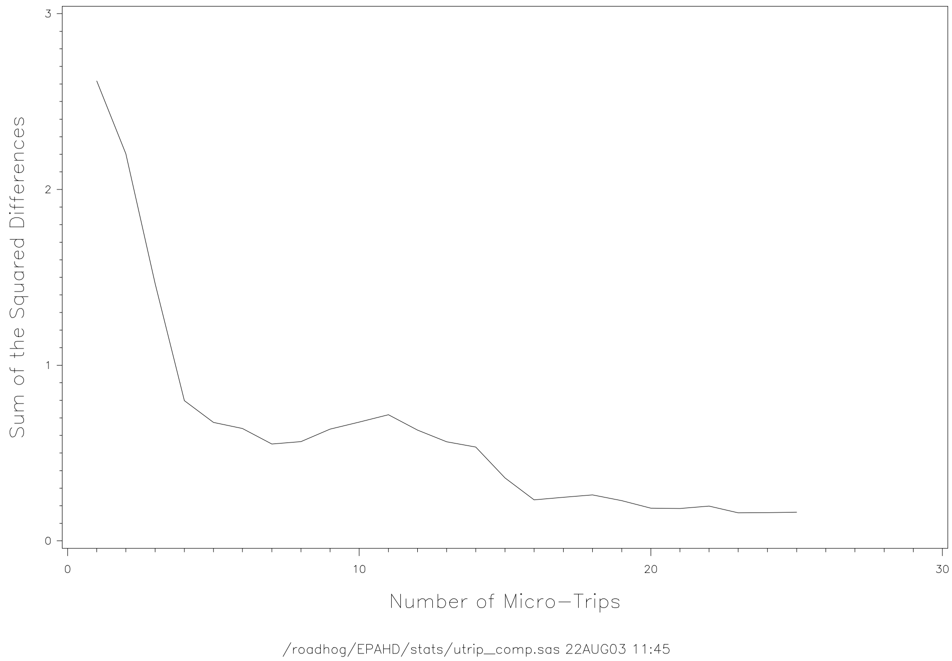


Figure 5-4. Speed vs. Time for the Candidate Cycle for Case H_1_50

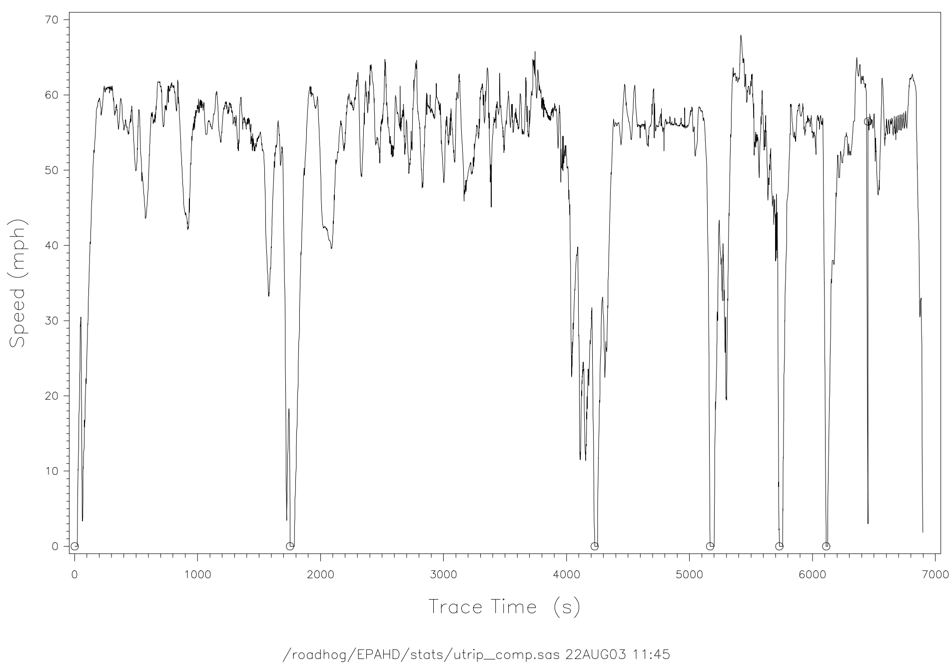
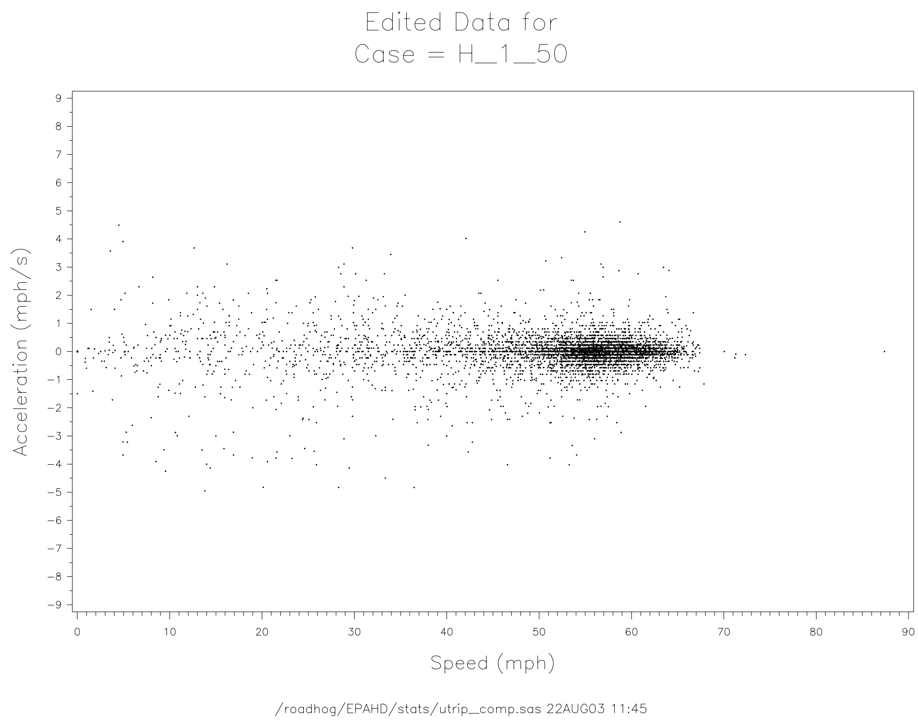


Figure 5-5. Acceleration vs. Speed for Case H_1_50

a) Target



b) Cycle

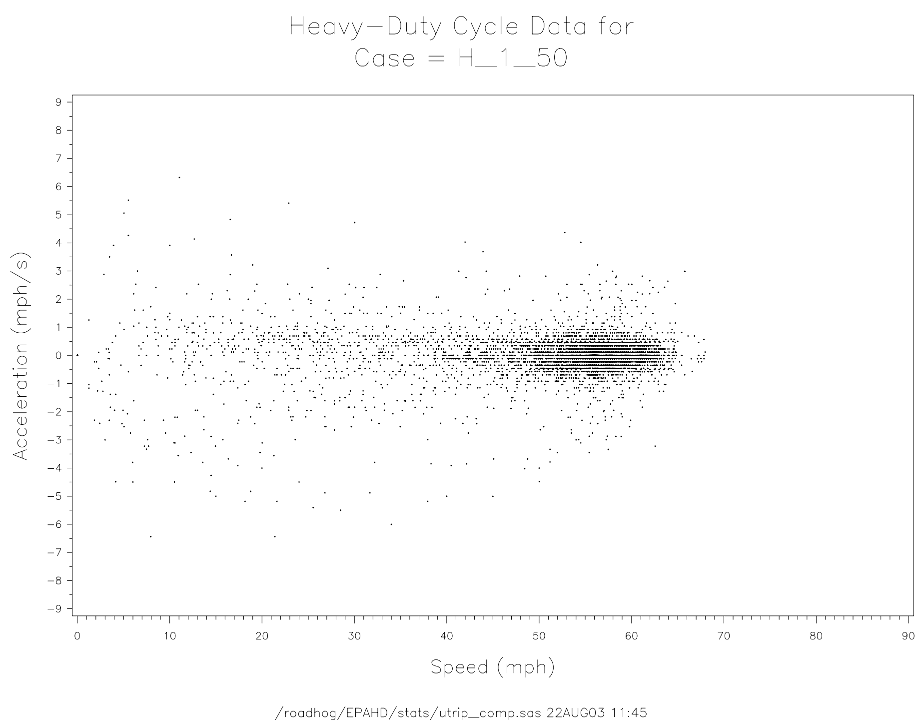
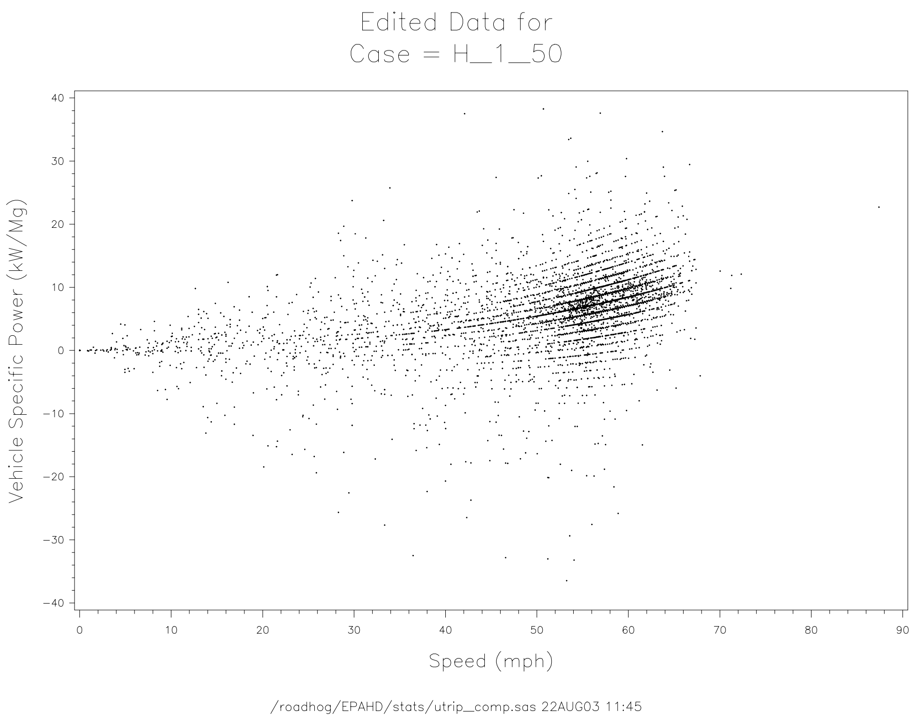
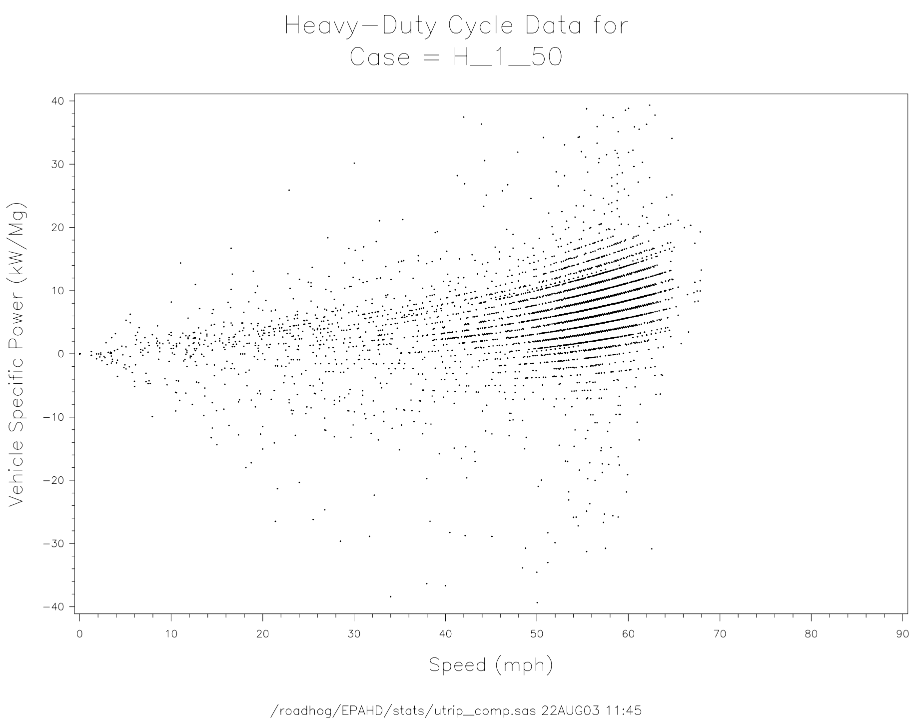


Figure 5-6. VSP vs. Speed for Case H_1_50

a) Target



b) Cycle



Several types of problems were identified in the micro-trips of the cycles:

- **Missing values –** While during cycle development micro-trips were selected only if they had less then 10 to 15 missing values, many of the micro-trips did have some missing values. In most cases, missing values were isolated to a one-second period.
- **Jumps in speed from idle to non-idle segments –** Because of the algorithm used to detect idles in the raw data, we frequently saw jumps in speed from zero speeds, which were the idles at the beginning of micro-trips, to significantly larger speeds than might really be expected in normal driving behavior when the vehicle started moving. Jumps in speed as large as 7 miles per hour were seen. However, when we examined the raw speed for those observations, in many instances we saw that the raw speed values were quite reasonable.
- **Speed shoulders at the end of micro-trips –** At the end of most micro-trips derived from the Battelle and Faucett datasets the speed traces displayed a shoulder where the speed was decreasing and almost came to a plateau and then abruptly dropped to zero. We believe this behavior was a result of the dither in the speed values being clipped by the idle detection algorithm when trips were converted to micro-trips. In other words, as the speeds got low, the influences of dither became more obvious in a speed trace and produced these shoulders, which are not at all typical of normal vehicle driving behavior.
- **"Stuck" speeds in the Battelle micro-trips –** Both the Faucett and Battelle datasets had many periods during which the reported vehicle speed was constant for periods of consecutive observations. We believe these periods were a result of the GPS units temporarily losing contact with the satellite during vehicle driving. In these instances, the datalogger retained the most recent vehicle speed until the GPS unit reacquired the satellite signal. Because the Faucett datalogging system reported vehicle speeds to much higher resolution (0.02 mph) than did the Battelle datalogging system (0.11 mph), the SAS quality checking program was able to successfully distinguish most periods of stuck speeds in the Faucett datasets but it could not do so in the Battelle datasets. In the Faucett dataset, stuck speeds were changed to missing values but in the Battelle dataset, the stuck speeds were left unedited and contained no edit flags. Examination of the cycles show numerous instances of stuck speed segment for up to perhaps 20 seconds. Replacement of these stuck speed values with realistic vehicle operating speeds becomes more difficult as the duration of the stuck speed segment increases. At some point, micro-trips containing stuck speeds probably should be deleted from consideration in cycle development.
- **Dither in micro-trips –** In spite of our efforts to detect dither during the quality checking of the entire dataset, periods of speed observations that are clearly dither remain in some micro-trips and are present in the final cycles. Whether a particular segment of observations is dither or not is not always clear-cut but, in most cases, the cycle developer can make reasonable guesses.

At the point in the project where we wanted to start making edits to the final micro-trips, we ran short on approved labor hours and budget. Therefore, the selected micro-trips could not be edited to solve most of the problems discussed above. Instead of performing detailed corrections on all of the observations in the 30 cycles where problems existed, we simply linearly interpolated values for missing speeds from the speed values before and after missing speed segments. This produced cycles that had all observations with non-missing speed values. However, the speed behavior during interpolated segments is not always representative of the manner in which heavy-duty vehicles are driven. The statistical results that are calculated for the tables in Section 7 and for the plots that are shown in Appendix F are all based on these final cycles with interpolated values replacing missing values. Section 8 presents recommendations for revisiting the development and editing of the 30 heavy-duty vehicle cycles to produce much improved cycles.

6.0 Heavy-Duty Vehicle Operating Characteristics

EPA wanted us to evaluate the characteristics of the datasets from the perspective of heavy-duty vehicle operation. This evaluation is really independent of the development of the cycles. Several pages of SAS printouts and plots are provided in Appendices A, B, C, D, and E to give an idea of the vehicle operation, trip characteristics, and micro-trip characteristics of the combined TxDOT, Faucett, and Battelle datasets.

7.0 Comparison of Dataset and Cycle Statistics

For each of the 30 cases, we produced a cycle from micro-trips selected from all micro-trips that were assigned to the case. A number of statistics were calculated for each case so that the characteristics of the cycle could be compared to the characteristics of the dataset, which we call the target. For each of the statistics that was requested by EPA, we provide a value for target and cycle. These statistics are shown in Tables 7-1, 7-2, and 7-3. It is important to remember when comparing any of these statistics in the three tables that the micro-trips in the cycles were selected only because their non-idle speed, VSP, and acceleration characteristics match those of the target. Any other statistics that are calculated and compared were not the basis, or at least not the direct basis, for choosing the micro-trips for each cycle. The fact that the micro-trip statistics for the targets and cycles come as close as they do is noteworthy, but perhaps not critical, to the applicability of the cycles for the calculation of emissions.

Table 7-1 shows that the average second-by-second speeds of the cases for the targeted cycle data agree well with each other. Probably the only exception to this statement is that for the lowest speed cases, the cycle had a substantially larger average speed then the target did. The reason for this is that the cycles were selected based only on the non-idle portion of the micro-trips in the target set. Therefore, differences in the fraction of idling can have an influence on the average speed and this influence is most easily seen in the micro-trips that have the lowest average speed. The influence of the difference of idles in the target and cycle can also be seen in Table 7-2 by comparing the percentages in the idle mode. The lowest speed cases have much larger percentages of idle in the target data then in the cycle data. Distributions for speed, acceleration, and VSP are provided in plots in Appendix F.

Table 7-1 also shows that the average micro-trip distance for cycles is usually smaller then the average micro-trip distance for the target. We have not determined why this appears to be the case.

Table 7-1. Comparison of Dataset and Cycle Operation Characteristics

Case Code Name	Total Time (s)		Average sxs Speed (mph)		Total Distance (mile)		Average Micro-Trip Time (s)		Average Micro-Trip Distance (mile)		Micro-Trips (count)	
	Target	Cycle	Target	Cycle	Target	Cycle	Target	Cycle	Target	Cycle	Target	Cycle
N_0_5	493,816	638	1.3	4.9	161.86	0.87	93	58	0.031	0.079	5288	11
N_0_10	84,840	883	10.0	10.5	217.66	2.58	96	63	0.245	0.185	887	14
N_0_15	97,070	1,100	15.1	15.6	382.21	4.75	117	79	0.461	0.340	829	14
N_0_20	111,862	2,336	19.9	20.4	588.22	13.22	148	117	0.776	0.661	758	20
N_0_25	83,014	1,536	24.7	24.4	550.31	10.39	167	96	1.107	0.650	497	16
N_0_30	55,621	1,702	31.1	30.8	464.13	14.58	189	142	1.579	1.215	294	12
N_1_30	63,189	5,905	28.2	29.9	469.96	48.27	780	484	5.802	4.022	81	12
N_1_40	117,229	5,646	41.0	41.2	1276.51	64.59	862	706	9.386	8.075	136	8
N_1_50	226,562	8,618	50.4	50.2	3030.66	120.17	1325	718	17.723	10.014	171	12
N_1_60	168,786	8,157	58.0	57.1	2496.70	129.49	2312	1360	34.201	21.582	73	6
P_0_5	726,990	988	1.7	5.3	283.08	1.45	52	47	0.020	0.069	14030	21
P_0_10	227,431	856	10.0	10.7	560.45	2.55	58	61	0.142	0.182	3942	14
P_0_15	242,596	988	15.0	15.5	921.11	4.25	72	82	0.273	0.354	3370	12
P_0_20	220,948	896	19.9	19.7	1148.56	4.91	95	75	0.495	0.409	2320	12
P_0_25	147,958	1,109	24.8	25.5	963.77	7.87	119	92	0.778	0.656	1239	12
P_0_30	148,539	1,657	33.1	32.5	1309.57	14.97	153	127	1.351	1.151	969	13
P_1_30	25,741	2,600	29.9	31.9	204.29	23.03	515	433	4.086	3.838	50	6
P_1_40	79,901	6,196	39.9	41.2	871.14	70.87	484	443	5.280	5.062	165	14
P_1_50	108,965	5,282	50.1	49.4	1498.06	72.48	623	440	8.560	6.040	175	12
P_1_60	97,217	8,586	59.9	59.5	1606.31	141.94	1389	1227	22.947	20.278	70	7
H_0_5	4,215,612	1,441	0.5	4.7	520.18	1.87	196	63	0.024	0.081	21554	23
H_0_10	239,821	1,997	9.9	10.8	619.07	5.96	97	80	0.252	0.239	2461	25
H_0_15	201,255	1,251	15.0	15.2	798.01	5.29	125	104	0.496	0.441	1610	12
H_0_20	208,233	1,797	20.0	19.8	1113.85	9.89	158	106	0.846	0.582	1317	17
H_0_25	177,071	2,166	24.8	24.9	1185.85	14.96	188	155	1.256	1.069	944	14
H_0_30	153,756	1,967	32.1	30.8	1304.50	16.84	202	164	1.714	1.403	761	12
H_1_30	303,878	6,670	27.4	31.9	2192.32	59.17	902	513	6.505	4.551	337	13
H_1_40	580,763	7,350	41.0	41.2	6248.94	84.15	976	613	10.502	7.012	595	12
H_1_50	2,391,077	6,898	51.1	50.6	32489.53	97.03	2067	985	28.081	13.862	1157	7
H_1_60	1,550,245	16,685	58.2	58.0	24616.30	268.60	3682	2781	58.471	44.770	421	6

Table 7-2 shows a comparison of the dataset and cycle operation modes. The operation modes were set by the program utrip_comp.sas. An observation was assigned to "Cruise" if the average difference between the previous observation and the following observation was less than 0.5 mph and the speed of the observation was greater then 0.00 mph. The observation was called a "Decel" if the average difference between the previous and the following observations was less then or equal to -0.5 mph and the observation had a speed greater then 0.00 mph. The observation was called an "Accel" if the average difference between the previous and following observation was greater then or equal to 0.5 mph and the speed of the observation was greater then 0.00 mph. If the difference between the previous and the following observation could not be determined because one or both were missing or the observation speed itself was missing, then the mode was determined to be "Missing." All other observations were assigned to the "Idle" mode.

Table 7-2. Comparison of Dataset and Cycle Operation Modes

	Accel (%)		Cruise (%)		Decel (%)		Idle (%)		Missing (%)	
Case Code Name	Target	Cycle	Target	Cycle	Target	Cycle	Target	Cycle	Target	Cycle
N_0_5	3.23	17.55	9.26	35.11	3.34	17.87	73.40	27.74	10.76	1.72
N_0_10	20.46	26.61	25.08	24.46	18.10	24.24	24.92	23.10	11.44	1.59
N_0_15	26.57	33.00	26.52	25.09	21.93	27.36	15.37	13.27	9.60	1.27
N_0_20	27.16	33.26	30.13	29.11	21.91	26.41	11.57	10.36	9.22	0.86
N_0_25	27.50	32.42	34.23	29.88	22.45	26.24	8.86	10.42	6.96	1.04
N_0_30	26.66	30.85	38.01	36.90	21.09	24.32	7.13	7.23	7.11	0.71
N_1_30	17.87	24.03	42.65	45.99	15.00	18.95	15.79	10.82	8.69	0.21
N_1_40	17.93	19.85	55.06	60.64	15.45	16.44	2.66	2.92	8.89	0.14
N_1_50	12.46	14.96	66.87	69.38	11.22	13.38	1.62	2.15	7.83	0.14
N_1_60	7.22	10.09	73.70	78.53	6.82	9.46	0.60	1.84	11.67	0.07
P_0_5	4.34	18.83	5.00	16.70	4.83	18.83	64.49	43.52	21.34	2.13
P_0_10	20.91	26.29	18.28	22.43	20.39	24.42	26.99	25.23	13.43	1.64
P_0_15	27.04	33.30	21.08	21.86	24.38	27.63	17.19	15.99	10.30	1.21
P_0_20	30.34	33.71	24.65	23.55	25.93	28.01	11.70	13.39	7.38	1.34
P_0_25	30.32	35.17	28.69	29.67	25.36	27.95	8.98	6.13	6.65	1.08
P_0_30	28.39	32.47	35.66	32.71	23.19	25.65	7.55	8.39	5.21	0.78
P_1_30	25.67	30.92	39.84	41.85	21.17	24.04	7.90	2.96	5.42	0.23
P_1_40	23.10	26.15	49.15	48.29	18.83	21.45	6.65	3.89	2.27	0.23
P_1_50	17.91	19.42	62.68	60.89	14.68	15.68	2.96	3.79	1.78	0.23
P_1_60	9.10	10.97	79.63	77.98	8.60	9.34	1.49	1.63	1.18	0.08
H_0_5	0.98	14.92	3.84	36.36	1.04	14.23	82.80	32.89	11.35	1.60
H_0_10	17.38	23.23	37.02	40.51	14.84	17.88	21.91	17.13	8.86	1.25
H_0_15	24.08	29.10	34.01	32.05	18.04	22.06	15.78	15.83	8.10	0.96
H_0_20	25.54	31.27	37.26	34.45	18.58	19.87	12.55	13.47	6.07	0.95
H_0_25	25.94	29.69	41.56	39.47	18.72	20.96	9.03	9.23	4.75	0.65
H_0_30	24.35	28.11	44.24	42.60	17.80	18.05	6.86	10.63	6.75	0.61
H_1_30	15.29	21.53	47.78	59.04	12.30	16.69	15.71	2.55	8.92	0.19
H_1_40	13.40	18.30	62.27	64.83	10.88	13.48	3.66	3.22	9.79	0.16
H_1_50	7.95	11.00	76.99	77.76	7.05	8.84	1.47	2.29	6.54	0.10
H_1_60	4.95	6.08	85.99	87.19	4.68	5.66	0.58	1.04	3.80	0.04

EPA also requested certain maximum and minimum values for speed, acceleration, and VSP. When we identified these values, we found that, as expected, they were highly variable since they were the extreme values in the dataset. So rather than reporting these values, we instead determined the 0.5 and 99.5 percentile values for the speed, acceleration, and VSP quantities of interest. These are reported in Table 7-3 for the target and cycle datasets for each case.

Table 7-3. Comparison of Data and Cycle Extreme Values

	99.5 %ile sxs Speed (mph)		0.5 %ile sxs Acceleration (mph/s)		99.5 %ile sxs Acceleration (mph/s)		0.5 %ile VSP (kW/Mg)		99.5 %ile VSP (kW/Mg)	
Case Code Name	**Target**	**Cycle**	**Target**	**Cycle**	**Target**	**Cycle**	**Target**	**Cycle**	**Target**	**Cycle**
N_0_5	24.27	28.21	-2.99	-3.57	2.50	3.57	-3.86	-10.06	6.24	12.45
N_0_10	39.78	35.28	-4.71	-4.60	4.25	5.01	-14.28	-15.10	15.11	17.17
N_0_15	44.18	41.30	-4.95	-5.29	4.26	5.99	-18.41	-20.41	18.63	20.32
N_0_20	47.84	44.45	-4.97	-5.40	4.14	4.71	-19.20	-19.71	19.34	20.31
N_0_25	51.79	48.42	-5.06	-4.95	4.14	6.62	-21.46	-22.77	21.86	25.79
N_0_30	61.10	55.88	-5.18	-5.63	4.07	4.50	-24.10	-24.73	23.74	26.10
N_1_30	64.68	61.41	-3.91	-4.20	2.89	3.17	-18.17	-22.89	18.87	24.85
N_1_40	66.96	67.28	-4.07	-4.37	3.00	3.00	-20.64	-24.82	23.76	29.63
N_1_50	69.02	68.31	-3.55	-4.02	2.65	2.87	-19.55	-18.78	26.00	26.75
N_1_60	77.34	67.39	-3.22	-3.50	2.53	2.76	-18.38	-22.33	27.50	31.37
P_0_5	25.99	27.95	-6.10	-7.25	5.41	7.94	-8.36	-15.06	12.87	25.03
P_0_10	41.17	34.39	-7.36	-7.36	11.39	12.77	-19.36	-22.04	29.73	34.25
P_0_15	43.93	39.45	-7.36	-8.45	11.04	13.57	-22.74	-32.11	31.89	53.94
P_0_20	48.88	46.12	-7.13	-6.79	8.06	13.23	-25.64	-26.22	31.84	36.01
P_0_25	55.09	52.56	-7.02	-7.13	6.56	9.89	-28.64	-32.14	32.27	34.82
P_0_30	64.75	61.30	-6.91	-7.13	5.86	7.07	-32.68	-37.65	33.50	42.59
P_1_30	64.29	64.52	-5.98	-5.52	4.71	4.72	-28.40	-25.60	28.75	29.23
P_1_40	67.62	67.51	-5.75	-5.52	4.26	4.37	-29.25	-31.01	30.36	32.33
P_1_50	73.14	71.53	-5.29	-5.64	3.79	4.14	-28.48	-30.77	31.30	34.07
P_1_60	73.83	72.34	-4.02	-4.60	2.76	2.88	-21.27	-23.61	31.11	31.54
H_0_5	15.07	26.87	-1.61	-3.68	1.25	5.06	-0.87	-7.42	2.50	10.36
H_0_10	39.68	34.00	-5.06	-5.29	4.83	7.59	-11.75	-13.62	13.82	17.76
H_0_15	42.35	41.28	-5.40	-4.94	4.72	6.79	-16.25	-20.11	17.34	20.43
H_0_20	47.38	43.70	-5.17	-5.52	4.26	4.49	-18.66	-22.05	19.05	21.03
H_0_25	53.48	50.72	-5.06	-4.71	4.15	5.06	-20.78	-21.30	21.30	22.01
H_0_30	61.76	64.06	-5.06	-4.14	4.03	4.25	-22.72	-21.14	24.21	27.58
H_1_30	62.53	60.49	-3.80	-3.80	2.76	3.10	-17.40	-16.74	19.27	19.19
H_1_40	64.86	63.02	-3.68	-4.37	2.64	2.88	-18.50	-20.70	22.32	25.50
H_1_50	66.17	64.63	-2.88	-3.68	2.19	2.65	-15.72	-21.84	24.69	31.92
H_1_60	70.73	69.69	-2.30	-2.42	1.95	2.18	-13.37	-13.36	26.10	26.44

For each of the cases, there are 20 plots in Appendix F. The plots in Appendix F will be described, in general, here so that the reader can get an idea of what they show. On the first page of the plots for each case, the top plot shows the final cycle speed trace. The small circles along the trace denote the beginning of each micro-trip that makes up the cycle. By counting the number of small circles on the speed trace, the number of cycles can be determined.

The second plot for each case shows the square of the length of the **T-C** vector as micro-trips were built up to create the candidate cycle that contained 25 micro-trips. Usually the first few micro-trips drop the sum of squares value substantially. Then, the addition of more micro-trips may reduce it slightly more or, in some cases, may increase the length of the **T-C** vector to a degree. The plot does not show the point at which the build up was truncated for the final micro-trip shown in the previous plot.

The next nine pages in each series show plots on the top of the page that represent the target data and plots on the bottom of the page that represent the same sort of plot but for the cycle data. Therefore, by comparing the plot on the top with the plot on the bottom, the representativeness of the cycle can be evaluated. Some judgment needs to be made in comparing the horizontal bar plots because auto-scaling was used to make those histogram plots. In the case of the scatter plots and the stair-step plots, the same scales were used for each pair of plots.

Horizontal bar plots are shown to denote the distributions of speeds, accelerations, and VSPs on the next six plots. The four scatter plots show a comparison of acceleration versus speed and VSP versus speed. In the case of the cycle data on the bottom of these plots, all of the data points are plotted. However, because of the much larger number of data points in the target data only a random subset of the data points are plotted for the target data.

The next several stair-step plots show distributions of micro-trip times, time in idle in the micro-trips, running time in the micro-trips which is the time when the vehicle is not idling, and micro-trip distance. Each of the stair-step plots shows the differential distribution with a solid line and a cumulative distribution with a dotted line. The horizontal scale of the stair-step plots is a log scale so that some of the detail at very low values and very high values can be seen at the same time on one plot.

8.0 Recommendations for Development of Final Cycles

As discussed above, the final cycles developed in this work assignment have minor problems. However, now that we have gone completely through the development of the heavy-duty cycles, we can look back on the development process to see where improvements can be made. We know that much better heavy-duty cycles can be easily and quickly developed using this same set of data and we think that a small follow-on work assignment can be used to produce these cycles in a short amount of time.

There are two major and two minor problems that we see in the final cycles developed here:

Major Problem 1: The Battelle dataset contains "stuck" speed values. Since these stuck speed values are present not only in the developed cycles but also, more importantly, in the dataset, they potentially cause a bias in the cycles because the cycles will try to simulate the speed, accelerations, and VSPs that are associated with the stuck values. **Solution:** We think that removal of the Battelle speeds that are associated with stuck speed values from the Battelle dataset is worth doing. Stuck speed values could be changed to missing values and a stuck speed value flag for those observations could be produced so that they could be quickly found in any micro-trip chosen for a cycle. Changing the stuck speed values to missing values would eliminate just those observations from consideration for comparison with the candidate micro-trips under review for a given cycle.

Major Problem 2: In the 30 cycles developed in this work assignment, many of the micro-trips have the minor problems that were presented in Section 5.3.5. However, fixing these problems after <u>all</u> of the micro-trips have been chosen for a given cycle will result in a cycle that does not match the dataset as well as the original cycle matched it before the edits were made. Yet, it is important to make these edits so that the cycle that is produced is a reasonable representation of the way heavy-duty vehicles are driven. **Solution**: If each micro-trip were evaluated and edited for the detailed second-by-second observation problems listed in Section 5.3.5 before the micro-trip selection algorithm proceeded to select the subsequent micro-trip, this problem would be avoided. If an iterative process of select-evaluate-edit for each micro-trip were used, by the time all 25 micro-trips were selected, each of them would have already been edited. In addition, if the edits on a selected micro-trip slightly degrade the ability of that micro-trip to match the dataset, the subsequent micro-trips would be chosen by the program to converge on the entire sequence of micro-trips that has a good match with the dataset.

Minor Problem 3: The allocation of speed, acceleration, and VSP bins to the continuous values of those variables does not reflect the importance of the variables on emissions. We used too many speed bins (80) and too few acceleration bins (8) and VSP bins (10). **Solution**: Since emissions are perhaps more closely associated with acceleration and VSP then with speed, the binning scheme for acceleration and VSP should contain more bins and fewer bins should be used for speed. This will produce cycles that better match the driving characteristics that are important to emissions.

Minor Problem 4: In the micro-trip selection program we forbade the selection of any given micro-trip more than once. This arbitrary rule provides a restriction, although perhaps minor, on the micro-trips that are allowed to be selected for a cycle. We could envision that a cycle could be made up of nine repeats of one micro-trip and one repeat of another micro-trip perhaps because the bulk of the dataset behavior is represented by the first micro-trip but the second kind of micro-trip is needed to a small degree. With the current restrictions on the micro-trip selection this mixture of different micro-trip richnesses cannot be provided. **Solution**: The only restriction we really need to make is that the second micro-trip cannot be the same micro-trip as the first one. From the third micro-trip on, any micro-trip should be allowed to be selected.

As far as the detailed second-by-second editing of selected micro-trips goes, the following techniques can be used to edit micro-trips as each one is selected for addition to the candidate cycle being built up:

- **Missing values** - Where missing values are present, hand editing can be used to replace missing values with numeric values for speed by examining the speeds in the seconds before and after the missing value segment.
- **Jumps in speed from idle to non-idle segments –** In many instances the decision by the idle detection algorithm can be overridden and the reported raw speed values can be used. This can produce reasonable speed transitions at the beginning of the non-idle portion of micro-trips.
- **Speed shoulders at the end of micro-trips** - In these instances, hand edits can be used to remove the shoulder by decreasing the speed values in two or three of the seconds at the end of each micro-trip.
- **Stuck speeds from Battelle data –** If periods of stuck speeds are relatively short (on the order of 5 seconds), manufactured speeds may be relatively easy to produce. For periods of stuck speeds longer than this, it may be more reasonable to eliminate the offending micro-trip from further consideration in building up the candidate cycle.

- **Dither in micro-trips –** In most cases, the dither that may be present as part of selected micro-trips may be simply changed to 0.00 mph speed values and one or two seconds of speed transition values can be manufactured to avoid large jumps in speed.