Design of Automatic Comprehensive Traffic Data Management System for Pennsylvania

JIN-HYUK CHUNG, KRISHNAN VISWANATHAN, AND KONSTADINOS G. GOULIAS

With the highway infrastructure under strain, there is a need to collect and analyze traffic volume, vehicle classification, and weight data in an integrated manner. Most research focuses on collection and quality of data with little emphasis on an integrated traffic data management system. Design of an integrated traffic data management system that allows for editing, storage, and reporting of traffic volumes, vehicle classification, and vehicle weight along with quality assurance of data is presented. The system includes the creation of new software and the use of existing software that has been modified to allow for the flagging of data errors, the reading of 1995 Traffic Monitoring Guide format data, and the aggregation of vehicle class data in suitable schemes to ease analysis. Procedures for the input of missing data and the correction of erroneous data are described. The system is user friendly and designed to operate under a Windows 3.1 or a Windows NT 4.0 environment. Refinements needed in the future are identified and presented.

The country's aging highway infrastructure requires agencies to collect and analyze data on traffic volumes, vehicle classification, and vehicle weight distributions to plan, design, and operate facilities in a more efficient way. The main focus of research has been on the availability of and access to comprehensive and diverse, yet consistent, traffic, materials, structural, and climatic data for various pavement types (*1*) and on the quality of such data. Whereas the quality of weigh-in-motion (WIM) data and methods to ensure their accuracy are emphasized (*2,3*), not much emphasis is given to the seamless integration of traffic volume, vehicle classification, and weight data for the purposes of storage, analysis, and reporting. To address these issues and to provide the Pennsylvania Department of Transportation (PennDOT) with a comprehensive and coordinated traffic management system that can retrieve, verify, edit, manage, store, and report volume, vehicle classification, and vehicle weight data with the ability to do quality control on this information, a system was designed and implemented to run under Windows 3.1 and Windows NT 4.0 environments by the Pennsylvania Transportation Institute at Penn State University. This paper presents the design of this coordinated automatic traffic data management system.

BACKGROUND

Pennsylvania currently has a 63-station automatic traffic recorder (ATR) system capable of collecting highway traffic volumes. Most of these sites have functional double-loop installations. Consequently, speed and vehicle length data may also be collected. Current data management processes include

J.-H. Chung, K. Viswanathan, and K. G. Goulias, Department of Civil and Environmental Engineering and Pennsylvania Transportation Institute, Pennsylvania State University, 201 Research Office Building, University Park, PA 16802. Current affiliation of J.-H. Chung: Department of Urban Engineering, Chung-Ang University, Ansung-Gun, Kyunggi-Do, 456-756, Republic of Korea.

- Manually initiated, daily, automatic data polling from PC platform;
- Manual data review on PC platform;
- Upload to mainframe platform;
- Semiautomatic data edit on mainframe platform;
- Report production on mainframe platform; and
- Data storage on mainframe platform.

Pennsylvania participates in the Strategic Highway Research Program (SHRP), providing vehicle classification and truck weight data for Long-Term Pavement Performance (LTPP) research. One WIM station provides continuous automatic vehicle classification (CAVC) and weigh data; eight semipermanent WIM systems will provide CAVC data and at least 1 week per quarter of WIM data. WIM data management processes include the following:

- Manually initiated, automatic, daily data polling from PC platform;
- Manual data review on PC platform;
- Manual data edit on PC platform;
- Report production on PC platform; and
- Data storage on PC platform.

Currently, PennDOT collects site, volume, vehicle classification, and weight data, which are maintained in Traffic Monitoring Guide (TMG) No. 1 and No. 2 Cards, No. 3 Card, No. 4 Card, and No. 7 Card formats, respectively (*4*). PennDOT has no system that allows it to retrieve ATR, CAVC, or WIM data based on month, year, site, and type of data required for analysis. No system exists that allows PennDOT personnel to retrieve No. 4 Card based on the type of aggregation of vehicle class required for analysis. In addition, PennDOT does not have a system capable of flagging errors in the data set and thereby checking the performance of the ATR and WIM stations and replacing the erroneous data with clean data for the purposes of analysis and storage. Integrating the ATR, CAVC, and WIM data allows for easy cross checks on the performance of the sites and helps determine whether any of the ATR, CAVC, or WIM data polls are malfunctioning at a particular site. Integration allows for use of the CAVC data to determine the hourly volumes of vehicles, and use of the WIM data allows determination of vehicle classification. Thereby, even if errors are found in the ATR or CAVC data, it is possible to rectify them using the CAVC and WIM data, respectively.

AVAILABLE DATA AND SOFTWARE

The data used for the development of the system were the 1995 and 1996 ATR data (No. 3 Card) and the 1994 vehicle classification

(No. 4 Card) and weight data (No. 7 Card) from stations on I-80 in Luzerne and Clearfield Counties in Pennsylvania. The LTPP QC interface (*5*) developed by Chaparral Systems is the most suitable existing software for quality control checks on data. The primary goal of the interface for truck weight data is to identify calibration drift in WIM scales. The quality assurance calibration test is based on the gross vehicle weight distribution of five-axle tractor-semitrailer vehicles. These vehicles are the most common long-haul trucks in the United States. When loaded they can legally weigh up to 36.32 t (80,000 lb); however, with special permits trucks can carry loads greater than 36.32 t. When empty, the range in weight is between 12.71 and 14.53 t (28,000 and 32,000 lb), depending on the type of tractor and trailer being used. Use of these trucks as calibration tools has two advantages: (*a*) achieving a sufficient sample size is not difficult, and (*b*) they have a number of very consistent characteristics that can be used to measure scale health. Summary statistics used in the process are computed by lane for each data set submitted for a site. For truck weights, the primary quality assurance statistic is the gross vehicle weight (GVW) frequency distributions of 3S2 vehicles. The quality assurance test examines the distribution to verify that the peak values are at expected locations of the frequency curve. For most WIM sites, the unloaded peak should occur between 12.71 and 14.53 t, and the loaded peak should occur at or somewhat below 36.32 t. If the peaks do not occur at the indicated values, calibration drift or errors may be inferred. A number of quality assurance checks that use this frequency distribution are possible. One check used by the LTPP is the total percentage of "overweight" vehicles. This is a site-specific value that raises questions about scale calibration when the percentage of 3S2 vehicles that are greater than 38 t (83,700 lb) exceeds a specific value. An alternative way to calibrate the scales is to use enforcement agents and static scales, which is done by PennDOT at regulation enforcement sites. These data are not currently used by the method described here because of institutional barriers and lack of an integrated effort as of the date this paper was written.

The quality assurance review of vehicle classification data is much more difficult than the quality assurance review of truck weight calibration. Because vehicle classification volumes by class vary so dramatically from site to site, the quality assurance tests applied must be calibrated to each site. The primary tests being applied to the submitted vehicle classification data by lane are as follows:

- The number of consecutive hours during which total traffic volumes are equal to zero must be less than 8.
- The traffic volumes at 1:00 p.m. should be greater than traffic volumes at 1:00 a.m.
- Daily traffic volumes for key vehicle classes should be within specified ranges.
- The percentage of trucks within each truck category for each day should fall within a specific range.

The first of these checks ensures that the equipment has not failed. The comparison of traffic volumes at 1:00 a.m. and 1:00 p.m. is done to ensure that the clock counter has been set correctly. The check for traffic volumes for key vehicle classes is done by the day of the week and season. The ranges used are broad, and therefore the analysis is imprecise, but this analysis is often successful in determining when axle sensor problems are causing errors in the classification process by shifting some vehicles into other categories.

The final major test for classification data comes from an analysis of the "errors" reported by the SHRP equipment. These errors are reported to LTPP as "vehicles that could be counted but not classified" or "vehicles that could be classified but not weighed" (*6*). Typically, these values represent vehicles that changed lanes while crossing over axle sensors of the data collection equipment.

Whereas the LTPP QC interface (*5*) ensures that calibration, equipment, and system errors were detected, it does not flag erroneous data. The output from the interface is graphical, making it difficult to correct these errors. Also, the interface can read data only in the 1993 TMG format, whereas the data being collected follow the 1995 TMG format. The LTPP QC interface has been written to determine equipment malfunction and calibration errors and does not address data quality issues. The procedure suggested by Han et al. (*3*) can be used to determine the quality of WIM data. These are drawbacks of the LTPP QC interface in its present form. Also, the interface is not designed to retrieve ATR, CAVC, or WIM data based on month, year, site, and type of data required for analysis, and for automatic vehicle classification. Other limitations of the interface are as follows (*2*):

- Since it is based on GVW, it offsets the positive and negative errors in individual axle-load measurements.
- It does not explicitly consider the extent of variation in dynamic axle loads.
- It is largely subjective in determining whether a GVW frequency distribution represents a significant departure from the expected norm at a particular WIM site, which makes it hard to integrate into diagnostic software.

To correct the system errors, autocalibration routines have been suggested (*3*), and methods for conducting quality assurance on WIM data (*2*) have been designed.

DESIGN OF AUTOMATIC TRAFFIC DATA MANAGEMENT SYSTEM

Figure 1 shows the complete flowchart of the automatic traffic data management system, from data collection at the WIM/CAVC/ATR stations to data storage, analysis, and reporting. The modified version of the LTPP QC interface allows for flagging of the data and reads the data in the 1995 TMG format. As shown in the flowchart, the sequence of operations from data collection to data storage, analysis, and reporting includes some additional data management system activities. These include running the No. 4 Card and No. 7 Card data through the Vehicle Travel Information System (VTRIS) (*7*), a program developed by FHWA to validate and facilitate editing, summarizing, and generating reports on vehicle travel characteristics. Also, the original No. 4 Card and No. 7 Card data sets are sent to SHRP for storage in the LTPP database. Whereas the modified interface does not use the No. 7 Card data for the other new routines, the original interface's characteristics are used to determine the quality of truck weight calibration.

The numbers in circles in Figure 1 provide the four-step data management process. The circled "1" indicates that data from the ATR, CAVC, and WIM data polls are converted to No. 3 Card, No. 4 Card, and No. 7 Card data files, respectively. If the data are stored in files separated on the basis of site, month, year, and card type, then the data are ready for analysis. If the data have not been sorted on the basis of these criteria, an extraction routine is run, and data are separated in files on the basis of site, month, year, and card type.

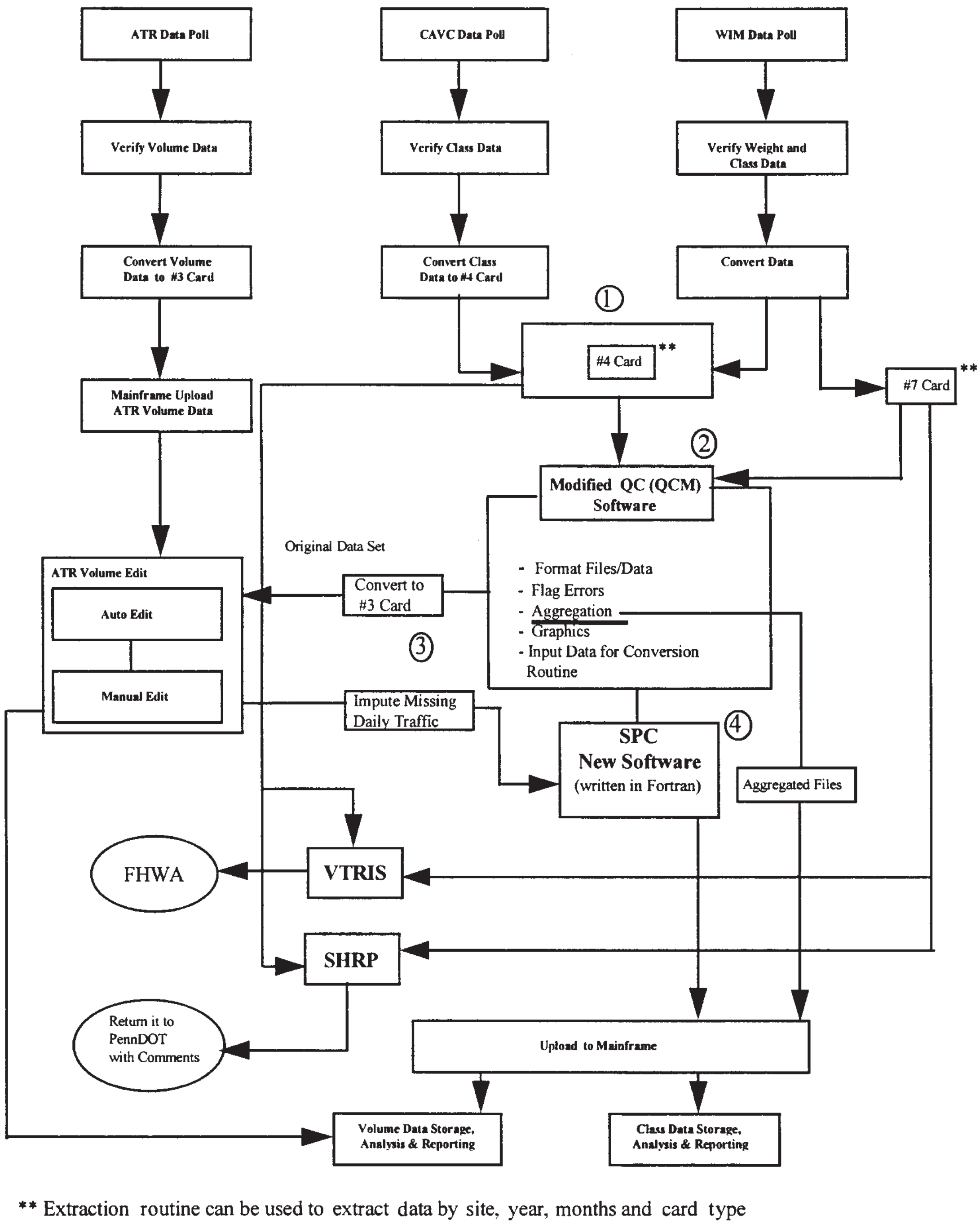

FIGURE 1 Design of automatic traffic data management system.

The circled "2" indicates that once the data have been sorted by site, year, month, and card type, the modified QC interface, called QCM herein, is run and produces a graph catalog and a flag file, aggregated files, and a No. 3 Card conversion file. These files are described in the following paragraphs.

The flag file contains the errors in the data set that have been flagged on the basis of equipment or system errors. Data records are flagged because (*a*) volumes at 1 a.m. are greater than volumes at 1 p.m., (*b*) the same volumes are "observed" for a continuous period of 4 h or longer, (*c*) zero volumes are "observed" for a continuous period of 8 h or longer, or (*d*) other data are missing.

The preceding criteria were selected to flag records because the LTPP QC interface (*5*) uses them as the basis to determine equipment calibration and system errors, and it was decided at this first phase of the project to make minimal changes to existing software and to maintain the integrity of the original interface. As experience accumulates, other modifications are expected.

Aggregated files: Four aggregation schemes were developed for vehicle class analysis:

- Scheme 1: All 13 classes maintained individually;
- Scheme 2: Add Classes 1 through 3 and Classes 4 through 13;

- Scheme 3: Add Classes 1 through 3, 4 through 8, and 9 through 13; and
- Scheme 4: Add together Classes 1 through 3, 4 through 5, 6 through 8, and 9 through 13.

The routine creates all the four aggregation schemes, with erroneous data, and one file, with all 13 FHWA vehicle classes used as an input for the next routine called PC_PSU. This routine allows the merging of the aggregated file with the flag file and the lookup table (for imputing missing data and correcting erroneous data) with volume distribution by class, time of day, and day of week to give a clean data set needed for analysis.

No. 3 Card conversion file: The input data file for conversion from No. 4 Card to No. 3 Card is created on the basis of the year, month, day, and direction and lane of travel.

The circled “3” indicates that once the input file for the conversion of the No. 4 Card data file to the No. 3 Card data file has been completed, the routine TMG3 is run. This creates a No. 3 Card data set, and missing traffic data can be imputed to get a complete volume data set.

The circled “4” indicates that once the data set is ready for aggregation, it is run through the SPC_PSU routine to create the Scheme 1 file on the basis of a lookup table, the diagnostics files, and the missing data imputed. The file containing the suggested values for each error diagnosed is run through the SPC_AGG routine, giving aggregated files for the four schemes.

The automatic traffic data management scheme here also allows use of the unmodified No. 4 Card and No. 7 Card data in VTRIS (*7*).

CORRECTION FOR MISSING OR ERRONEOUS DATA

In this section the routines to create suggested values for missing or erroneous data are described. The CAVC data editing scheme with provision for inputting suggested values is documented.

Recall that the QCM is used to first create the four aggregation schemes as required by PennDOT (Schemes 1 through 4 as described in the preceding section) and then to flag the data for errors on the basis of the four criteria described in the preceding section. At this point in the process values need to be found for erroneous or missing data items. They are produced in the form of suggested values that are reviewed and communicated to QCM concerning the type of action to be taken. Table 1 gives the errors caught by the QCM interface.

Creation of the Lookup Table

The purpose of the lookup table is to determine the volume distribution by class, time of day, and day of the week. The lookup table was also used in the edited No. 3 Card data file to impute the missing values. The No. 4 Card data from 1994 and from site SHRP No. 1 was used to create a lookup table. The lookup table uses only the data from the particular site and does not consider the surrounding sites because of lack of such information. The procedure for creation of the lookup table follows.

Step 1. From the yearly data file, a column containing the days of the week was created. The date on which the observation was recorded was used as a parameter to extract the day of the week.

Step 2. The proportion of vehicles for each hour of the day and class of vehicle was calculated as a fraction of the 24 h of the day. For example, if on the first day of the week at noon there are 12 vehicles of Class 3 and the total number of vehicles for all 24 h of the first day of the week is 480, the value in the cell for that particular class (i.e., Class 3 at noon for the first day of the week) is 12/480 = 0.025.

Step 3. Once the proportions have been created for the first day of the week, the same process is repeated for the remaining days of the week. This process can be done using any spreadsheet package, such as Lotus 1-2-3 or Microsoft Excel.

Step 4. Once the proportions have been calculated for each day of the week, they are put together, giving a lookup table based on day of the week and time of day. The lookup table contains 14 columns and 168 rows. The first 24 rows are for 24 h of the first day of the week (Sunday), the next 24 rows are for the 24 h of the second day of the week, and so on for all days in a week. Thirteen columns are for the 13 classes of vehicles, and the 14th column is for the sum of the first 13 columns. The 14th column data are used for converting edited No. 4 Card to No. 3 Card.

Step 5. The table is saved as an ASCII (text) file so that it is readable with other editors.

Table 2 shows a section of the lookup table. Figures 2 and 3 show the total volumes per class by the day of the week and the logarithm of total volumes per class by the day of the week, respectively, and Figures 4 and 5 show the total volumes per class by time of day and the logarithm of total volumes per class by time of day, respectively, used in the creation of the lookup table. As new data become available for each site, additional lookup tables will be created.

TABLE 1 Errors Caught by QCM Interface (Flag File with Diagnostics)

OBS	SHRPID	DIRECT	IANE	YEAR	MONTH	DAY	DOW	EDIT	STTIME	ENDTIME	NUMB
1	5	7	1	1994	6	2	5	4+ Consec Nonzeros	5	11	6
2	5	7	2	1994	6	2	5	4+ Consec Nonzeros	5	11	6
3	5	3	1	1994	6	2	5	4+ Consec Nonzeros	5	11	6
4	5	3	2	1994	6	2	5	4+ Consec Nonzeros	5	11	6
5	5	7	1	1994	6	3	6	8+ Consec Zeros	2	12	10
6	5	3	2	1994	6	3	6	8+ Consec Zeros	2	12	10
7	5	3	1	1994	6	3	6	8+ Consec Zeros	2	12	10
8	5	3	2	1994	6	3	6	8+ Consec Zeros	2	12	10
9	5	7	1	1994	6	1	4	Time Check	2	14	12

TABLE 2 Section of Lookup Table for Missing Value Imputation

0.00000332	0.005414	0.0009835	0.00008302	0.0002533	0.0001537	0.00000392	0.000184	0.0042234	0.00005312	0.00074372	0.00004352	0.00033312	0.01247556
0.00000322	0.003814	0.0007032	0.00009352	0.0001833	0.0001333	0.00001312	0.000164	0.0033835	0.00005392	0.00069302	0.00003382	0.00033302	0.00960426
0.00000322	0.003153	0.0005831	0.00006352	0.0001531	0.0001234	0.00000382	0.000143	0.0031139	0.00007302	0.00065392	0.00004312	0.00027352	0.00838386
0.00000332	0.002883	0.0006236	0.00007372	0.0001335	0.0001233	0.00001312	0.000104	0.0026534	0.00005362	0.00068352	0.00002392	0.00029322	0.00766506
0.00000322	0.002683	0.0005233	0.00007332	0.0001437	0.0001037	0.00000382	0.000104	0.0026133	0.00007342	0.00069362	0.00004352	0.00026362	0.00732536
0.00000322	0.002953	0.0006433	0.00006352	0.0001536	0.0001239	0.00000382	0.000123	0.0026635	0.00005372	0.00072312	0.00005302	0.00021362	0.00777516
0.00000362	0.004633	0.0010938	0.00007372	0.0001839	0.0001231	0.00001322	0.000193	0.0029333	0.00005342	0.00088382	0.00004352	0.00025392	0.01048616
0.00001342	0.007903	0.0015937	0.00006322	0.0002639	0.0001536	0.00001322	0.000244	0.0042736	0.00005362	0.00099352	0.00004372	0.00030372	0.01591606
0.00001352	0.013503	0.0025637	0.00007302	0.0004038	0.0001831	0.00001352	0.000303	0.0059638	0.00006382	0.00102382	0.00005312	0.00036332	0.02452546
0.00001382	0.021523	0.004023	0.00007302	0.000623	0.0002537	0.00002332	0.000494	0.0076339	0.00010362	0.00094352	0.00004352	0.00059372	0.03634476
0.00001312	0.030683	0.0057233	0.00010312	0.0007234	0.0002638	0.00003352	0.000653	0.0101437	0.00013352	0.00096332	0.00004352	0.00049352	0.04997446
0.00002392	0.038883	0.007163	0.00008362	0.0008333	0.0003336	0.00001382	0.000673	0.0119837	0.00009322	0.00106322	0.00003392	0.00047302	0.06165456
0.00003322	0.042733	0.0075531	0.00012312	0.0008234	0.0002831	0.00002342	0.000664	0.0128237	0.00015302	0.00085312	0.00004322	0.00063312	0.06674266
0.00004322	0.047764	0.0086336	0.00016322	0.0009534	0.0003531	0.00003332	0.000714	0.0142336	0.00017312	0.00098332	0.00006302	0.00055322	0.07466356
0.00003352	0.051344	0.0089936	0.00014312	0.0010036	0.0003735	0.00003302	0.000733	0.0155537	0.00017342	0.00100332	0.00004382	0.00053352	0.07996486
0.00003392	0.052884	0.0094237	0.00016322	0.0010539	0.000423	0.00005302	0.000743	0.0172534	0.00020342	0.00093352	0.00004352	0.00049322	0.08370506
0.00003382	0.050903	0.0090131	0.00017322	0.0010136	0.0004235	0.00002372	0.000724	0.0180538	0.00022372	0.00084382	0.00003392	0.00061392	0.08207696
0.00002372	0.045473	0.0079733	0.00015332	0.0010338	0.0004238	0.00002362	0.000743	0.0184734	0.00026332	0.00090372	0.00004342	0.00074372	0.07627586
0.00004372	0.038013	0.0067335	0.00017322	0.0008634	0.0004133	0.00003352	0.000684	0.0190232	0.00034332	0.00080352	0.00007372	0.00098322	0.06818466
0.00003382	0.030634	0.0052633	0.00016362	0.0007434	0.0004038	0.00001352	0.000734	0.0200538	0.00041342	0.00085382	0.00006352	0.00119362	0.06056706
0.00001372	0.023604	0.0043636	0.00011332	0.0006736	0.0003537	0.00001342	0.000603	0.0189338	0.00044392	0.00083392	0.00003392	0.00099312	0.05097706
0.00001342	0.016704	0.0028638	0.00011352	0.0006238	0.0002935	0.00000392	0.000544	0.0180736	0.00039362	0.00079392	0.00005372	0.00088352	0.04135816
0.00001312	0.011884	0.0022335	0.00008332	0.0005133	0.0002336	0.00001312	0.000513	0.0166035	0.00043332	0.00077392	0.00008332	0.00070392	0.03408526
0.00001312	0.008524	0.0016231	0.00008352	0.0004538	0.0002638	0.00001342	0.000534	0.0152735	0.00037372	0.00077302	0.00007302	0.00090362	0.02890536
0.00000438	0.005714	0.0012644	0.00010438	0.0004744	0.0002544	0.00002438	0.000424	0.0133644	0.00030438	0.00081438	0.00009438	0.00083438	0.02367694
0.00000438	0.004204	0.0009544	0.00008438	0.0003844	0.0002144	0.00002438	0.000404	0.0107344	0.00030438	0.00071438	0.00005438	0.00066438	0.01874694
0.00000438	0.003364	0.0008844	0.00007438	0.0003144	0.0002244	0.00003438	0.000364	0.0094244	0.00022438	0.00069438	0.00006438	0.00058438	0.01625694
0.00000438	0.003164	0.0009644	0.00007438	0.0003944	0.0002844	0.00011438	0.000364	0.0089444	0.00024438	0.00071438	0.00004438	0.00045438	0.01576694

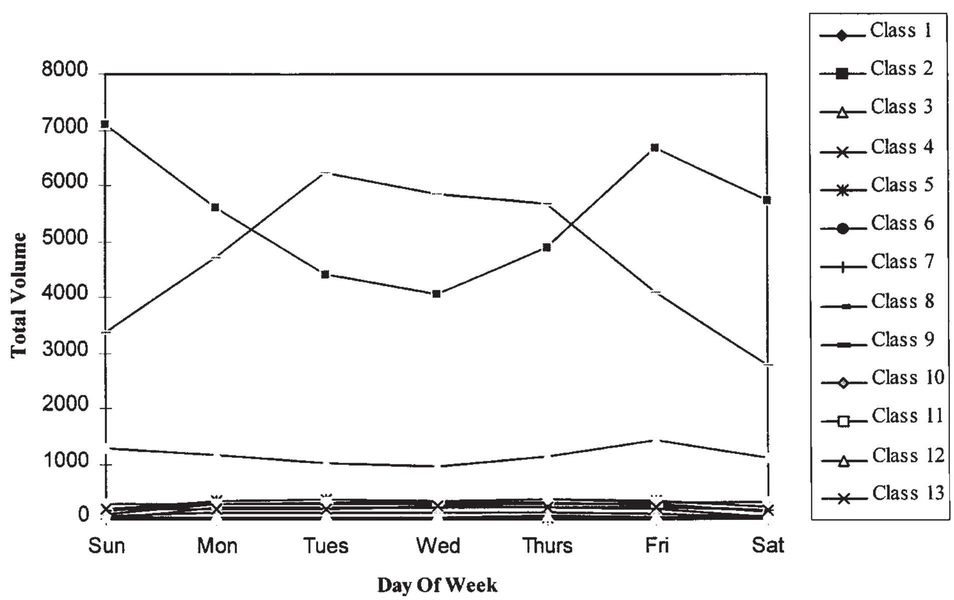

FIGURE 2 Total volume per class by day of week.

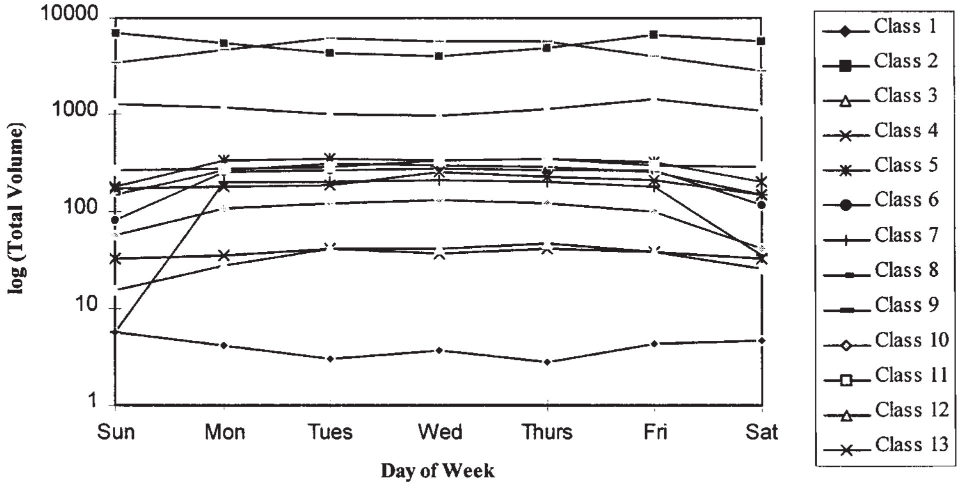

FIGURE 3 Log (total volume) per class by day of week.

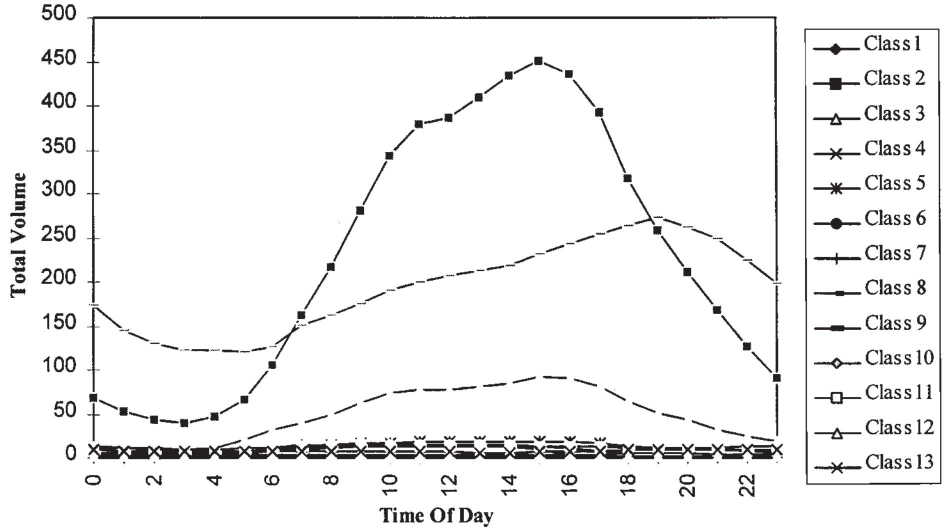

FIGURE 4 Total volume per class by time of day.

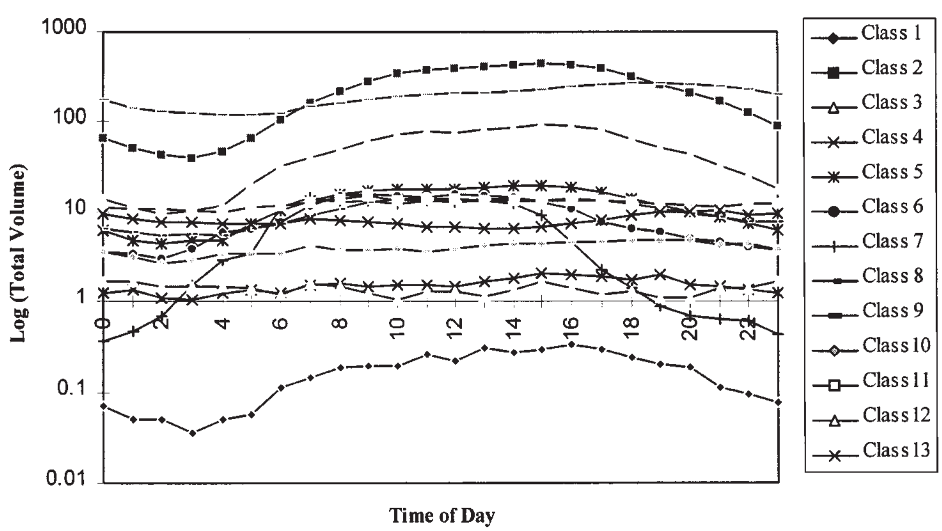

FIGURE 5 Log (total volume) per class by time of day.

Error Data Suggested Values

In conjunction with the QCM two other routines were developed, SPC_PSU and SPC_AGG, to remove the erroneous data and update the original records. In addition, the data are aggregated in the four schemes mentioned earlier. Figure 6 shows the flowchart of the sequence of operations to update the four aggregation schemes with suggested values.

After the QCM is run, it produces six files, one file containing the errors caught by the interface, one file for creating the No. 3 Card data set from the No. 4 Card data set, and four files containing the four aggregation schemes. The flag file with diagnostics and the Scheme 1 files are run through the routine SPC_PSU along with the lookup table with volume distribution by class, time of day, and day of week. In the same routine the table format and header are created. Details concerning the creation of the lookup table were explained earlier. The routine is designed to give a file with suggested values. This is accomplished as follows. When the flag file with diagnostics and the Scheme 1 files are run through the routine, the routine looks for the starting and ending time of the errors in the flag file. The routine refers to the lookup table and suggests a value for the error on the basis of the proportion for each class for the same time and day of week present in the lookup table. A new file is created with suggested values for the errors. The software is also designed to accommodate statistical procedures that will be created in a future application for missing data and the calculation of trends, cycles, and seasonality in the data.

Table 3 gives the aggregation scheme for CAVC with suggested values. In the column labeled "error," the number 1 indicates the error caught by the QCM interface. If no change is needed (i.e., a value can be accepted as is), the 1 should be changed to 0. Changes

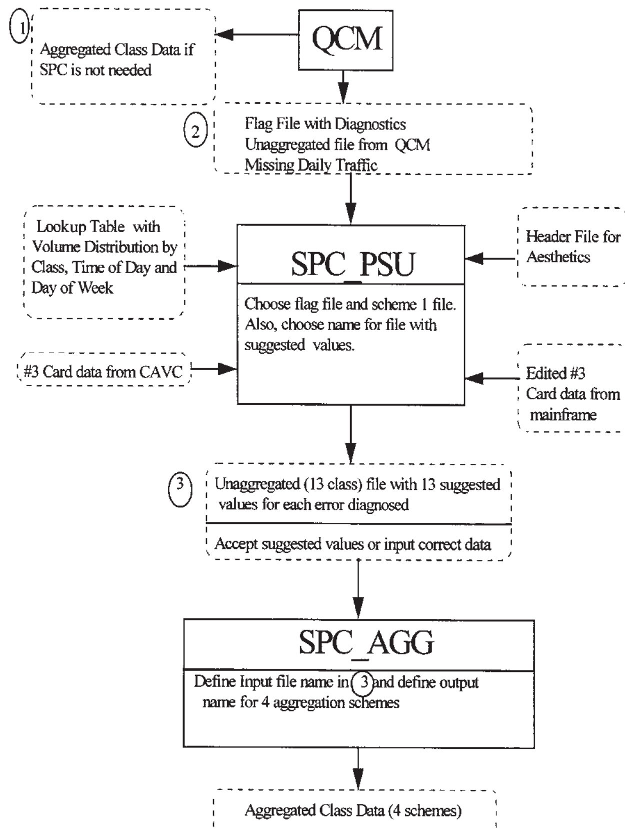

FIGURE 6 Flowchart for suggested values.

TABLE 3 Aggregation Scheme for CAVC with Suggested Values

SHRPID	DIRECT	LANE	YR	MN	DAY	DOW	TIME	CLASS1	CLASS2	CLASS3	CLASS4	CLASS5	CLASS6	CLASS7	CLASS8	CLASS9	CLASS10	CLASS11	CLASS12	CLASS13	ERROR	SUGC1	SUGC2	SUGC3	SUGC4	SUGC5	SUGC6	SUGC7	SUGC8	SUGC9	SUGC10	SUGC11	SUGC12	SUGC13
5	7	2	1994	6	1	4	0	0	14	0	0	0	0	0	2	5	0	0	0	0	0													
5	7	1	1994	6	1	4	0	0	42	4	2	0	1	0	19	55	1	2	0	0	0													
5	3	2	1994	6	1	4	0	0	6	2	0	0	1	0	1	17	0	0	0	0	0													
5	3	1	1994	6	1	4	0	0	37	5	0	3	1	0	2	61	0	5	0	0	0													
5	7	2	1994	6	1	4	1	0	5	0	0	1	1	0	7	4	0	0	0	0	0													
5	7	1	1994	6	1	4	1	0	32	8	0	3	1	0	21	61	0	1	0	0	1	0	48	12	1	7	5	0	11	284	5	17	3	14
5	3	2	1994	6	1	4	1	0	3	1	0	1	0	0	0	8	0	0	1	0	0													
5	3	1	1994	6	1	4	1	0	22	3	1	0	0	0	4	50	0	2	0	0	0													
5	7	2	1994	6	1	4	2	0	7	0	0	0	0	0	3	7	0	0	0	0	0													
5	7	1	1994	6	1	4	2	1	13	2	2	0	2	0	25	85	0	1	0	0	0													
5	3	2	1994	6	1	4	2	0	5	1	1	0	0	0	0	4	0	0	0	0	0													

TABLE 4 Aggregation Scheme 1 for CAVC

SHRPID	DIRECT	LANE	YR	MN	DAY	DOW	TIME	CLASS1	CLASS2	CLASS3	CLASS4	CLASS5	CLASS6	CLASS7	CLASS8	CLASS9	CLASS10	CLASS11	CLASS12	CLASS13	MARK
5	7	2	1994	6	1	4	0	0	14	0	0	0	0	0	2	5	0	0	0	0	
5	7	1	1994	6	1	4	0	0	42	4	2	0	1	0	19	55	1	2	0	0	
5	7	2	1994	6	1	4	0	0	6	2	0	0	1	0	1	17	0	0	0	0	
5	7	1	1994	6	1	4	0	0	37	5	0	3	1	0	2	61	0	5	0	0	
5	7	2	1994	6	1	4	1	0	5	0	0	1	1	0	7	4	0	0	0	0	
5	7	1	1994	6	1	4	1	0	48	12	1	7	5	0	11	284	5	17	3	14	*
5	3	2	1994	6	1	4	1	0	3	1	0	1	0	0	0	8	0	0	1	0	
5	3	1	1994	6	1	4	1	0	22	3	1	0	0	0	4	50	0	2	0	0	
5	7	2	1994	6	1	4	2	0	7	0	0	0	0	0	3	7	0	0	0	0	
5	7	1	1994	6	1	4	2	1	13	2	2	0	2	0	25	85	0	1	0	0	
5	3	2	1994	6	1	4	2	0	5	1	1	0	0	0	0	4	0	0	0	0	
5	3	1	1994	6	1	4	2	0	14	4	0	7	0	0	2	52	0	1	0	0	
5	7	2	1994	6	1	4	3	0	8	0	1	0	1	0	3	5	0	0	0	0	
5	7	1	1994	6	1	4	3	0	19	2	4	2	1	0	28	75	0	2	0	0	
5	3	2	1994	6	1	4	3	0	6	0	0	0	0	0	1	7	0	0	0	0	
5	3	1	1994	6	1	4	3	0	11	2	0	3	1	0	6	53	1	6	2	0	
5	7	2	1994	6	1	4	4	0	10	1	1	1	0	0	6	4	0	0	0	0	
5	7	1	1994	6	1	4	4	0	33	16	0	0	3	1	29	83	0	2	0	0	
5	3	2	1994	6	1	4	4	0	4	0	0	0	0	0	0	6	0	0	0	0	
5	3	1	1994	6	1	4	4	0	20	1	0	2	1	1	2	42	1	4	2	0	
5	7	2	1994	6	1	4	5	0	8	1	0	1	0	0	4	5	0	0	0	0	
5	7	1	1994	6	1	4	5	0	61	16	0	1	2	0	18	47	0	2	0	0	
5	3	2	1994	6	1	4	5	0	5	1	0	0	0	0	1	12	0	2	0	0	
5	3	1	1994	6	1	4	5	0	36	6	0	2	1	0	5	58	1	8	4	0	
5	7	2	1994	6	1	4	6	0	34	4	1	1	0	0	2	3	0	0	0	0	
5	7	1	1994	6	1	4	6	0	138	25	2	2	3	0	23	62	0	5	0	0	
5	3	2	1994	6	1	4	6	0	27	2	0	0	1	0	0	9	0	0	0	0	
5	3	1	1994	6	1	4	6	0	84	20	0	7	0	0	6	71	2	2	0	0	
5	7	2	1994	6	1	4	7	0	36	2	3	0	5	0	7	5	0	0	0	0	
5	7	1	1994	6	1	4	7	1	160	22	4	1	8	0	36	74	0	1	1	0	
5	3	2	1994	6	1	4	7	0	64	5	0	0	0	0	0	9	0	1	0	0	
5	3	1	1994	6	1	4	7	0	166	20	0	17	1	0	6	81	0	5	0	0	
5	7	2	1994	6	1	4	8	0	38	3	1	0	3	0	3	9	0	0	0	0	
5	7	1	1994	6	1	4	8	0	158	23	3	1	6	1	27	73	0	3	0	0	
5	3	2	1994	6	1	4	8	0	57	4	0	3	0	0	1	9	0	1	0	0	
5	3	1	1994	6	1	4	8	0	236	28	0	13	0	0	9	58	1	3	1	0	
5	7	2	1994	6	1	4	9	0	38	4	2	0	0	0	6	5	0	1	0	0	

can be made on an as-needed basis to reflect the nature of the data. This allows any type of data editing (manual, semiautomatic, and completely automatic) and allows for imputation of simulated data (the suggested values). As Table 3 indicates, the suggested values appear higher than expected. This is because the lookup table is site specific, and the example uses the lookup table created from Site 1 for Site 5 data. After work on the file containing the suggested values is done, the routine SPC_AGG is run, and new files are created for the four aggregation schemes. Tables 4 through 7 give portions of the final aggregated files for the four schemes. The column "mark" indicates the row for which suggested values were imputed or changes were made by an operator.

CONCLUSIONS AND RECOMMENDATIONS

Whereas PennDOT collects data on site, volume, vehicle classification, and weight, it does not have a system to retrieve data based on month, year, site, and type of analysis required. No system exists that allows for retrieval of No. 4 Card on the basis of the type of aggregation of vehicle classification required for analysis. In addition, no routine is available to flag the erroneous data within LTPP QC that gives the type and duration of errors in a numeric format and replaces the erroneous data checking the performance of ATR and WIM stations. A system was designed that allows for retrieval of ATR, CAVC, and WIM data based on month, year, site, and type of data required for analysis and for the integration of ATR and CAVC data. The QCM

- Reads No. 3 Card, No. 4 Card, and No. 7 Card data in the 1995 TMG format;
- Gives an error flag file along with the time of occurrence of the error;
- Aggregates the No. 4 Card data in four schemes for wider analysis options; and
- Converts No. 4 Card to No. 3 Card for imputing missing daily traffic.

TABLE 5 Aggregation Scheme 2 for CAVC

SHRPID	DIRECT	LANE	YR	MN	DAY	DOW	TIME	C1_3	C4_13	MARK
5	7	2	1994	6	1	4	0	14	7	
5	7	1	1994	6	1	4	0	46	80	
5	3	2	1994	6	1	4	0	8	19	
5	3	1	1994	6	1	4	0	42	72	
5	7	2	1994	6	1	4	1	5	13	
5	7	1	1994	6	1	4	1	60	347	*
5	3	2	1994	6	1	4	1	4	10	
5	3	1	1994	6	1	4	1	25	57	
5	7	2	1994	6	1	4	2	7	10	
5	7	1	1994	6	1	4	2	16	115	
5	3	2	1994	6	1	4	2	6	5	
5	3	1	1994	6	1	4	2	18	62	
5	7	2	1994	6	1	4	3	8	10	
5	7	1	1994	6	1	4	3	21	112	
5	3	2	1994	6	1	4	3	6	8	
5	3	1	1994	6	1	4	3	13	72	
5	7	2	1994	6	1	4	4	11	12	
5	7	1	1994	6	1	4	4	49	118	
5	3	2	1994	6	1	4	4	4	6	
5	3	1	1994	6	1	4	4	21	55	
5	7	2	1994	6	1	4	5	9	10	
5	7	1	1994	6	1	4	5	77	70	
5	3	2	1994	6	1	4	5	6	15	
5	3	1	1994	6	1	4	5	42	79	
5	7	2	1994	6	1	4	6	38	7	
5	7	1	1994	6	1	4	6	163	97	
5	3	2	1994	6	1	4	6	29	10	
5	3	1	1994	6	1	4	6	104	88	
5	7	2	1994	6	1	4	7	38	20	
5	7	1	1994	6	1	4	7	183	125	
5	3	2	1994	6	1	4	7	69	10	
5	3	1	1994	6	1	4	7	186	110	
5	7	2	1994	6	1	4	8	41	16	
5	7	1	1994	6	1	4	8	181	114	
5	3	2	1994	6	1	4	8	61	14	
5	3	1	1994	6	1	4	8	264	85	
5	7	2	1994	6	1	4	9	42	14	

TABLE 6 Aggregation Scheme 3 for CAVC

SHRPID	DIRECT	LANE	YR	MN	DAY	DOW	TIME	C1_3	C4_8	C9_13	MARK
5	7	2	1994	6	1	4	0	14	2	5	
5	7	1	1994	6	1	4	0	46	22	58	
5	3	2	1994	6	1	4	0	8	2	17	
5	3	1	1994	6	1	4	0	42	6	66	
5	7	2	1994	6	1	4	1	5	9	4	
5	7	1	1994	6	1	4	1	60	24	323	*
5	3	2	1994	6	1	4	1	4	1	9	
5	3	1	1994	6	1	4	1	25	5	52	
5	7	2	1994	6	1	4	2	7	3	7	
5	7	1	1994	6	1	4	2	16	29	86	
5	3	2	1994	6	1	4	2	6	1	4	
5	3	1	1994	6	1	4	2	18	9	53	
5	7	2	1994	6	1	4	3	8	5	5	
5	7	1	1994	6	1	4	3	21	35	77	
5	3	2	1994	6	1	4	3	6	1	7	
5	3	1	1994	6	1	4	3	13	10	62	
5	7	2	1994	6	1	4	4	11	8	4	
5	7	1	1994	6	1	4	4	49	33	85	
5	3	2	1994	6	1	4	4	4	0	6	
5	3	1	1994	6	1	4	4	21	6	49	
5	7	2	1994	6	1	4	5	9	5	5	
5	7	1	1994	6	1	4	5	77	21	49	
5	3	2	1994	6	1	4	5	6	1	14	
5	3	1	1994	6	1	4	5	42	8	71	
5	7	2	1994	6	1	4	6	38	4	3	
5	7	1	1994	6	1	4	6	163	30	67	
5	3	2	1994	6	1	4	6	29	1	9	
5	3	1	1994	6	1	4	6	104	13	75	
5	7	2	1994	6	1	4	7	38	15	5	
5	7	1	1994	6	1	4	7	183	49	76	
5	3	2	1994	6	1	4	7	69	0	10	
5	3	1	1994	6	1	4	7	186	24	86	
5	7	2	1994	6	1	4	8	41	7	9	
5	7	1	1994	6	1	4	8	181	38	76	
5	3	2	1994	6	1	4	8	61	4	10	
5	3	1	1994	6	1	4	8	264	22	63	
5	7	2	1994	6	1	4	9	42	8	6	
5	7	1	1994	6	1	4	9	175	29	105	

The SPC_PSU routine uses the lookup table to suggest values to be imputed to the data. The SPC_AGG routine gives the aggregated files, with the desired data, with or without errors at the discretion of the operator, in four aggregation schemes. In addition, missing data are imputed using ATR information.

Future research is expected in using time series for inputting missing data and using the quality assurance methods (*2,3*) for WIM data to obtain a totally automatic comprehensive traffic data management system for analysis, storage, and reporting of traffic volumes, vehicle classification, and weights. This, however, requires first the collection of data at the various PennDOT sites and then the creation of site-specific models that reflect cycles, seasonality, and trends that characterize each site. The approach taken here accounts for the diversity of traffic monitoring sites (e.g., site-specific lookup tables) and attempts to integrate data on traffic volumes, vehicle classification, and vehicle weights. In addition, a data collection scheme to verify the sampling scheme and the data items needs to be designed along the lines proposed elsewhere (*8*).

ACKNOWLEDGMENTS

This research was sponsored by PennDOT and the Mid-Atlantic Universities Transportation Center. Particular thanks are expressed to Vaughn Barnhart of PennDOT for the data used in the design of this system.

REFERENCES

1. Hadley, W. O., C. Copeland, and S. Rowshan. Strategic Highway Research Program–Long-Term Pavement Performance Information Management System. In *Transportation Research Record 1435,* TRB, National Research Council, Washington, D.C., 1994, pp. 1–7.

TABLE 7 Aggregation Scheme 4 for CAVC

SHRPID	DIRECT	LANE	YR	MN	DAY	DOW	TIME	C1_3	C4_5	C6_8	C9_13	MARK
5	7	2	1994	6	1	4	0	14	0	2	5	
5	7	1	1994	6	1	4	0	46	2	20	58	
5	3	2	1994	6	1	4	0	8	0	2	17	
5	3	1	1994	6	1	4	0	42	3	3	66	
5	7	2	1994	6	1	4	1	5	1	8	4	
5	7	1	1994	6	1	4	1	60	8	16	323	*
5	3	2	1994	6	1	4	1	4	1	0	9	
5	3	1	1994	6	1	4	1	25	1	4	52	
5	7	2	1994	6	1	4	2	7	0	3	7	
5	7	1	1994	6	1	4	2	16	2	27	86	
5	3	2	1994	6	1	4	2	6	1	0	4	
5	3	1	1994	6	1	4	2	18	7	2	53	
5	7	2	1994	6	1	4	3	8	1	4	5	
5	7	1	1994	6	1	4	3	21	6	29	77	
5	3	2	1994	6	1	4	3	6	0	1	7	
5	3	1	1994	6	1	4	3	13	3	7	62	
5	7	2	1994	6	1	4	4	11	2	6	4	
5	7	1	1994	6	1	4	4	49	0	33	85	
5	3	2	1994	6	1	4	4	4	0	0	6	
5	3	1	1994	6	1	4	4	21	2	4	49	
5	7	2	1994	6	1	4	5	9	1	4	5	
5	7	1	1994	6	1	4	5	77	1	20	49	
5	3	2	1994	6	1	4	5	6	0	1	14	
5	3	1	1994	6	1	4	5	42	2	6	71	
5	7	2	1994	6	1	4	6	38	2	2	3	
5	7	1	1994	6	1	4	6	163	4	26	67	
5	3	2	1994	6	1	4	6	29	0	1	9	
5	3	1	1994	6	1	4	6	104	7	6	75	
5	7	2	1994	6	1	4	7	38	3	12	5	
5	7	1	1994	6	1	4	7	183	5	44	76	
5	3	2	1994	6	1	4	7	69	0	0	10	
5	3	1	1994	6	1	4	7	186	17	7	86	
5	7	2	1994	6	1	4	8	41	1	6	9	
5	7	1	1994	6	1	4	8	181	4	34	76	
5	3	2	1994	6	1	4	8	61	3	1	10	
5	3	1	1994	6	1	4	8	264	13	9	63	
5	7	2	1994	6	1	4	9	42	2	6	6	
5	7	1	1994	6	1	4	9	175	7	22	105	
5	3	2	1994	6	1	4	9	72	0	3	16	
5	3	1	1994	6	1	4	9	207	15	9	91	

2. Ott, W. C., and A. T. Papagiannakis. Weigh-in-Motion Data Quality Assurance Based on 3-S2 Steering Axle Load Analysis. In *Transportation Research Record 1536,* TRB, National Research Council, Washington, D.C., 1996, pp. 12–18.
3. Han, C., W. T. Boyd, and M. Marti. Quality Control of Weigh-in-Motion Systems Using Statistical Process Control. In *Transportation Research Record 1501,* TRB, National Research Council, Washington, D.C., 1995, pp. 72–80.
4. *Traffic Monitoring Guide.* Office of Highway Management Systems, FHWA, U.S. Department of Transportation, 1995.
5. *LTPP Quality Control Analysis, User's Manual.* 1996.
6. Hallenbeck, M. A Description of the Long-Term Pavement Performance Project Traffic Database. *Post-Proceedings, First European Conference on Weigh-in-Motion of Road Vehicles,* Swiss Federal Institute of Technology, Zurich, Switzerland, 1995, pp. 285–293.
7. *Vehicle Travel Information System (VTRIS), User's Guide.* Office of Highway Management Systems, FHWA, U.S. Department of Transportation, 1996.
8. Goulias, K. G., K. Viswanathan, and J. Chung. *Truck Weight Data Collection in Pennsylvania.* Draft final report. The Pennsylvania Transportation Institute, Pennsylvania State University, University Park, Pa., 1997.

The paper does not constitute a policy or regulation of the Commonwealth of Pennsylvania.

Publication of this paper sponsored by Committee on Statewide Transportation Data and Information Systems.