TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JTRP-2003/16		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
Weigh-In-Motion Data Checking and Imputation	5. Report Date: August 2003 6. Performing Organization Code
7. Author(s) Ting Wei and Jon D. Fricker	8. Performing Organization Report No. FHWA/IN/JTRP-2003/16
9. Performing Organization Name and Address Joint Transportation Research Program 1284 Civil Engineering Building Purdue University West Lafayette, IN 47907-1284	10. Work Unit No. 11. Contract or Grant No. SPR-2470
12. Sponsoring Agency Name and Address Indiana Department of Transportation State Office Building 100 North Senate Avenue Indianapolis, IN 46204	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

Prepared in cooperation with the Indiana Department of Transportation and Federal Highway Administration.

16. Abstract

There are about 46 weigh-in-motion (WIM) stations in Indiana. When operating properly, they provide valuable information on traffic volumes, vehicle classifications, and axle weights. Because there are great amounts of WIM data collected everyday, the quality of these data should be monitor without further delay. The first objective of this study is to develop effective and efficient methods to identify missing or erroneous WIM data. The second objective is to develop a data imputation method to update the missing or erroneous data.

This report describes the WIM data checking process on both a monthly and a daily basis. The Weigh-In-Motion Daily Data Checking (WDDC) program is introduced. The whole procedure requires very little human intervention, and provides a convenient way to check daily summary data. This report also describes several imputation methods in the experiment of imputing 7-day data.

17. Key Words	18. Distribution Statement
Weigh-In-Motion, data quality, data checking, imputation.	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	66	

Form DOT F 1700.7 (8-69)

INDOT Research

TECHNICAL *Summary*

Technology Transfer and Project Implementation Information

TRB Subject Code: 53-9 Weigh-In-Motion
Publication No.: FHWA/IN/JTRP-2003/16, SPR-2470

August 2003
Final Report

Weigh-In-Motion Data Checking and Imputation

Introduction

WIM systems are usually used to detect illegally overweight vehicles and to collect traffic data for highway planning and management purposes. The purpose of this study is to identify missing or erroneous data and to develop a data imputation method to update these data. WIM data from the year of 1997 to 2001 for sites 4260 and 4270 are used in the analysis.

Findings

WIM data checking should be conducted on both a monthly basis and a daily basis. The three methods using unclassified vehicle rate, front axle distribution and Class 9 vehicle GVW are widely accepted. The WDDC (Weigh-In-Motion Daily Data Checking) program is developed for INDOT to facilitate the daily checking process.

In the experiment of imputing 7-day data, the MAPE for the factor method and regression methods are within the range of 15-20 percent.

Implementation

Throughout the data analysis for this project, we realize how much important information the Weigh-In-Motion system can provide. However, the data quality often suffers from equipment problems. In addition, this project has been hampered by the lack of historical data. As more historical data can be retrieved, the ability to impute data can be more comprehensively assessed.

In the meantime, the data checking procedures developed in this project should facilitate the prompt detection of apparent data anomalies and the application of appropriate connective action. In the process, the amount of poor data can be reduced, with a corresponding reduction in the need for data imputation.

Contacts

For more information:

Prof. Jon Fricker
Principal Investigator
School of Civil Engineering
Purdue University
West Lafayette IN 47907
Phone: (765) 494-2205
Fax: (765) 496-7996
E-mail: fricker@ecn.purdue.edu

Indiana Department of Transportation
Division of Research
1205 Montgomery Street
P.O. Box 2279
West Lafayette, IN 47906
Phone: (765) 463-1521
Fax: (765) 497-1665

Purdue University
Joint Transportation Research Program
School of Civil Engineering
West Lafayette, IN 47907-1284
Phone: (765) 494-9310
Fax: (765) 496-7996

Final Report

FHWA/IN/JTRP-2003/16

Weigh-In-Motion Data Checking and Imputation

By

Ting Wei
Graduate Research Assistant

and

Jon D. Fricker
Professor

School of Civil Engineering
Purdue University

Joint Transportation Research Program
Project No: C-36-17FFF
File No: 8-4-58
SPR-2470

Conducted in Cooperation with the
Indiana Department of Transportation
and the
U.S. Department of Transportation
Federal Highway Administration

Purdue University
West Lafayette, IN 47907
August 2003

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1. Weigh-In-Motion in Indiana

Weigh-in-Motion (WIM) is defined as "the process of measuring the dynamic tire forces of a moving vehicle and estimating the corresponding tire loads of the static vehicle" by ASTM (1994). WIM systems serve two very important functions:

1. Detection of illegally overweight vehicles
2. Data collection for highway planning and management purposes

It is widely accepted that heavy trucks cause major damage to highways. Weigh stations with static scales have traditionally been used to weigh commercial carriers using the highways. When a weigh station is in operation, all passing trucks are required to stop at the weigh station and have their weights recorded. Based on the industry standard cost of one dollar per minute of delay at a weigh station, it was estimated that weigh station stops could cost the nation's trucking industry over 10 billion dollars (Bergana et al. 1998).

A WIM system automatically weighs trucks when they are traveling at highway speed, so it can be an important enforcement tool for screening trucks and targeting potential violators. Because only suspected trucks are directed to a static scale, WIM can focus enforcement resources by concentrating on non-compliant trucks, while reducing the delay costs to compliant trucks. Green et al. (2002) described some enforcement practices using a "Virtual Weigh Station" by Indiana State Police.

Another advantage of WIM systems is that they can be used to collect traffic data at all times, even when static weigh station is closed. Drivers of overweight vehicles may try to find the operation schedule of weigh station and delay their travel until the station is closed. However, if a WIM system is installed near the weigh station, the overweight trucks will be recorded.

The data continuously collected from WIM systems has proven to be valuable information for highway planning and management. Indiana has about 46 WIM stations on major highways, forming a statewide WIM data collection system. When operating properly, they provide valuable information on traffic volumes, vehicle classifications, and axle weights. This information is an important part of the data collected for the Highway Performance Monitoring System (HPMS) and is a key element of INDOT's warranty projects. The WIM data are also collected for the nationwide Long-Term Pavement Performance (LTPP) project. Gulen et al. (2000) updated the ESAL values for single unit trucks and multiple unit trucks, using the 1998 and 1999 traffic data collected from Indiana WIM stations.

There are three common technologies used in WIM systems: bending plates, piezoelectric sensors, and load cells. Bushman and Pratt (1998) compared the three types of technologies with respect to accuracy, life span and cost, as summarized in Table 1.1.

Table 1.1 Comparison of Common WIM Technologies

	Piezoelectric	Bending Plate	Single Load Cell
Accuracy (95% confidence)	± 15 %	± 10 %	± 6 %
Expected Life	4 Years	6 Years	12 Years
Initial Installation Cost	$9000	$21500	$48700
Annual Life Cycle Cost	$4750	$6400	$8300

Most of Indiana WIM stations are using piezoelectric sensors or single load cells. INDOT has contracts with International Road Dynamics (IRD), Inc. for all the WIM systems in Indiana. IRD also provides a software package for INDOT to download the vehicle record files from remote WIM sites and to process the data to generate summary reports.

1.2. Definition of WIM Data Quality Problem

There are two kinds of data quality problems discussed in this report: missing data and erroneous data.

1.2.1. Missing Data

It is not unusual for a WIM site to fail to record any data for a specific period. This kind of problem can be caused by a power outage or a loop malfunction. Because INDOT downloads the vehicle record file on a daily basis, if the situation lasts a whole day, there will no file for this day; if the situation lasts for a few hours in a day, the file will only have the records for the other hours of the day.

1.2.2. Erroneous Data

Erroneous data is defined as the data that don’t reflect the actual traffic conditions. By “erroneous”, we mean the WIM data differ significantly from what we expect from a properly performing WIM station. There are many possible reasons for erroneous data, such as pavement condition, improper calibration, or lack of calibration.

In order to detect data that are present but erroneous, we can convert the raw data to some kind of summary report, and look for suspicious values of representative variables (e.g., gross vehicle weight). An extreme situation can be that a WIM site records all vehicles with 0 gross vehicle weight. Such a situation can be easily characterized as erroneous. In real life, we may see other situations with suspiciously high or low gross vehicle weights, and we have to conduct further analysis before making a conclusion.

1.3. Purpose of Project

There are two major objectives of this study. The first objective is to develop effective and efficient methods to identify missing or erroneous WIM data. Because highway planners and designers rely on WIM data, the quality of the WIM data must be assured. Currently INDOT personnel inspect only the monthly summary report to identify possible data quality problems, because daily checking would require more time than is available to the personnel responsible. In this report, we will explore the techniques used in monthly data checking (Chapter 2) and the methods possible for daily data checking (Chapter 3). As a result of the checking process, all suspicious data should be flagged for appropriate action. The second objective is to develop a data imputation method to update the missing or erroneous data. We will test and compare several candidate methods in data imputation.

There has been much debate about data imputation and base data integrity. AASHTO Guidelines for Traffic Data Programs states: “Some current traffic editing programs estimate missing or edit rejected data. This practice, termed ‘imputation,’ is not recommended” (AASHTO 1992). This recommendation is based on the concern of base data integrity and the justification that imputing missing values introduces “errors which cannot be quantified” (AASHTO 1992).

However, leaving gaps where the missing data and erroneous data are identified may cause the whole data set to be far from complete, and may introduce biases. The Mobility Monitoring Program of the Texas Transportation Institute (TTI) reports that, after screening erroneous data, Transportation Management System (TMS) data archives can be anywhere from 16 percent to 93 percent complete. The median value in this study was 67 percent (Lomax et al. 2001).

The use of the remaining undiscarded data can introduce bias into the incomplete data set, because the data that have been discarded may represent a certain kind of traffic pattern that doesn't exist in the remaining dataset. For example, if we estimate average annual daily ESAL for a WIM station that has good data through the year except in January and December, the ESAL value could be higher than the actual value, because January and December are the two months that have the least truck traffic during the year.

Although INDOT has been using ESAL values for pavement design, INDOT currently doesn't employ any techniques for "replacing" missing or erroneous data. In a recent effort to update the average ESAL estimate values for multiple unit trucks and single unit trucks for INDOT, WIM data were refined by deleting the data based on several quality criteria (Gulen et al. 2000).

1.4. Literature Review

It has been recognized that the quality of WIM data can affect pavement design dramatically. It was reported (FHWA 1998) that the basic trend is that every 1 percent error by which a scale is under-calibrated results in slightly more than a 3 percent under-estimation of the true ESAL value. Every 1 percent over-estimation in axle weight represents a 4.5 percent over-estimation of ESAL values. Thus, even an over-calibration of only 10 percent would result in a 45 percent error in estimated damage.

Therefore, it is important to do the system calibration properly. The American Society for Testing and Materials (1994) has published standard specifications for the calibration procedure for highway WIM system, which includes the acceptance and initial calibration processes. States' Successful Practices Weigh-In-Motion Handbook (McCall and Vodrazka 1997) also provides a standardized step-by-step procedure for WIM calibration.

Moreover, calibration must be done routinely because electronics do drift over time. Zhi et al. (1999) evaluate the performance of a WIM system in Manitoba, Canada, and found that about 90 percent of truck weights were underestimated and the degree of underestimation was higher than 50 percent of the static weights, due to the reasons of unstandardized calibration procedures and a drift in calibration over time. This finding highlights the importance of quality control and corrections on WIM data prior to their use in research or engineering practice.

In order to detect possible WIM data quality problems, there are two kinds of methods used to check WIM data. The first type involves comparing data values to a confirmed range. For example, the WIM axle-spacing data sets should fall with the range specified by ASTM Standard E1572. There are other variables that can be selected, such as front axle weight or gross vehicle weight (GVW). The second type involves plotting data values of a specified period for serial check or graphic inspection. Generally we can use daily GVW or ESAL. Schmoyer and Hsu (1997) used a change-point algorithm based on the statistic:

$$T = 200 \times \left| \frac{\text{mean for 2weeks post} - \text{mean for 2weeks prior}}{\text{mean for 2weeks post} + \text{mean for 2weeks prior}} \right|$$

Schmoyer and Hsu stated that, “A change in the series is suggested at any point for which the mean for the last two weeks is appreciably different from the mean for the next two weeks. “Appreciably different” must be defined, of course, and should achieve a reasonable balance of false positives and false negatives.” The advantage of this method is to avoid the reliance on human eyes to do the serial check.

There is a large amount of research related to missing values. In the book Statistical Analysis with Missing Data, Little and Rubin (2002) summarized most of the up-to-date approaches that handle missing data. There is some research addressing how missing data are handled by transportation practitioners. Nihan and Holmesland (1980) used Box-Jenkins techniques to predict short-term traffic for urban freeways. Their model was able to predict average weekday volumes for two months, in which entire monthly data was missing. Zhong et al. (2002) analyzed the data from Alberta DOT and Minnesota DOT, and found that genetically designed regression models based on data from before and after the failure had the most accurate results. Average errors for refined models were lower than 1 percent, and the 95th percentile errors were below 2 percent for counts with stable patterns. Smith et al. (2003) provided a comparison of heuristic techniques and statistical techniques and indicated that the more sophisticated statistical techniques may generate better imputations. Smith et al. also suggested that the transportation profession seriously reconsider the AASHTO "policy" of not imputing traffic data in order to provide the user as much information as possible.

1.5. Data Used in this Study

In order to analyze the characteristics and patterns of WIM data, it has been recognized that there must be enough WIM data available, especially for several continuous years. INDOT provided us with 5 years (from 1997 to 2001) of WIM data for Stations 4260 and 4270. These data were used throughout the study. The following figures are the plots of the average daily GVW for class 9 vehicles from summary reports.

For station 4260, as we can see in Figure 1.1-1.5, there are a few periods that have missing data or zero GVW. The year of 1998 seems to have the most complete data set. However, there is a big change happening during July 1998 (Figure 1.2). This kind of change also happened in March 1999 and July 2001.

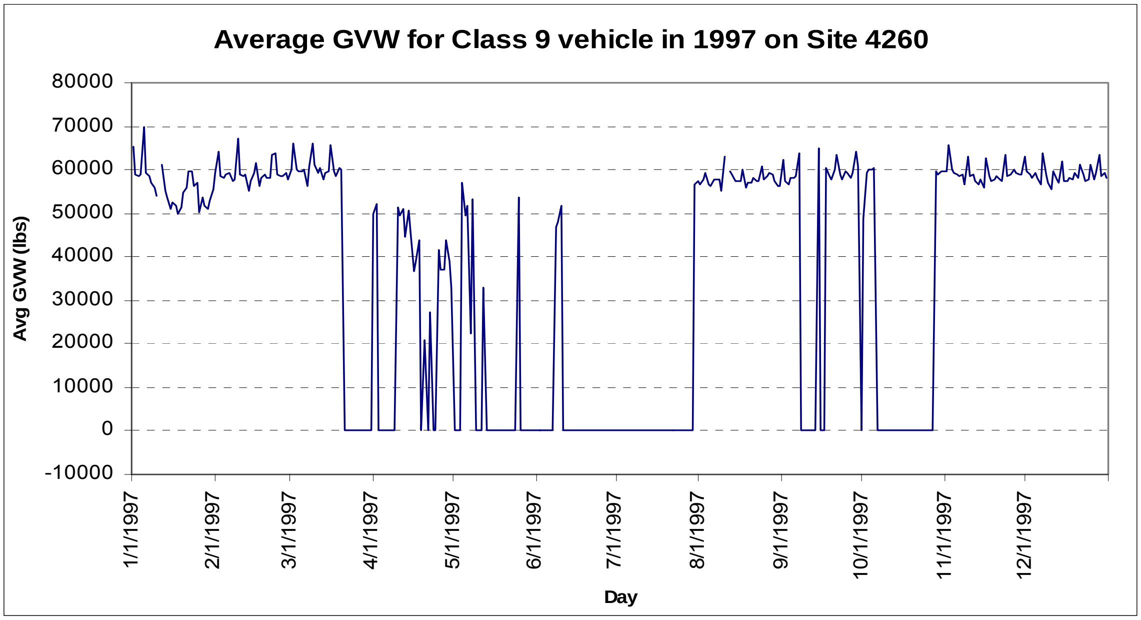


Figure 1.1 AGVW in 1997 on Site 4260

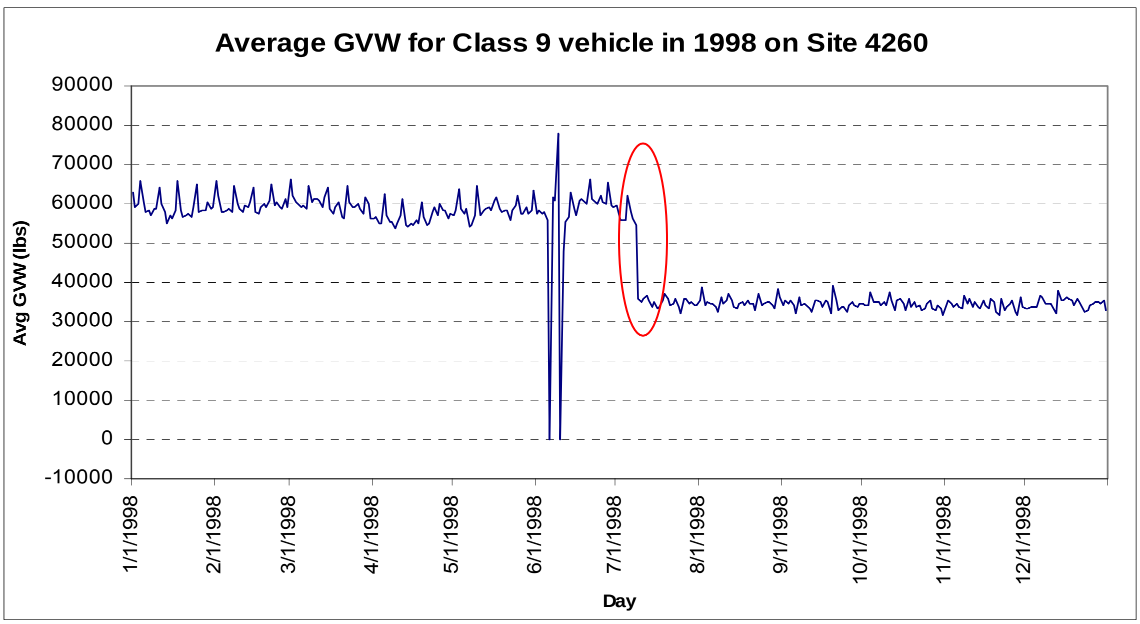


Figure 1.2 AGVW in 1998 on Site 4260

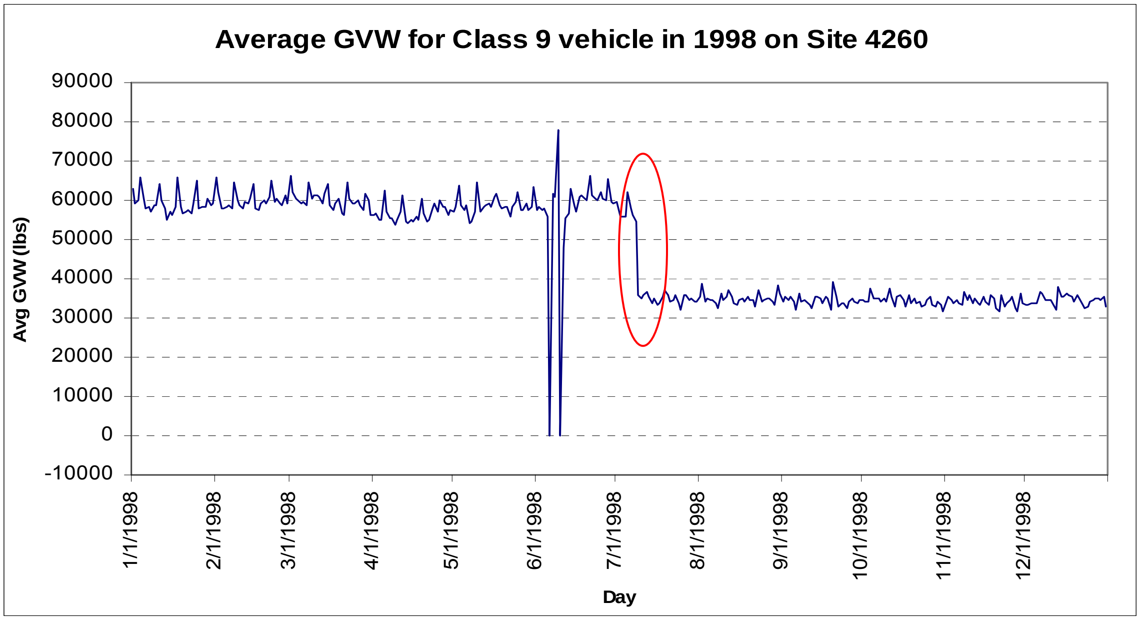


Figure 1.3 AGVW in 1999 on Site 4260

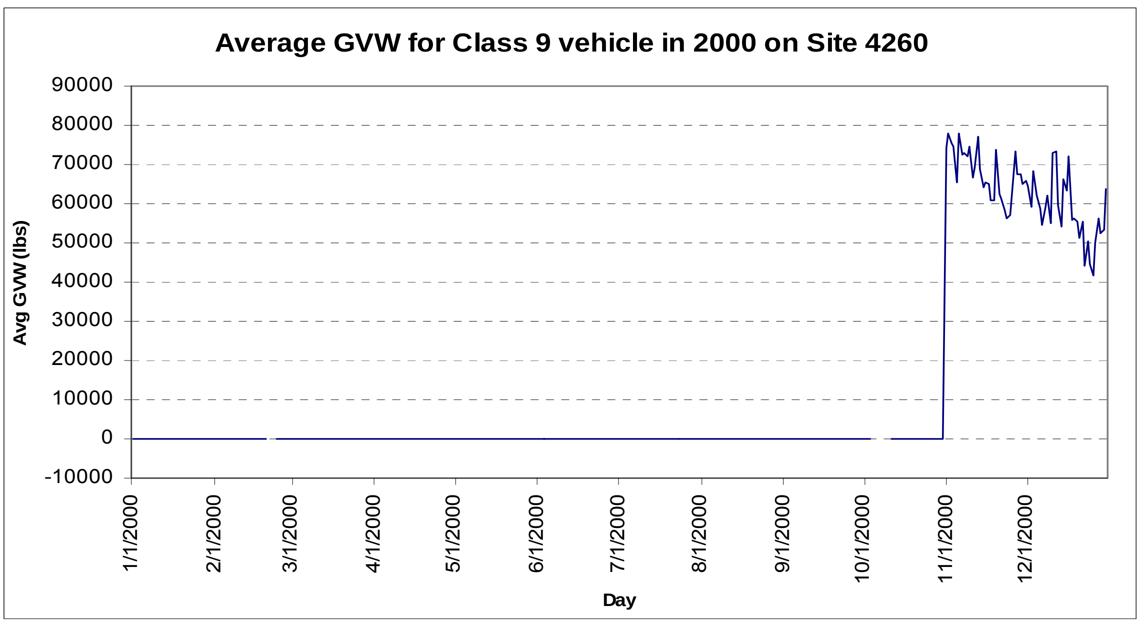


Figure 1.4 AGVW in 2000 on Site 4260

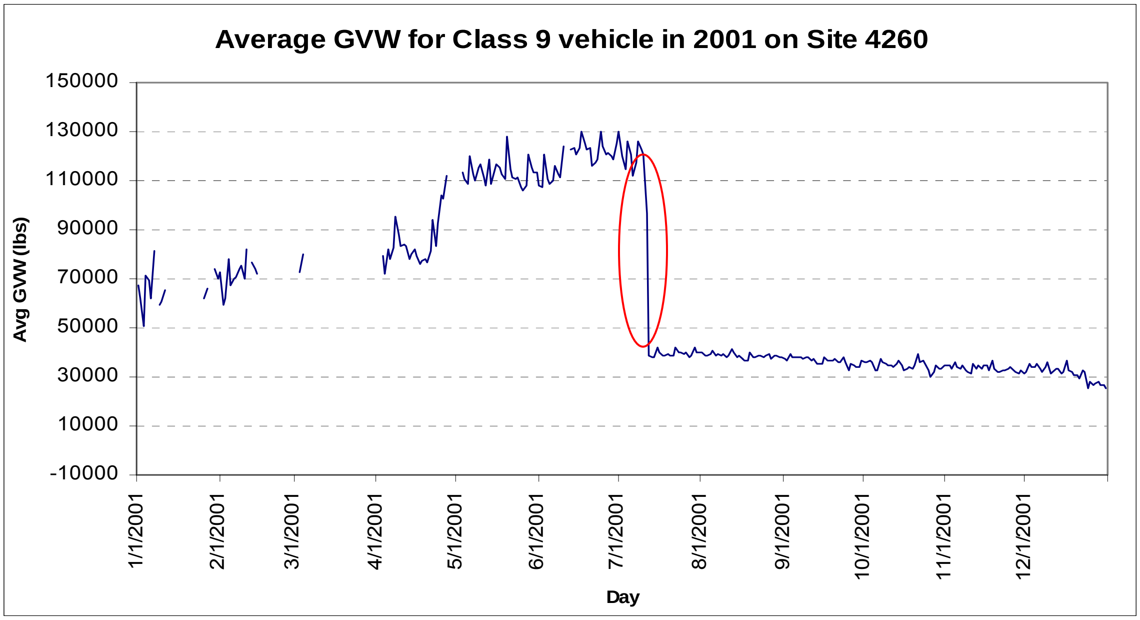


Figure 1.5 AGVW in 2000 on Site 4260

For station 4270, as we can see in Figures 1.6-1.10, the WIM data has overall better quality, except a 5-month period of zero GVW (Figures 1.7 and 1.8) and a 2-month period of missing data (Figure 1.10).

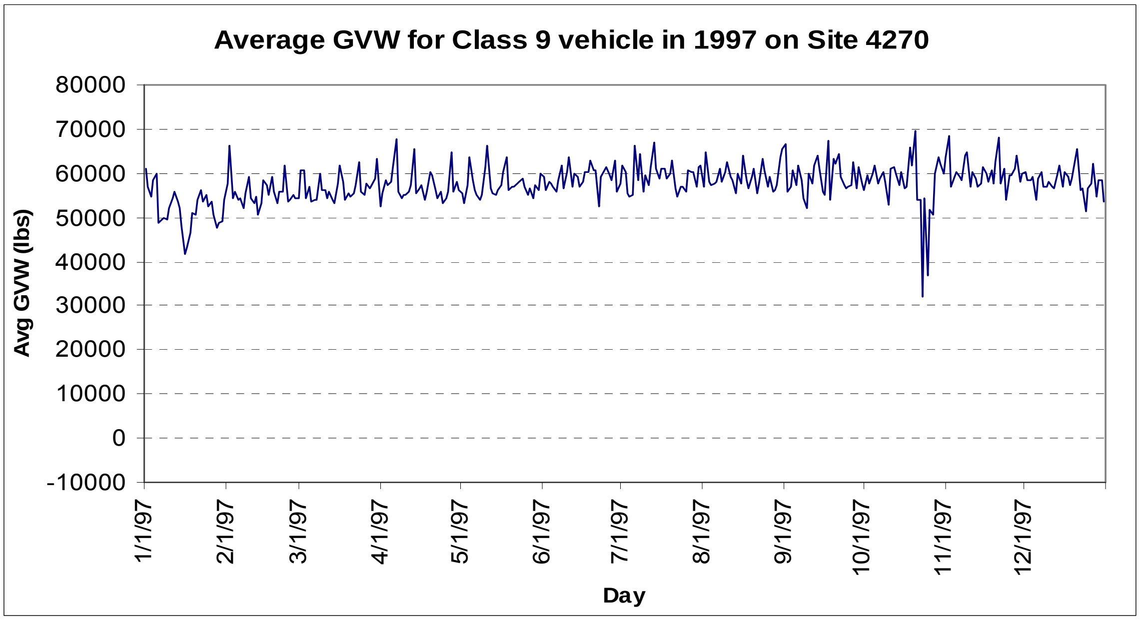


Figure 1.6 AGVW in 1997 on Site 4270

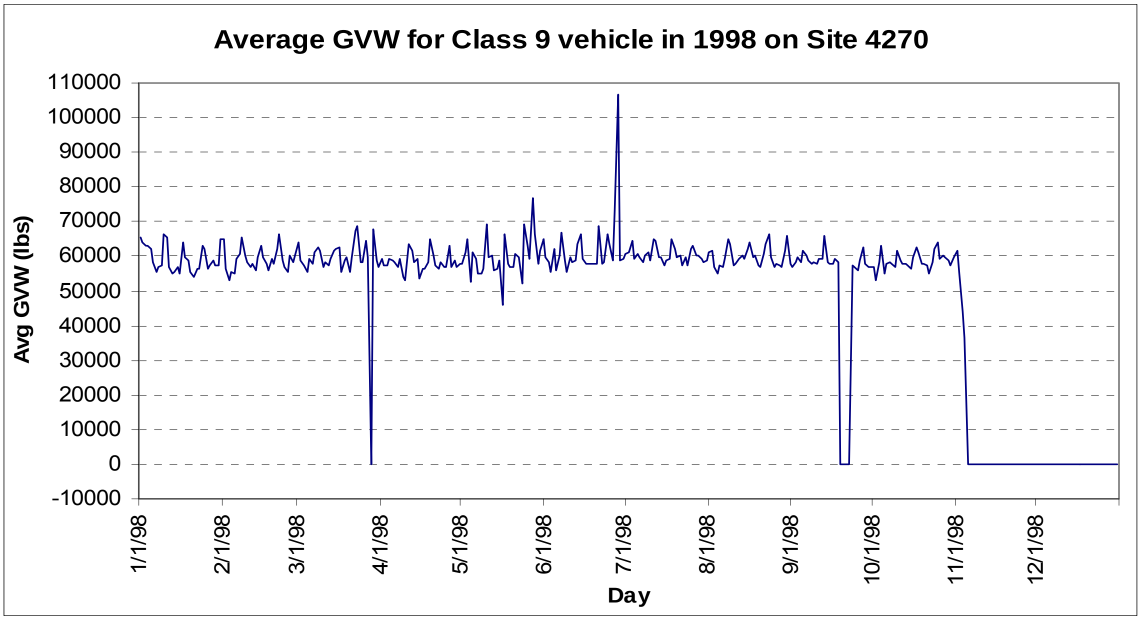


Figure 1.7 AGVW in 1998 on Site 4270

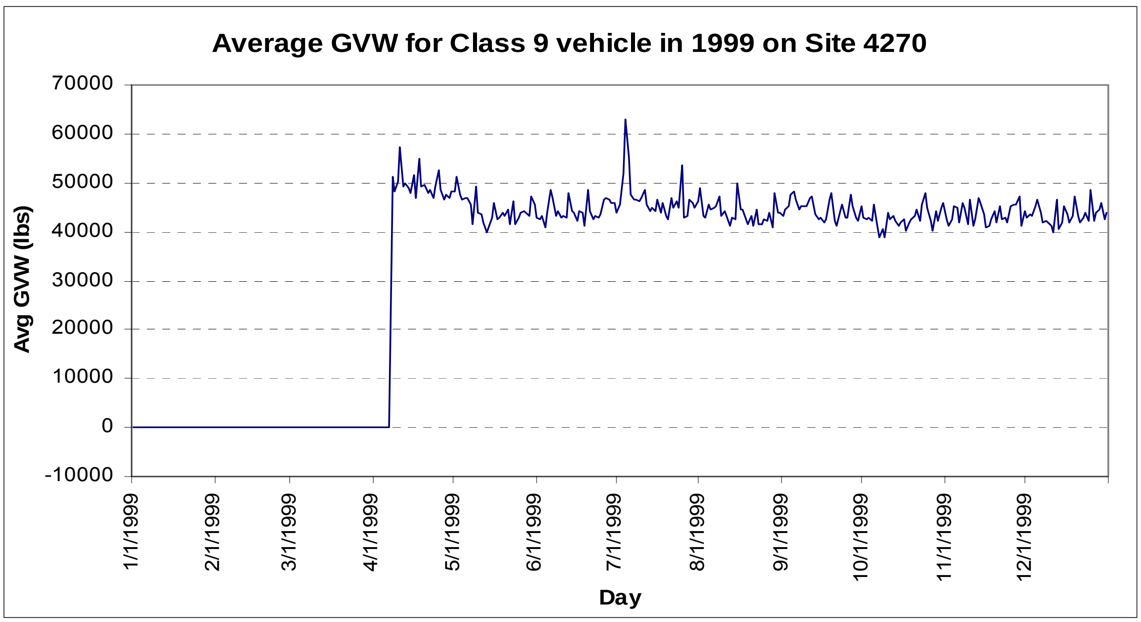


Figure 1.8 AGVW in 1999 on Site 4270

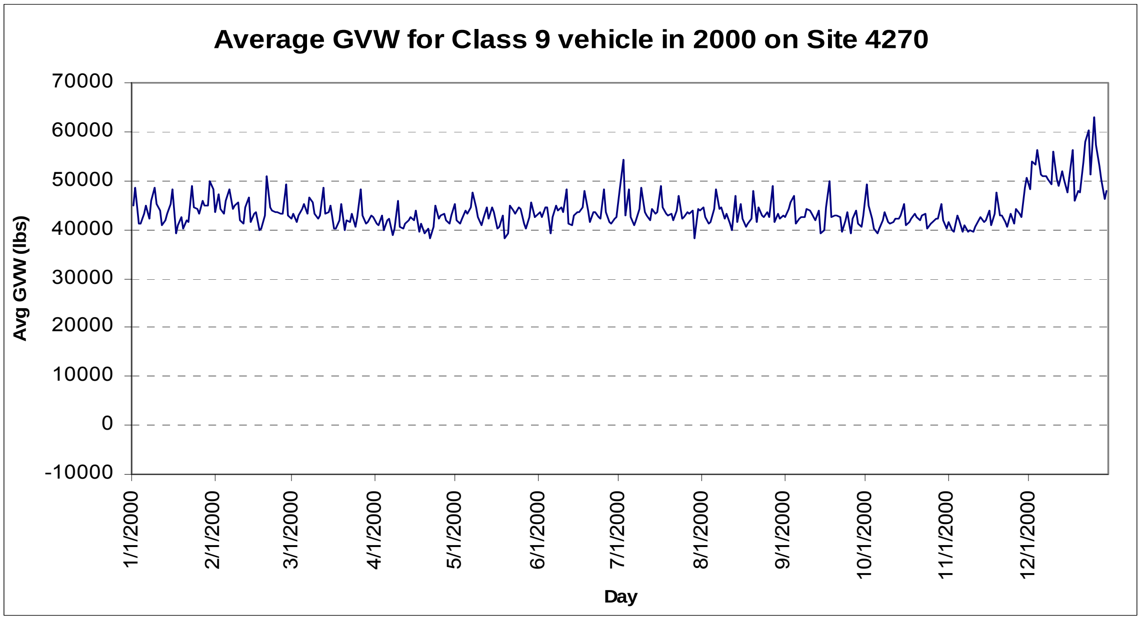


Figure 1.9 AGVW in 1999 on Site 4270

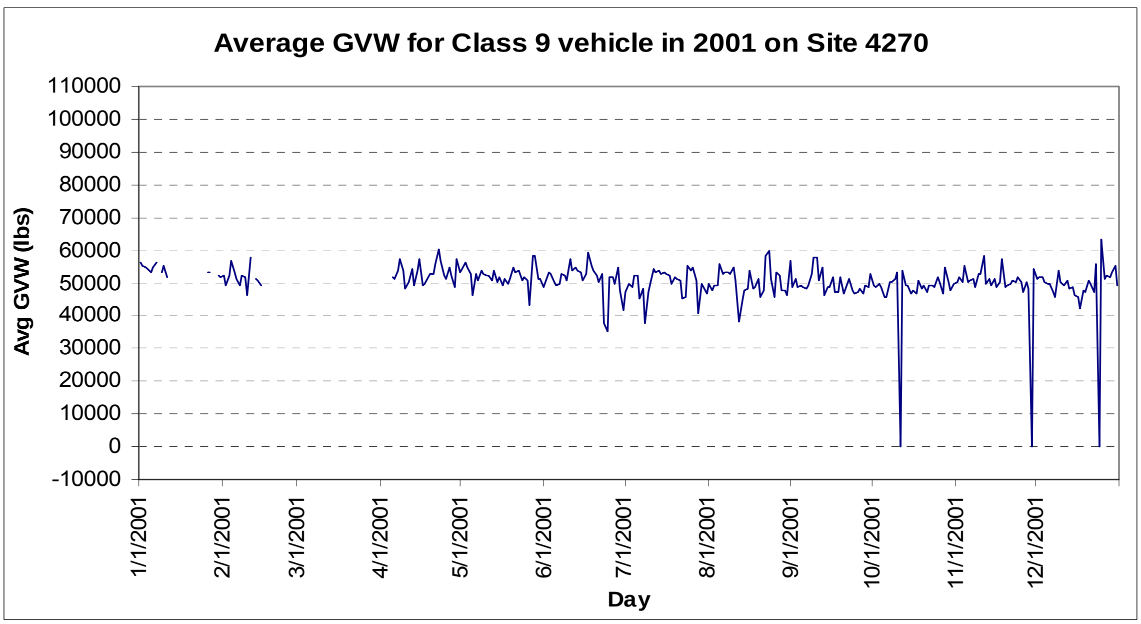


Figure 1.10 AGVW in 2001 on Site 4270

During the year of 2002, INDOT began to continuously provide WIM data for all Indiana WIM stations to Professor Darcy Bullock at Purdue University. The use of these data will be discussed in the following chapters.

CHAPTER 2. MONTHLY CHECKING FOR WIM DATA

2.1. Monthly Summary Reports

2.1.1. Selection of Monthly Summary Reports

The raw data downloaded from WIM stations have been encrypted by IRD, and can't be viewed directly. In this project, we use the IRD Office (version 7.5.0) to process these data, focusing on truck information. One way to do this is to retrieve vehicle-by-vehicle information, as shown in Figure 2.1.

```
Site: RENSSELAER-411

(69) LANE #1  TYPE 9  GVW 70.8 kips  LENGTH 60 ft
18-K ESAL 1.467  SPEED 55 mph  MAX GVW 80.0 kips  Sat Mar 02 00:04:22.32 1996
     UNIT    SEPARATION     WEIGHT    ALLOWABLE
                (ft)        (kips)      (kips)
       1                     11.1        20.0
       2         9.0         15.4        17.0
       3         4.2         14.0        17.0
       4        35.2         13.3        17.0
       5         4.0         17.0        17.0
```

Figure 2.1 Vehicle Passage Information by IRD Office

We can now view all the information the WIM system has recorded for each individual truck. The ESAL is automatically calculated using the AASHTO load equivalency table or formula (selected by user) in the AASHTO Guide for Design of Pavement Structures, 1986 (IRD 1993).

Another way to retrieve the truck information is to generate customized summary reports. In this project, because we are specifically interested in daily statistics, we only use summary reports. The user can select a preferred format and appropriate variables. Table 2.1 lists the report formats provided by IRD Office:

Table 2.1 List of Available Report Formats

Class by Hour	FHWA TMG Card Reports
Speed by Hour	Autocalibration
Lane by Hour	Site Summary
Lane by Class	IRD ASCII Vehicles Records
Error Vehicle by Hour	Single Axles
Class by Day of Month	Tandem Axles
Truck Count by Day of Month	Tridem Axles
Class by Front Axle Weight	Quadrem Axles
Class by Gross Vehicle Weight	Axle Count by Axle Weight
Class by Overweight Vehicles	Site History
Weight Violations by Hour	Power Log
18 KIP ESALs by Hour	Calibration Log

Users can select any combination of these formats in their summary report according to their needs. In this project, "site summary" contains most of the information we need. Table 2.2 is an excerpt of a site summary report:

Table 2.2 Example of Site Summary Report

Site Summary Report								
Site: RENSSELAER-411	Lanes: #1							
Classification: FHWA			Start Class 0 End Class 16					
FROM: Wed Jan 01 00:00:00 1997 TO: Thu Jan 02 00:00:00 1997								
Classification								
	0	1	2	……	9	……	13	Total
Veh Cnt	1496	0	509	……	547	……	196	6334
% Count	23.6	0	8	……	8.6	……	3.1	100
Recorded	2024	165	803	……	470	……	33	4911
Warn Cnt	315	6	10	……	233	……	19	781
% Warn	15.6	3.6	1.2	……	49.6	……	57.6	15.9
Valid Wt	1709	159	793	……	237	……	14	4130
TGW(000)	27074	845	7105	……	6314	……	657	59976
Avg GVW	15842	5317	8960	……	26641	……	46955	14522
Tot ESAL	89.2	0.4	2.4	……	59.1	……	3.4	189.9
Avg ESAL	0.05	0	0	……	0.25	……	0.25	0.05
Ovrwt	0	0	0	……	8	……	0	11
% Ovrwt	0	0	0	……	3.4	……	0	0.3
Vehicle records used in this report include:								
Good + Warning + Error vehicles								
Total Counts As Defined By Report Parameters:								
Error Warning Stored Total								
842(13.3%) 781(12.3%) 5753(90.8%) 6334								

In Table 2.2, we can see that IRD Office uses the FHWA classification scheme to classify vehicles (see Appendix A). If a vehicle doesn't fit any of the 13 classes, it will be assigned to class 0 vehicle.

The summary reports can be generated for the intervals of 5 min, 15 min, hourly, daily, weekly, and monthly, depending on the user's needs.

2.1.2. Generating Monthly Summary Reports

There are two ways to generate summary reports. One is to use IRD Office's menu system; the other is to use a batch file. In this project, we use a batch file to generate summary reports, because a batch file can handle the data from all sites, while the menu system can handle only one site's data at a time. For detailed instructions on how to generate summary reports, refer to IRD Software User's Manual, 1993.

The summary reports generated are stored in ASCII files with file extensions corresponding to the station numbers minus the last zero. For example, the summary reports for station 7340 will be stored in an ASCII file as *. 734. If we generate a monthly summary report using daily interval, the ASCII file will have all the reports of each day in the month, unless the data are missing.

2.2. Evaluating WIM Data Quality

In the rest of this chapter, we will explore how to evaluate the WIM data quality by reviewing the monthly summary reports.

2.2.1. Using Unclassified Vehicle Rates

Unclassified vehicles, or class 0 vehicles, can be used to represent how well the WIM system is working. If the rates of unclassified vehicles exceed a specified level, say 10 percent, for several continuous days, this is evidence of possible malfunctioning of the WIM site. For example, during February 1999 (Figure 2.2), the unclassified vehicle rates for station 4270 are all below 10 percent. However, during February 2000 (2.3), there are several days that have unclassified vehicle rates over 30 percent. By inspecting unclassified vehicle rates, we find the data from February 2000 for station 4270 are suspicious.

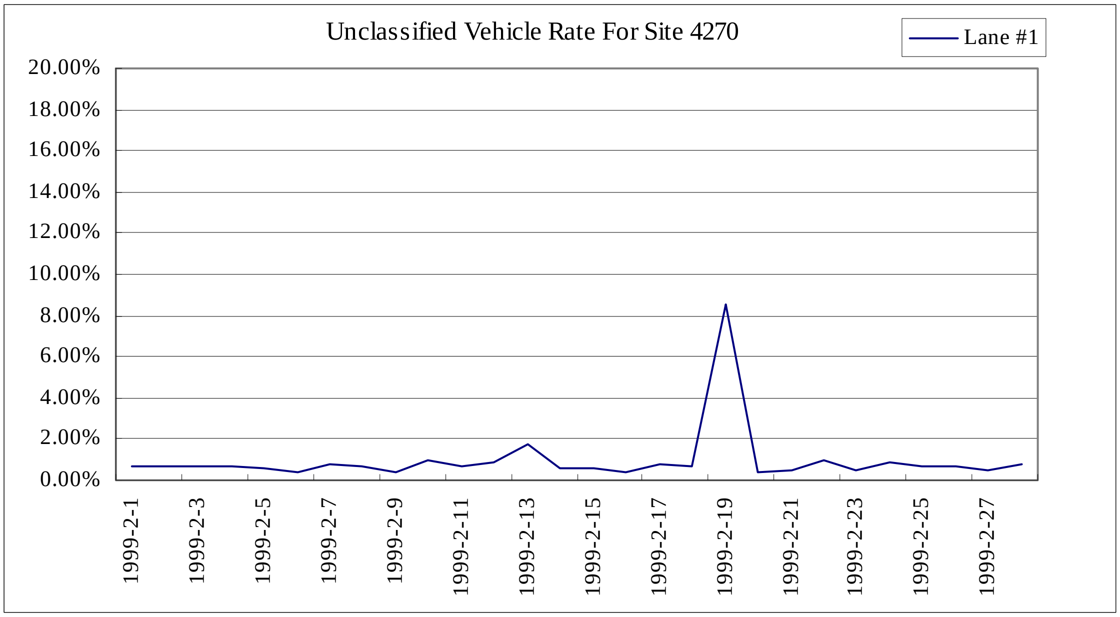


Figure 2.2 Unclassified Vehicle Rate In February 1999

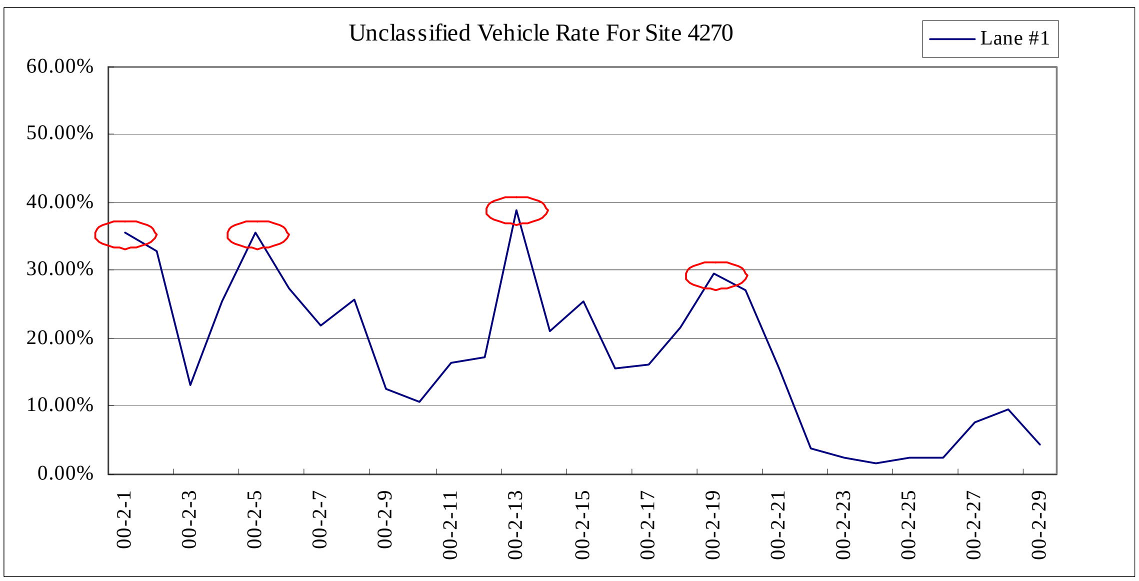


Figure 2.3 Unclassified Vehicle Rate In February 2000

As we further our analysis on the data in February 2000. We find that the changes in the unclassified vehicle rates are often related to the changes of class 2 and class 9 vehicle rates.

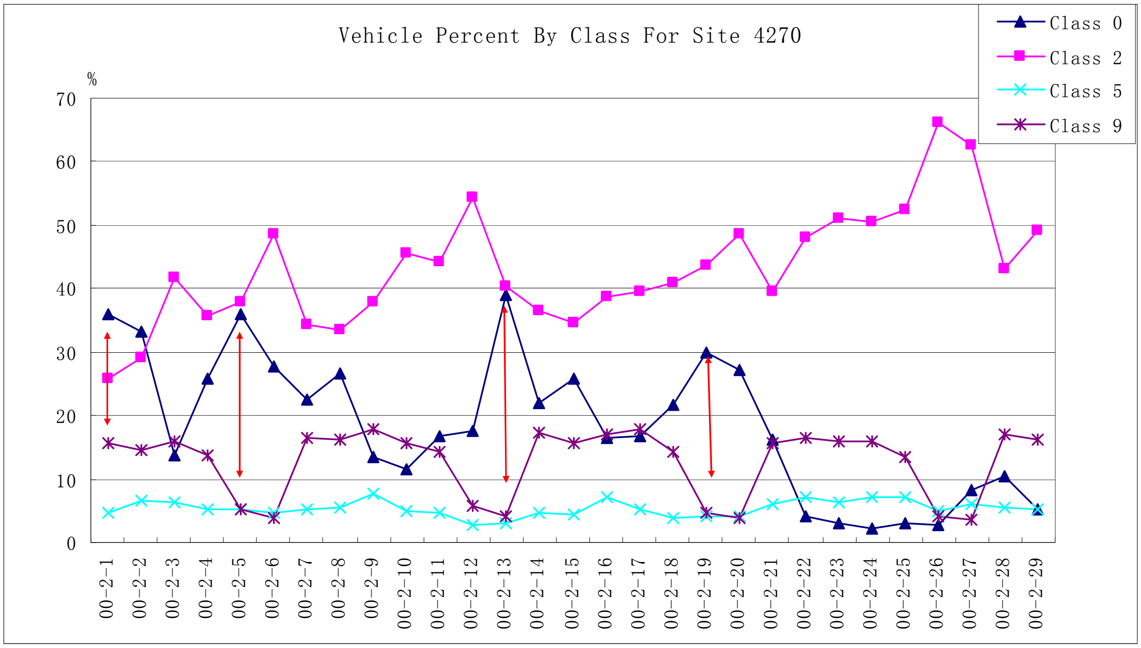


Figure 2.4 Vehicle Percent by Class In February 2000

As we can see in Figure 2.4, the class 0 vehicle rates often reach peaks during the weekends. Because the class 2 vehicle rates also have high values during weekends, it is possible that the increase of unclassified vehicles is from class 2 vehicles.

2.2.2. Using Front Axle Weight Distribution

For a properly working WIM station, the front axle weight of most class 9 vehicles should fall within the 8-12 kip range. Based on this criterion, we can inspect the front axle weight distribution, as shown in Figure 2.5 and Figure 2.6.

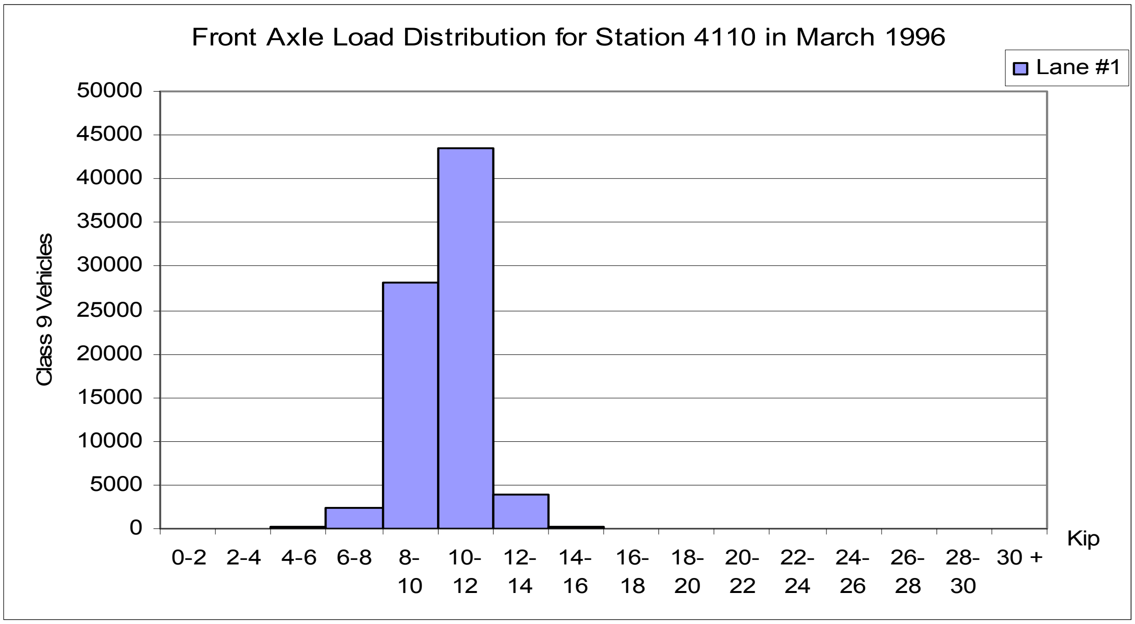


Figure 2.5 Front Axle Load Distribution for Site 4110 in March 1996

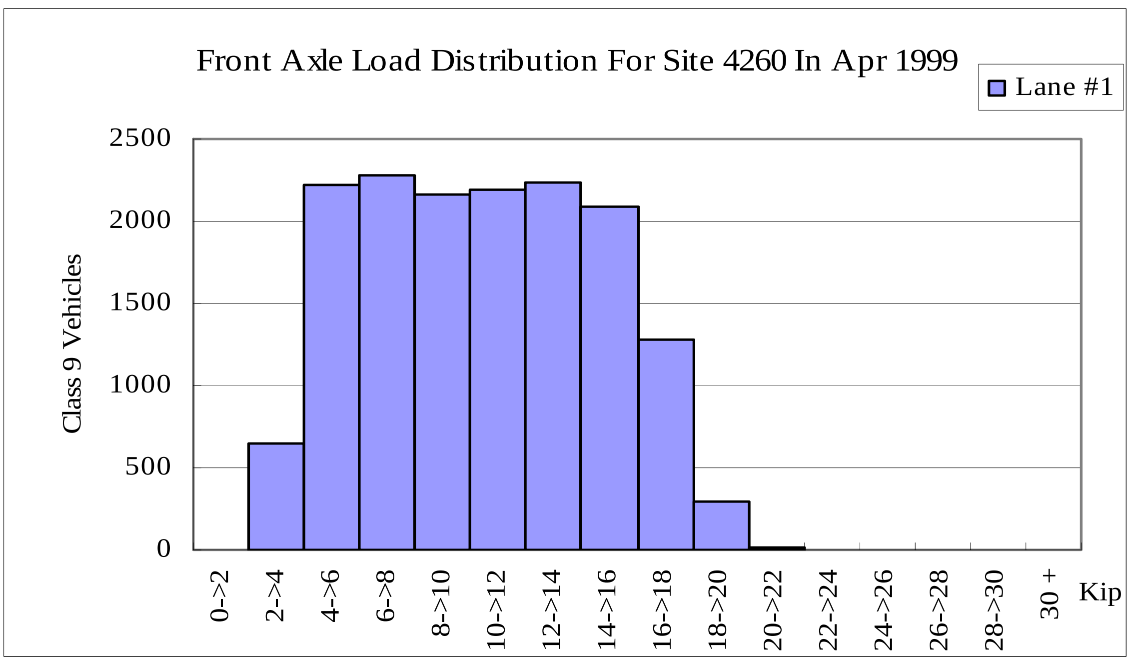


Figure 2.6 Front Axle Load Distribution for Site 4260 in April 1999

Figure 2.5 fits the criterion very well, while the data in Figure 2.6 look suspicious. Because the distribution expand to both lower and higher ranges, we suspect that the problem can be related to the sensor installed.

2.2.3. Using Average Gross Vehicle Weight

The average daily GVW for class 9 vehicles is plotted over the month. The average gross vehicle weight of class 9 vehicles should fall within the range of 25-80 kip. 80K is the legal limit; 25 kip is the weight that even an unloaded class 9 vehicle should exceed. In the plots, we can look for any appreciable changes in GVW, which may indicate a situation that needs immediate investigation.

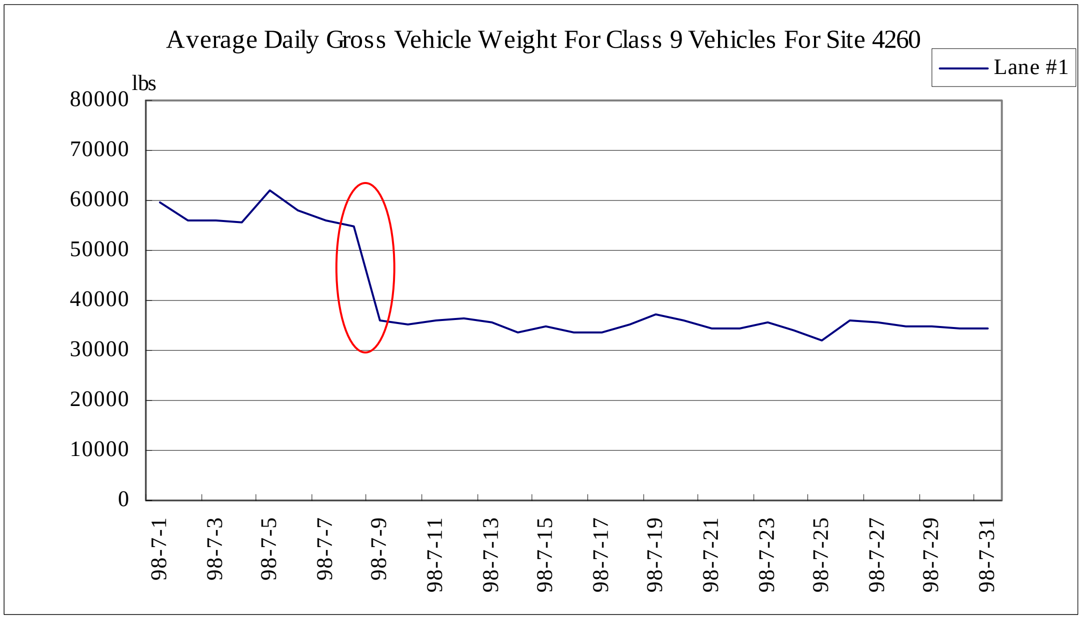


Figure 2.7 Average Daily Gross Vehicle Weight In July 1998

As we see in Figure 2.7, on July 9th 1998, the GVW decreased from the 50k-60k range to the 30k-40k range. This change is more than 40 percent.

For some situations, a shift in GVW patterns may not be immediately detected by a manual inspection. For example, the daily GVW values may be a function of the day of the week. The fluctuations between workdays and weekends may hinder a manual inspection to identify a shift. In this case, we would recommend calculating the 7-day average for each day i:

$$\overline{GVW_i} = \frac{1}{7} \times (GVW_{i-6} + GVW_{i-5} + GVW_{i-4} + GVW_{i-3} + GVW_{i-2} + GVW_{i-1} + GVW_i)$$

Where $\overline{GVW_i}$ is the weekly average GVW for the 7 days ending with day i.

The results of using this method are shown in Figure 2.8.

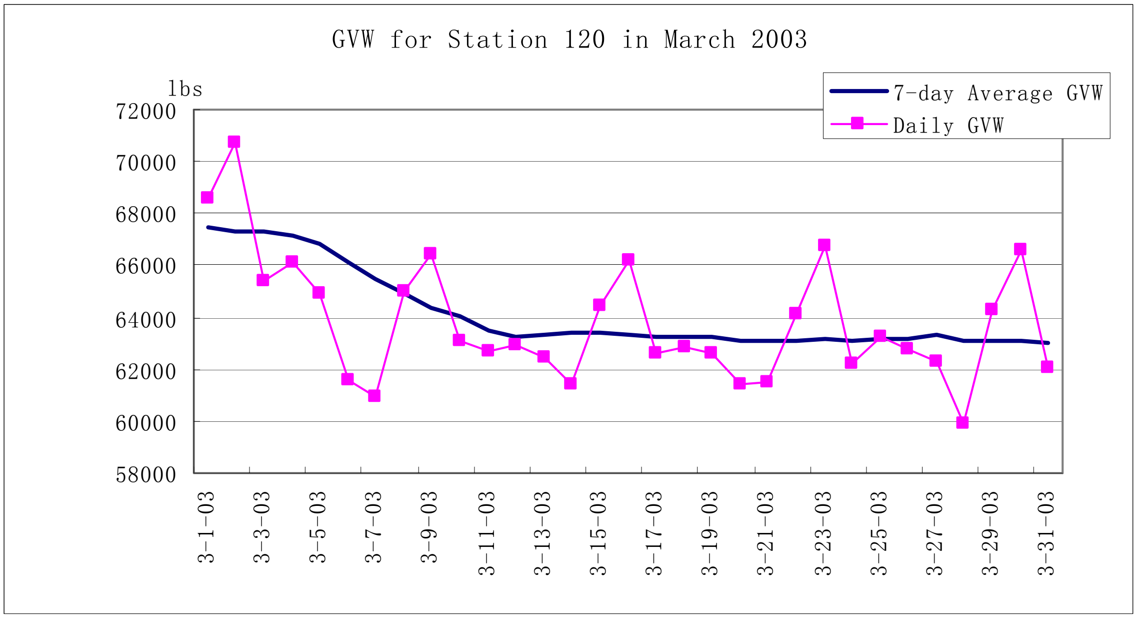


Figure 2.8 AGVW for Station 120 in March 2003

In this way, we can observe that the values decrease from the 6600-6800 range to the 62000-64000 range during March 2003.

CHAPTER 3. DAILY CHECKING FOR WIM DATA

3.1. Introduction to the WDDC Program

In Chapter 2, we have discussed the WIM data checking on a monthly basis. When we analyze the data at the end of a month, it is possible that a data quality problem has existed for several weeks before we find it. To improve the data quality, it is desirable to detect the problem soon after it happens, and fix the problem as soon as possible. Daily checking for WIM data can provide a way to achieve this objective. Instead of waiting until the current month ends, we can check the data from yesterday or the day before yesterday. If the daily checking procedure is to be repeated each day, it should involve as little human effort as possible, and preferably be an automated program.

The Weigh-In-Motion Daily Data Checking (WDDC) program was specifically developed in this project to facilitate the daily checking process for INDOT. This program should be run on a PC with Microsoft Excel and IRD Office software installed. Microsoft Excel version 2002 and IRD Office version 7.5.0 were used during the program development.

Table 3.1 listed the functional files in the WDDC program.

Table 3.1 Files in the WDDC Program

File	Location	Function
“Dailychecking.bat”	C:\Ird\Ird\Dataanalysis\750*	Generate batch reports
“Dailychecking.xls”	C:\Temp\Ird	Import the summary reports
Report parameter files	C:\Ird\Ird\Dataanalysis\750\Reports*	Specify the report formats
Site parameter files	C:\Ird\Ird\Dataanalysis\750\Params*	Provide the site configurations

*Actual location may vary, depending on the location where IRD office is installed.

3.2. Procedure for Using the WDDC Program

The procedure for using the WDDC program is described in Figure 3.1.

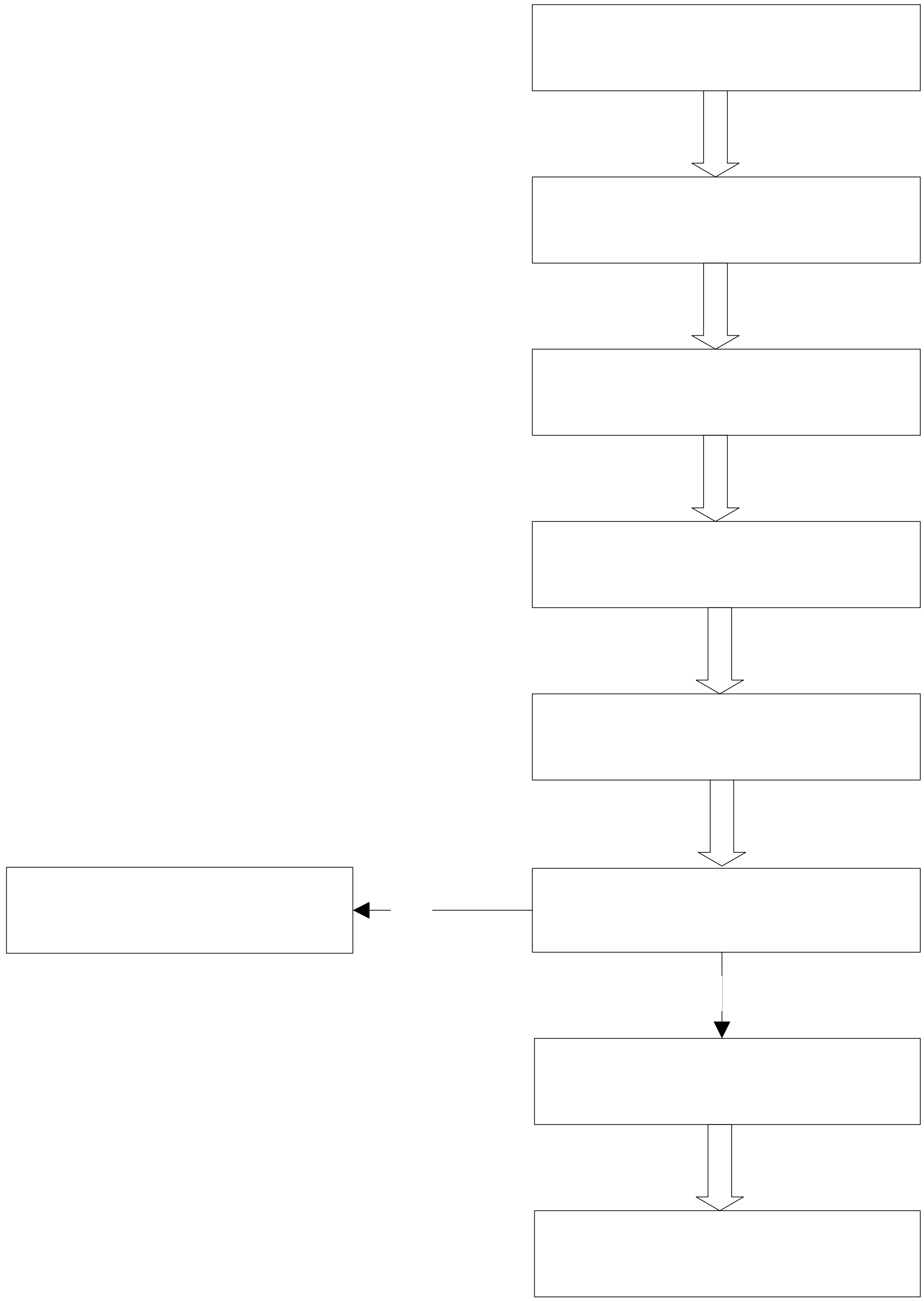


Figure 3.1 Step by Step Daily Checking Procedure

Step 1: The folder “newdata” is for holding IRD-encrypted raw data files, and should be placed in the same folder as “office.exe” (current location: C:\IRD\IRD\DATAANALYSIS\750). If it is the first time to use the WDDC program, please create this folder. Before proceeding to Step 2, please delete all data files in this folder.

Step 2: Copy the data files for the day of interest for each site into the folder "newdata". Let us say that we are interested in WIM data for the day November 20, 2002. All the files in the folder "newdata" should have a name like "11202002.*" (where * is called the site extension). Actually, the site extensions are the WIM station number without the last zero. For example, 11202002.240 is the raw data file from WIM Station 2400. The program currently can handle 45 sites. Table 3.2 shows the list of these sites.

Table 3.2 Site Extensions Supported in WDDC Program

Stations	Lane #1	Lane #2	Lane #3	Lane #4	Lane #5	Lane #6
100	Y	N	*	*	*	*
110	Y	N	Y	N	*	*
120	Y	Y	Y	Y	*	*
130	Y	N	Y	N	*	*
200	Y	Y	Y	Y	*	*
210	Y	Y	Y	Y	*	*
220	Y	N	N	N	*	*
230	Y	N	N	N	*	*
240	Y	Y	*	*	*	*
300	Y	N	Y	N	*	*
310	N	N	Y	N	*	*
320	Y	Y	Y	Y	*	*
330	Y	Y	Y	Y	Y	Y
340	Y	Y	Y	Y	Y	Y
350	Y	Y	Y	*	*	*
351	Y	Y	Y	*	*	*
352	Y	Y	Y	*	*	*
353	Y	Y	Y	*	*	*
360	Y	Y	Y	Y	*	*
370	Y	Y	Y	Y	*	*
400	Y	Y	Y	Y	*	*
401	Y	Y	Y	Y	*	*
410	Y	N	Y	N	*	*
420	Y	Y	Y	*	*	*
421	Y	Y	Y	*	*	*
430	Y	Y	Y	Y	Y	Y
440	Y	Y	Y	Y	Y	Y
450	N	N	Y	N	*	*
460	Y	N	N	N	*	*
470	Y	Y	Y	Y	*	*
500	Y	N	N	N	*	*
510	Y	Y	Y	Y	*	*
520	Y	Y	*	*	*	*
530	Y	N	Y	N	*	*
540	Y	N	Y	N	*	*
600	Y	Y	N	N	*	*
610	Y	N	Y	N	*	*
620	Y	N	*	*	*	*
630	Y	N	N	N	*	*
640	Y	Y	Y	Y	*	*
650	Y	N	Y	N	*	*
660	N	N	Y	N	*	*
730	Y	Y	Y	Y	*	*
732	Y	Y	Y	Y	*	*
734	Y	Y	Y	Y	*	*

Y--Lane Supported (WIM Scale Installed) N--Lane Not Supported (Classification Only)

Step 3. “Dailychecking.bat” is the batch file that will call “office.exe” and create batch reports. This file has been created for this program and should be placed in the same folder as “office.exe” (current location: C:\IRD\IRD\DATAANALYSIS\750). As long as we put only one day’s data in the folder “newdata”, there is no need to change the batch file. In order to get output files, there must be corresponding site parameter files located in: C:\IRD\IRD\DATAANALYSIS\750\PARAMS for each site. Otherwise, there will be no output files for this site. The output files generated are named as “Output1.*” and “Output2.*” (* is the site extension), and they will be located in: C:\Temp\IRD. All report parameters have been set in “WDDC1A(B,C,D).rpt” and “WDDC2A(B,C,D).rpt”, which should be placed in: C:\IRD\IRD\DATAANALYSIS\750\REPORTS.

Step 4. “Dailychecking.xls” is an EXCEL file that has pre-imported the output files “Output1.*” and “Output2.*” (* is the site extension) in a fixed format and provides some summarized data for checking. Each time “Dailychecking.xls” is opened, the pre-imported data will be automatically updated with the new output files. Currently there are 50 worksheets in this EXCEL file. The “Summary” worksheet provides all the important variables for daily checking. The “Hourly Counts” worksheet records the class 9 vehicle counts for each hour of the day. The pre-imported output files are stored in the worksheets named as “Site*” (* is the site extension). For example, the output file for Site 240 is stored in the “Site240” worksheet. “Plot1” and “Plot2” worksheets compare the Average Gross Vehicle Weight (GVW) for Class 9 Vehicle for the day of interest with the historical data, which consist of 6 months’ data in 2002 (May, June, September, October, November and December). These are the only data available to us during the program development. All the historical data are stored in the “Historical” worksheet.

Step 5. The “Summary” worksheet has four columns with summarized data. Column A is the site extension. Column F is the start time linked from the imported output file. Column B contains the unclassified vehicle rates and Column D shows the Avg GVW for Class 9 Vehicles during the specified period. We should check first whether Column F in the “Summary” worksheet is the date we are interested in. For example, let’s say in Step 2 we have copied raw data for November 20, 2002 into the folder “newdata”. When we proceed to Step 5, we find all the sites show “Nov 20, 2002” at Column F except that Site 410 shows “Nov 19, 2002”, which means the “Dailychecking.bat” doesn’t generate output file for Site 410 for November 20, 2002. Generally, this is because Site 410 doesn’t have the raw data for November 20, 2002 in the folder “newdata”. In this way, we can be alerted to sites that have missing data for the day of interest.

Step 6. Columns B and D in the “Summary” worksheet use two easy criteria to check the reliability of the data. Criterion 1 is to check if the number of unclassified vehicles exceeds 10%. If the unclassified vehicle rates at a site are over 10% for several consecutive days, the WIM system at that site may need to be checked out. Criterion 2 is to check if the Average GVW for Class 9 Vehicle is out of the 25,000-80,000 pounds range. As we can see in Figure 3.2, all the data exceeding the limits will be highlighted.

	Column B	Column D	Column F
Stations	Unclassified Vehicle Rate	Class 9 Avg GVW	From:
100#1	56.0%	0	Wed Nov 20 00:00:00 2002
110#1	8.0%	47163	Wed Nov 20 00:00:00 2002
120#1	11.7%	75260	Wed Nov 20 00:00:00 2002
130#1	32.3%	39502	Wed Nov 20 00:00:00 2002
200#1	100.0%	0	Wed Nov 20 00:00:00 2002

Figure 3.2 The “Summary” Worksheet （Part）

Step 7. It is possible that a WIM station may work improperly only a few hours in a day. The “Hourly Counts” worksheet provides a way we can check the hourly counts for class 9 vehicles. As we can see in Figure 3.3, all the hourly counts with zero will be highlighted.

Class 9 Vehicle Count					
	Site				
Hours	100#1	110#1	120#1	130#1	200#1
0-1	0	198	105	84	0
1-2	0	221	97	86	0
2-3	0	192	77	54	0
3-4	0	203	60	79	0
4-5	0	188	71	75	0
5-6	0	175	85	98	0
6-7	0	178	87	87	0
7-8	0	177	111	80	0
8-9	0	218	131	104	0
9-10	0	286	130	95	0
10-11	0	270	147	127	0
11-12	0	253	138	123	0
12-13	0	293	156	110	0
13-14	0	291	145	119	0
14-15	0	248	153	118	0
15-16	0	232	144	146	0
16-17	0	283	146	115	0
17-18	0	292	158	131	0
18-19	0	265	160	124	0
19-20	0	249	168	116	0
20-21	0	246	148	120	0
21-22	0	227	151	102	0
22-23	0	226	146	118	0
23-24	0	233	110	107	0
Total	0	5644	3024	2518	0

Figure 3.3 The “Hourly Counts” Worksheet (Part)

There are some situations in which one day’s data may not be enough to judge the performance of a site’s WIM system. The comparison of one day’s value to historical values can substantially help the user to make the judgment. In the WDDC program, “Dailychecking.xls” provides several ways to compare the GVW for Class 9 Vehicles for the day of interest to the historical values.

In the “Plot1” worksheet, each data point represents one site, with the x coordinate as a site’s historical average and the y coordinate as the site’s value for the day of interest. There are a total of 45 data points in Figure 3.4 (Lane #1 data only). If we move the cursor to any data point, we can see the name for the series, which is also the site extension for the data. If a data point is above the upper dashed line (blue on the screen), it means the ratio of the GVW for the day of interest to the historical average is more than 1.2. If a point is below the lower dashed line (red on the screen), it means the ratio is less than 0.8. The data points between the two lines have a ratio between 0.8 and 1.2.

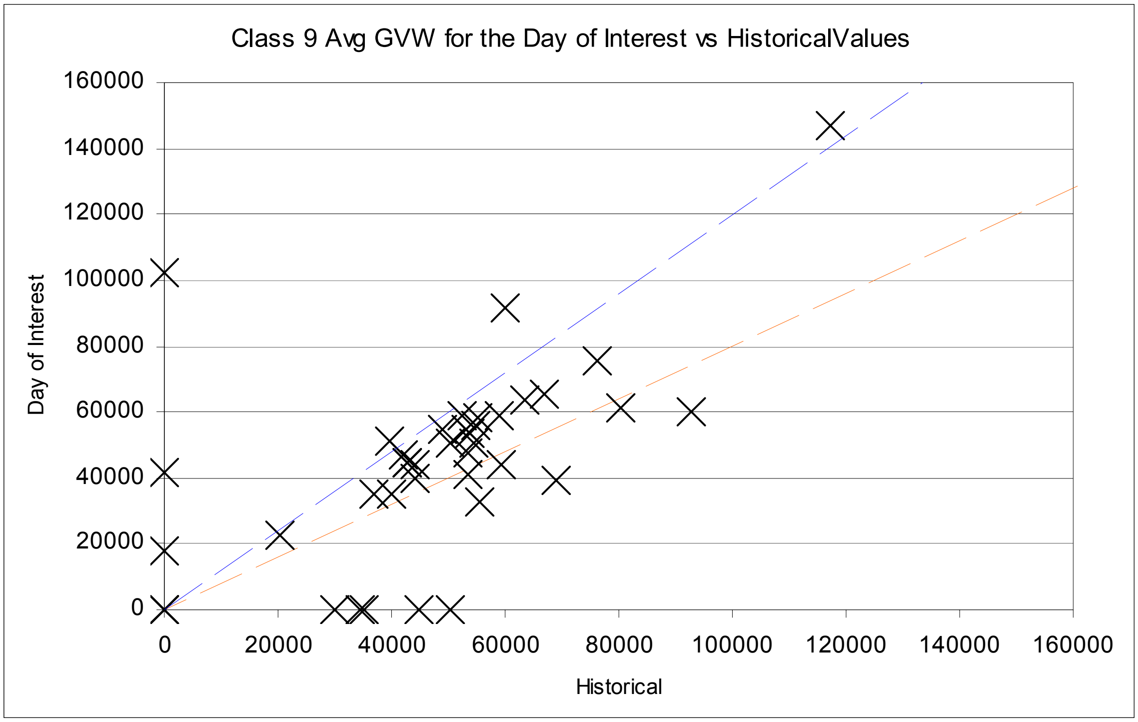


Figure 3.4 The “Plot1” Worksheet

In the “Plot2” worksheet, the confidence intervals (α=0.05) have been added to the historical average for each site, with the x coordinate as the site extension and y coordinate as GVW. We can then identify the data points that fall within or outside the confidence interval site by site. In Figure 3.5, we can see for the day of November 20, 2002, the average GVW for lane 1 in site 220 is near the lower limit of the confidence interval, while the average GVW for lane 1 in site 130 is far below the lower limit of the interval.

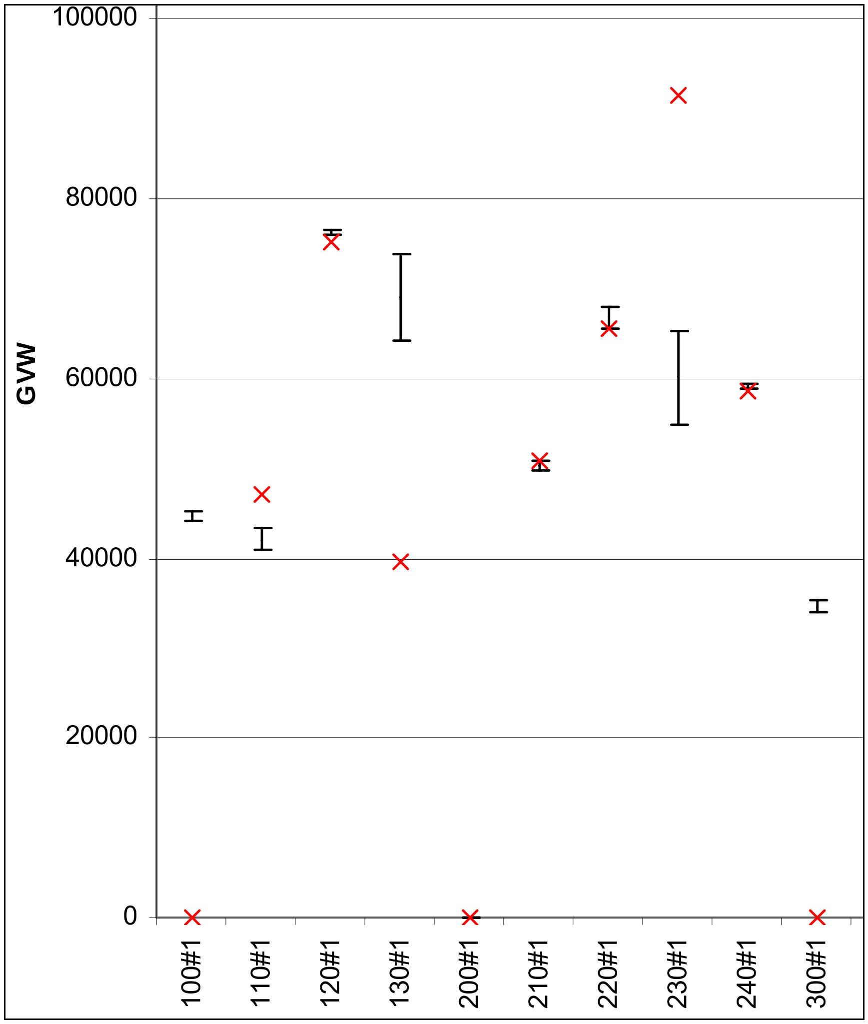


Figure 3.5 The “Plot2” Worksheet (Part)

If we need to check more information for the day of interest, we can go to the output files imported in the spreadsheet or to the original ASCII output files located at: C:\TEMP\IRD.

CHAPTER 4. IMPUTATION FOR WIM DATA

4.1. Pattern Analysis

In this chapter we will begin to focus on ESALs (Equivalent Single Axle Loads). The ESAL data may have larger variance than GVWs, because ESAL is calculated based on the damage of the axle weights to the pavement. However, we will use ESALs because ESAL values are used widely in highway maintenance and pavement design.

Before we go to data imputation methods, it is desirable to analyze the WIM data to identify any patterns in the data. Pattern analysis is a tool to characterize the variability in WIM data, and provides a basis for data imputation. In order to produce a reasonable result, the data used in this analysis should have good quality throughout a year. After inspecting the WIM data for stations 4260 in the years 4270 from 1997 to 2001, we have select the data from July 1, 1999 to June 30, 2000 for station 4270 (lane #1) as the basis for our pattern analysis (See Appendix B).

4.1.1. Day of the Week Pattern

Generally the daily ESALs on weekdays are higher, because there is more truck activity on highways on the weekdays than on weekends and holidays. The recreational traffic that occurs during weekends and holidays has very little effect on the total ESALs. The daily ESAL data are divided into 8 day groups. Groups 1-7 correspond to Monday, Tuesday, …, Sunday. Group 8 corresponds to Holidays. In this analysis, Group 8 has a total of 14 days, as listed in Table 4.1

Table 4.1 Holiday Periods from July 1 1999 to June 30 2002

Holiday	Holiday Period
Independence Day	July 4, 5
Labor Day	September 6
Thanksgiving Day	November 25-28
Christmas Day	December 24-26
New Year's Day	December 31, January 1-2
Memorial Day	May 29

The ANOVA model has only one main factor: day (day of the week factor). The model is analyzed using SAS GLM procedure:

```
Proc glm data=a1;
class day;
model TotalEsal=day;
means day/duncan;
run;
```

The output is shown in Figure 4.1:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	2366971.047	338138.721	234.81	<.0001
Error	358	515547.405	1440.077		
Corrected Total	365	2882518.451			

R-Square	Coeff Var	Root MSE	TotalEsal Mean
0.821147	21.37293	37.94834	177.5533

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Day	7	2366971.047	338138.721	234.81	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Day	7	2366971.047	338138.721	234.81	<.0001

Figure 4.1 Output of ANOVA for Day of the Week Analysis

In Figure 4.1 we can see the factor “day” is significant (=0.05). To further analyze the day factor, we need to look at the output of the Duncan Test in Figure 4.2.

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	Day
A	239.329	52	3
A			
A	236.821	52	4
A			
A	226.514	50	5
A			
A	226.127	52	2
A			
A	222.222	49	1
B	69.004	49	6
C	46.957	14	8
C			
C	46.102	48	7

Figure 4.2 Duncan Test on Day Factor

As we can see in Figure 4.2, Wednesday has the highest ESALs, while Sunday has the lowest values. For simplicity, group 8 will be included in group 7 from now on.

The day of the week factor can be defined as:

$$d_i = \frac{MD_i \times 7}{\sum_{k=1}^{7} MD_k}$$

Where d_i is the day of the week factor

MD_i is the mean value of the day group i

Table 4.2 and Figure 4.3 show the values of the day of the week factors, based on the mean values in Figure 4.2.

Table 4.2 Day of the Week Factor

Day	1	2	3	4	5	6	7
d_i	1.229	1.250	1.323	1.309	1.252	0.382	0.255

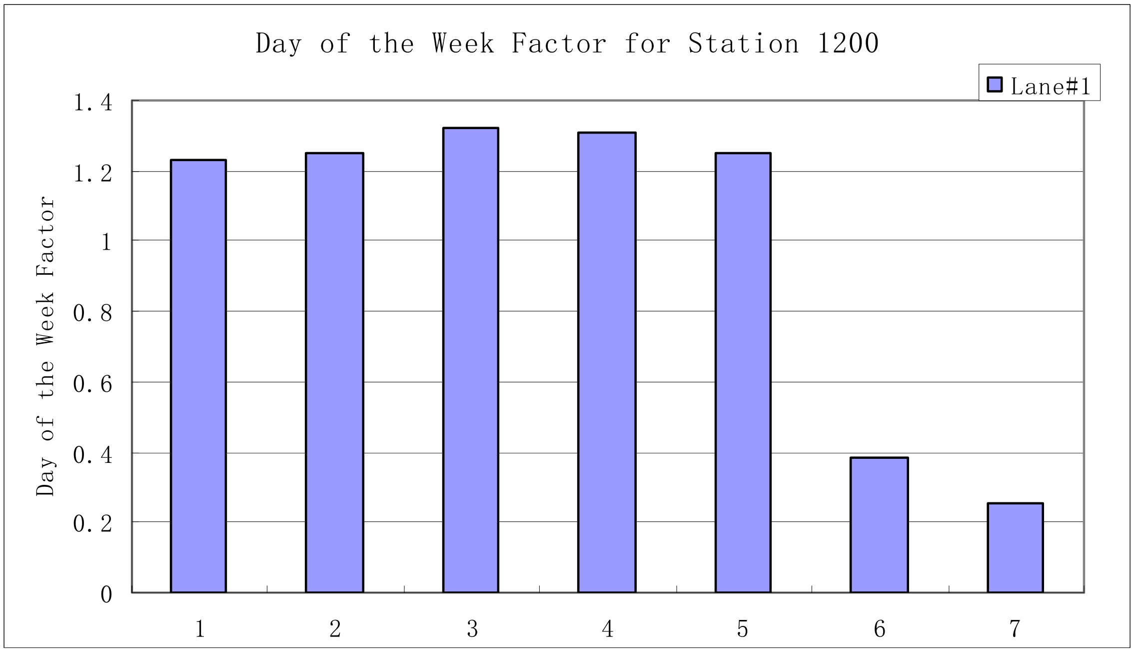


Figure 4.3 Day of the Week Factor

4.1.2. Month of the Year Pattern

Because weather can have significant effect on truck activity, we expect that the daily ESALs reflect the yearly weather pattern. Especially in winter, the daily ESALs should be lower, because of the severe weather. The daily ESAL data are divided into 12 monthly groups. These groups correspond to each calendar month. The ANOVA model now has two main factors: day (day of the week factor) and month (month of the year factor). The model is analyzed using SAS GLM procedure:

```
Proc glm data=a1;
class day month;
model TotalEsal=day month day*month;
means month/duncan;
run;
```

The result of ANOVA by SAS is shown in Figure 4.4.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	89	2561886.869	28785.246	24.78	<.0001
Error	276	320631.582	1161.709		
Corrected Total	365	2882518.451			

R-Square	Coeff Var	Root MSE	TotalEsal Mean
0.888767	19.19641	34.08385	177.5533

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Day	7	2366971.047	338138.721	291.07	<.0001
Month	11	88188.860	8017.169	6.90	<.0001
Day*Month	71	106726.962	1503.197	1.29	0.0752

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Day	7	2223059.938	317579.991	273.37	<.0001
Month	11	82924.166	7538.561	6.49	<.0001
Day*Month	71	106726.962	1503.197	1.29	0.0752

Figure 4.4 Output of ANOVA for Month of the Year Analysis

In Figure 4.4 we can see the factors “day” and “month” are significant, while the interaction factor is not significant (α=0.05). To further analyze the month factor, we need to look at the output of the Duncan Test in Figure 4.5.

Means with the same letter are not significantly different.

Duncan Grouping		Mean	N	Month
	A	217.470	30	6
	B	197.035	31	7
	B			
	B	195.061	31	3
	B			
	B	190.671	31	5
	B			
C	B	187.821	29	2
C				
C	D	172.477	30	9
	D			
	D	166.297	30	4
	D			
	D	163.774	31	1
	D			
	D	161.432	31	12
	D			
	D	161.139	31	10
	D			
	D	159.387	30	11
	D			
	D	158.913	31	8

Figure 4.5 Duncan Test on Month Factor

As we can see in Figure 4.5, June has the highest average ESAL values while August has the lowest ESAL values.

The month of the year factor can be defined as:

$$m_i = \frac{MM_i}{MM_7}$$

m_i is the month the year factor

MM_i is the mean values of the month group i

Table 4.3 and Figure 4.6 show the values of the month of the year factors, based on the mean values in Figure 4.5.

Table 4.3 Month of the Year Factor

Month	7	8	9	10	11	12	1	2	3	4	5	6
m_i	1.000	0.807	0.875	0.818	0.809	0.819	0.831	0.953	0.990	0.844	0.968	1.104

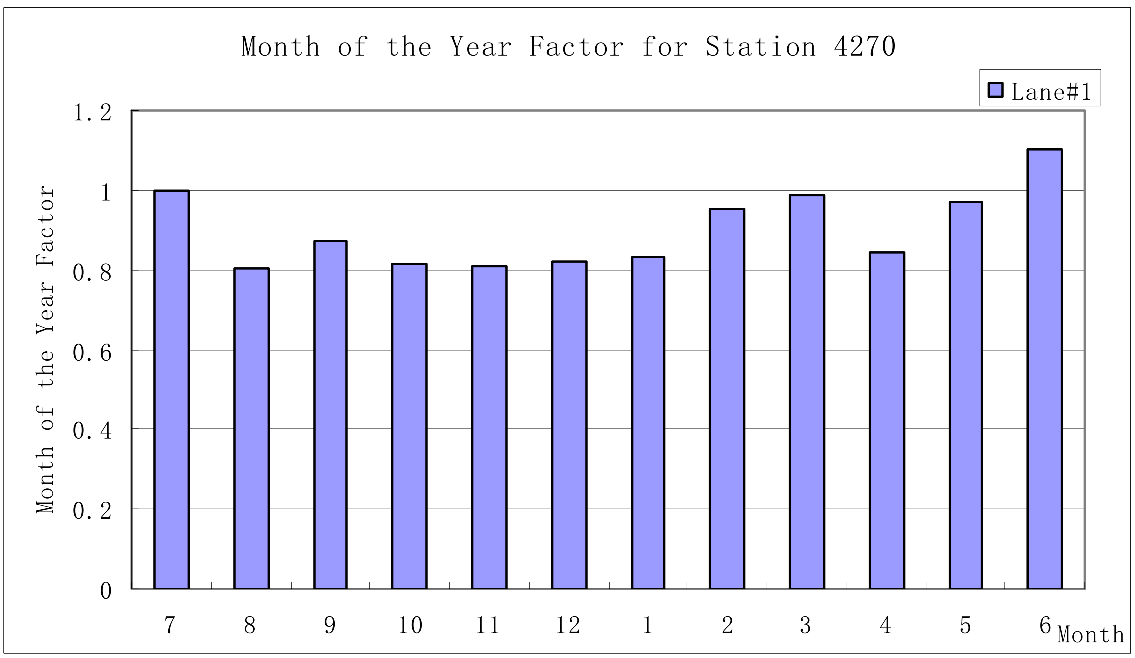


Figure 4.6 Month of the Year Factor

4.2. Imputation Methods

In this section, we will discuss several data imputation methods. To compare the candidate methods, we have the WIM data from August 1, 2000 to October 31, 2000 available (See Appendix C). First, we assume that the data from

September 24, 2000 to September 30, 2000 are missing. Then we will impute the 7 days of “missing” ESAL values and compare the imputed values to their original values.

4.2.1. Factor Method

Based on pattern analysis, we can impute the data using the day of week factor and the month of the year factor. For any day, which is the ith day of a week and is in the jth month of a year, the ESAL value for this day can be calculated as:

$$ESAL_{ij} = MM_j \times d_i$$

d_i is the day of the week factor.

MM_j is the average ESAL value for the month j, and can be calculated as:

$$MM_j = m_j \times \frac{MM_{j-1} + MM_{j+1}}{m_{j-1} + m_{j+1}}$$

If MM_{j+1} is not available, the equation can be simplified as:

$$MM_j = m_j \times \frac{MM_{j-1}}{m_{j-1}}$$

m_i is the month of the year factor.

In this case: $ESAL_{i9} = m_9 \times \frac{MM_8 + MM_{10}}{m_8 + m_{10}} \times d_i$

Here j=9, and $MM_8 = 202.26$, $MM_{10} = 197.45$, d_i and m_i are in Table 4.2 and 4.3.

The ESAL values are estimated in Table 4.4.

Table 4.4 ESAL Values Estimated by Factor Method

Date	Sep. 24	Sep. 25	Sep. 26	Sep. 27	Sep. 28	Sep. 29	Sep. 30
i	7	1	2	3	4	5	6
ESAL(Estimated)	54.86	264.45	269.1	284.81	281.83	269.56	82.12
ESAL(Actual)	40.8	182.4	243.1	303.9	274.5	214.4	94.1

There are several performance measures that could be used to evaluate the quality of imputed data. In this report we will use RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percent Error).

$$RMSE = \sqrt{\frac{\sum e_i^2}{n}}$$

$$MAPE = \frac{\sum |PE_i|}{n}$$

Here, $PE_i = \frac{e_i}{V_i} \times 100\%$, $e_i = \hat{V}_i - V_i$, V_i is the measured variable.

For the factor method, RMSE=40.0, MAPE=19.7 percent.

4.2.2. Ordinary Regression

As we saw before, the day of the week factor has a very important effect on ESALs. We can use this factor as the only regressor. The basic model is:

$$ESAL_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \varepsilon_i$$

ε_i are independent $N(0, \sigma^2)$

For any day $\overline{X}_i$, which is the jth day in a week,

$$X_{ik} = \begin{Bmatrix} 1, k = j \\ 0, k \neq j \end{Bmatrix}, k = 1,2,3,4,5,6$$

This model is analyzed using SAS REG procedure:

```
proc reg data=a1;
model TotalEsal = d1 d2 d3 d4 d5 d6;
output out=p;
run;
```

The output is shown in Figure 4.7.

The REG Procedure
Model: MODEL1
Dependent Variable: TotalEsal

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	643927	107321	78.65	<.0001
Error	78	106436	1364.55826		
Corrected Total	84	750362			

Root MSE	36.93993	R-Square	0.8582
Dependent Mean	196.86118	Adj R-Sq	0.8472
Coeff Var	18.76446		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	48.90833	10.66364	4.59	<.0001
d1	1	178.69167	15.08066	11.85	<.0001
d2	1	219.72244	14.78781	14.86	<.0001
d3	1	213.75000	15.08066	14.17	<.0001
d4	1	204.78333	15.08066	13.58	<.0001
d5	1	190.71667	15.08066	12.65	<.0001
d6	1	22.02500	15.08066	1.46	0.1482

Figure 4.7 Output of the REG Analysis

Based on the parameters in Figure 4.7, the ESAL values can be estimated as shown in Table 4.5.

Table 4.5 ESAL Values Estimated by Ordinary Regression

Date	Sep. 24	Sep. 25	Sep. 26	Sep. 27	Sep. 28	Sep. 29	Sep. 30
i	7	1	2	3	4	5	6
ESAL(Estimated)	48.91	227.6	268.63	262.66	253.69	239.63	70.93
ESAL(Actual)	40.8	182.4	243.1	303.9	274.5	214.4	94.1

For Ordinary Regression, RMSE=29.4, MAPE=16.1 percent.

4.2.3. Regression with Autocorrelated Errors

When a regression is performed on the WIM data, which are also time series data, the errors may not be independent. There is big chance that the errors are correlated, which means each error may be correlated with the error immediately before it. Because one of the assumptions of ordinary regression requires the residuals to be independent, it may be desirable if we can make some correction to the ordinary model. The model used in section 4.2.2 can now be modified to be:

$$ESAL_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \varepsilon_i$$

$$\varepsilon_i = \rho\varepsilon_{i-1} + u_i$$

u_i are independent $N(0,\sigma^2)$

For any day $\overline{X}_i$, which is the jth day in a week,

$$X_{ik} = \begin{Bmatrix} 1, k = j \\ 0, k \neq j \end{Bmatrix}, k = 1,2,3,4,5,6$$

This model is analyzed using SAS AUTOREG procedure:

```
proc autoreg data=a1;
model TotalEsal = d1 d2 d3 d4 d5 d6 / nlag=2 method=ml;
output out=p p=yhat pm=trendhat;
run;
```

The output is shown in Figure 4.8 and Figure 4.9.

The AUTOREG Procedure

Dependent Variable TotalEsal

Ordinary Least Squares Estimates

SSE	106435.544	DFE	78
MSE	1365	Root MSE	36.93993
SBC	878.592816	AIC	861.494257
Regress R-Square	0.8582	Total R-Square	0.8582
Durbin-Watson	1.3703		

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	48.9083	10.6636	4.59	<.0001
d1	1	178.6917	15.0807	11.85	<.0001
d2	1	219.7224	14.7878	14.86	<.0001
d3	1	213.7500	15.0807	14.17	<.0001
d4	1	204.7833	15.0807	13.58	<.0001
d5	1	190.7167	15.0807	12.65	<.0001
d6	1	22.0250	15.0807	1.46	0.1482

Figure 4.8 Output of Ordinary Least Squares Estimates

In Figure, the Ordinary Least Squares Estimates give the same estimated parameters as in session 4.2.2. The D value for the Durbin-Watson test is 1.3703, which indicates the errors are correlated.

```
                         The AUTOREG Procedure

                     Maximum Likelihood Estimates

     SSE                  97323.5598    DFE                       77
     MSE                        1264    Root MSE            35.55197
     SBC                  875.612172    AIC               856.070962
     Regress R-Square         0.8498    Total R-Square        0.8703
     Durbin-Watson            1.9371


                                           Standard                 Approx
  Variable         DF     Estimate         Error    t Value     Pr > |t|

  Intercept         1      49.3683       10.7088       4.61       <.0001
  d1                1     179.8939       12.7138      14.15       <.0001
  d2                1     219.2629       14.2210      15.42       <.0001
  d3                1     212.0674       14.9150      14.22       <.0001
  d4                1     203.9982       14.9342      13.66       <.0001
  d5                1     190.2864       14.5269      13.10       <.0001
  d6                1      21.9993       12.9028       1.70       0.0922
  AR1               1      -0.2965        0.1101      -2.69       0.0087
```

Figure 4.9 Output of Maximum Likelihood Estimates

The results show that AR1 is significant ($\alpha = 0.05$). Based on the parameters given in Figure 4.9, the ESAL values can be estimated as in Table 4.6.

Table 4.6 ESAL Values Estimated by Regression with Autocorrelated Errors

Date	Sep. 24	Sep. 25	Sep. 26	Sep. 27	Sep. 28	Sep. 29	Sep. 30
i	7	1	2	3	4	5	6
ESAL(Estimated)	49.05	229.17	268.60	261.43	253.36	239.65	71.37
ESAL(Actual)	40.8	182.4	243.1	303.9	274.5	214.4	94.1

For Regression with autocorrelated errors, RMSE=30.0, MAPE=16.3 percent.

4.2.4. Other Time Series Forecasting Methods

In some situations, we can use time series forecasting methods to update the missing values. These methods are usually more advanced, and can give a more accurate prediction for time series data. Table 4.7 lists a few forecasting methods in time series analysis. Their prediction errors are calculated based on the data from September 24 to September 30, 2000. The data used to train models are from August 1, 2000 to September 23, 2000. These models are analyzed by SAS.

Table 4.7 Results for Time Series Forecasting Methods

Model	RMSE	MAPE
Seasonal Exponential Smoothing	26.8	15.9%
Winters Method-Additive	26.7	15.3%
ARIMA(2,0,0)(1,0,0)s	28.8	20.5%
Log Seasonal Exponential Smoothing	27.7	15.2%
Log Winters Method-Additive	27.7	15.2%
Log ARIMA(2,0,0)(1,0,0)s	25.2	17.5%

4.2.5. Summary

The factor method is simple and easy to apply. However, it may be difficult to find enough good data to develop the month of the year factor, and there is no guarantee that the factor will not change year-to-year. So errors in imputation are likely to be caused by inappropriate factors.

Compared to the factor method, regression methods use only “local” data to develop the regression model. The results are better. Regression with autocorrelated errors accounts for data that are correlated, and incorporates the correlation into the error terms, which makes the model suitable for time series data.

Although some time series forecasting methods can produce a lower RMSE and MAPE, it is not recommended to use these methods in most situations. Imputation is different from forecasting: imputation can use the data before and after a gap; forecasting uses only the data before a gap.

CHAPTER 5. CONCLUSIONS

WIM data checking should be conducted on both a monthly basis and a daily basis. Chapter 2 discusses the methods that can be used in monthly checking. Although different states use different methods, the three methods using unclassified vehicle rate, front axle distribution and Class 9 vehicle GVW are widely accepted. In Chapter 3, the WDDC (Weigh-In-Motion Daily Data Checking) program is introduced. The WDDC program is developed for INDOT to facilitate the daily checking process. This program can automatically process the raw data to produce daily reports using IRD Office. The reports will then be imported into EXCEL. The whole procedure requires very little human intervention, and provides a convenient way to check daily summary data.

There are debates about data imputation and data integrity. In the experiment in Chapter 4, the factor method and regression methods can do imputation with a MAPE from 15 percent to 20 percent. Due to a lack of data from more WIM sites, we have not been able to test these models extensively and make recommendations. These tests should be carried out as data are made available.

Throughout the data analysis for this project, we realize how much important information the Weigh-In-Motion system can provide. However, the data quality often suffers from equipment problems. In addition, this project has been hampered by the lack of historical data. As more historical data can be retrieved, the ability to impute data can be more comprehensively assessed.

In the meantime, the data checking procedures developed in this project should facilitate the prompt detection of apparent data anomalies and the application of appropriate connective action. In the process, the amount of poor data can be reduced, with a corresponding reduction in the need for data imputation.

LIST OF REFERENCES

American Association of State Highway and Transportation Officials. “Guidelines for Traffic Data Programs”. Washington, D.C.:AASHTO, 1992. As cited in Smith et al. (2003).

American Society for Testing and Materials. “Standard Specification for Highway Weigh-In-Motion (WIM) Systems with User Requirements and Test Method”. ASTM Designation E 1318-94, 1994. As cited in McCall and Vodrazka (1997).

Bergen, A.T., Lindgren, N., Berthelotc, C. and Woytowich, B. “Preserving Highway Infrastructure Using Weigh-In-Motion”. Saskatoon, SK, Canada: IRD, 1998.

Bushman, R. and Pratt A.J. “Weigh In Motion Technology – Economics and Performance”. Charlotte, NC: North American Travel Monitoring Exhibition and Conference, 1998.

Federal Highway Administration. “WIM Scale Calibration: A Vital Activity for LTPP Sites”. McLean, VA: FHWA, 1998.

Green J., Nichols, A., Allen, E., Nuber, L., Thomaz, J., Bullock, D., Boruff, G., Wasson, J. and Newland M. “Virtual Weigh Station: Final Report”. FHWA/IN/JTRP-2001/09, 2002.

Gulen, S., Nagle, J., Weaver, J. and Gallivan, V. “Determination Of Practical ESALs Per Truck Values On Indiana Roads: Final Report”. Indianapolis, IN: INDOT, December 2000.

International Road Dynamics, Inc. “Software User’s Manual IRD Weigh-In-Motion Data Collection System”. Version 7.4.3. Saskatoon, SK, Canada: IRD, 1993, 33 of section 5.

Katz J.K. and Rakha, H.A. “Weigh-In-Motion Evaluation: Final Report”. Blacksburg, VA: Virginia Tech Transportation Institute, 2002.

Lingras, P.J., Sharma, S,C. and Zhong M. “Effect of Missing Value Imputations on Traffic Parameters Estimations from Permanent Traffic Counts”. Washington DC: Transportation Research Board, 82th Annual Meeting Compendium of Papers CD-ROM, 2003.

Little, R.J.A. and Rubin, D.B. “Statistical Analysis with Missing Data”. Hoboken, N.J. : Wiley, 2002.

Lomax, T., Turner, S. and Margiotta, R. "Monitoring Urban Roadways in 2000: Using Archived Operations Data for Reliability and Mobility Measurement". College Station, TX: Texas Transportation Institute, 2001.

McCall, B. and Vodrazka, W.C. "States' Successful Practices Weigh-In-Motion Handbook". Washington, DC: FHWA, 1997.

Nihan, N.L. and Holmesland, K.O. "Use of the Box and Jenkins Time Series Technique in Traffic Forecasting". Transportation, 9, 1980, 125-143. As cited in Zhong et al. (2002).

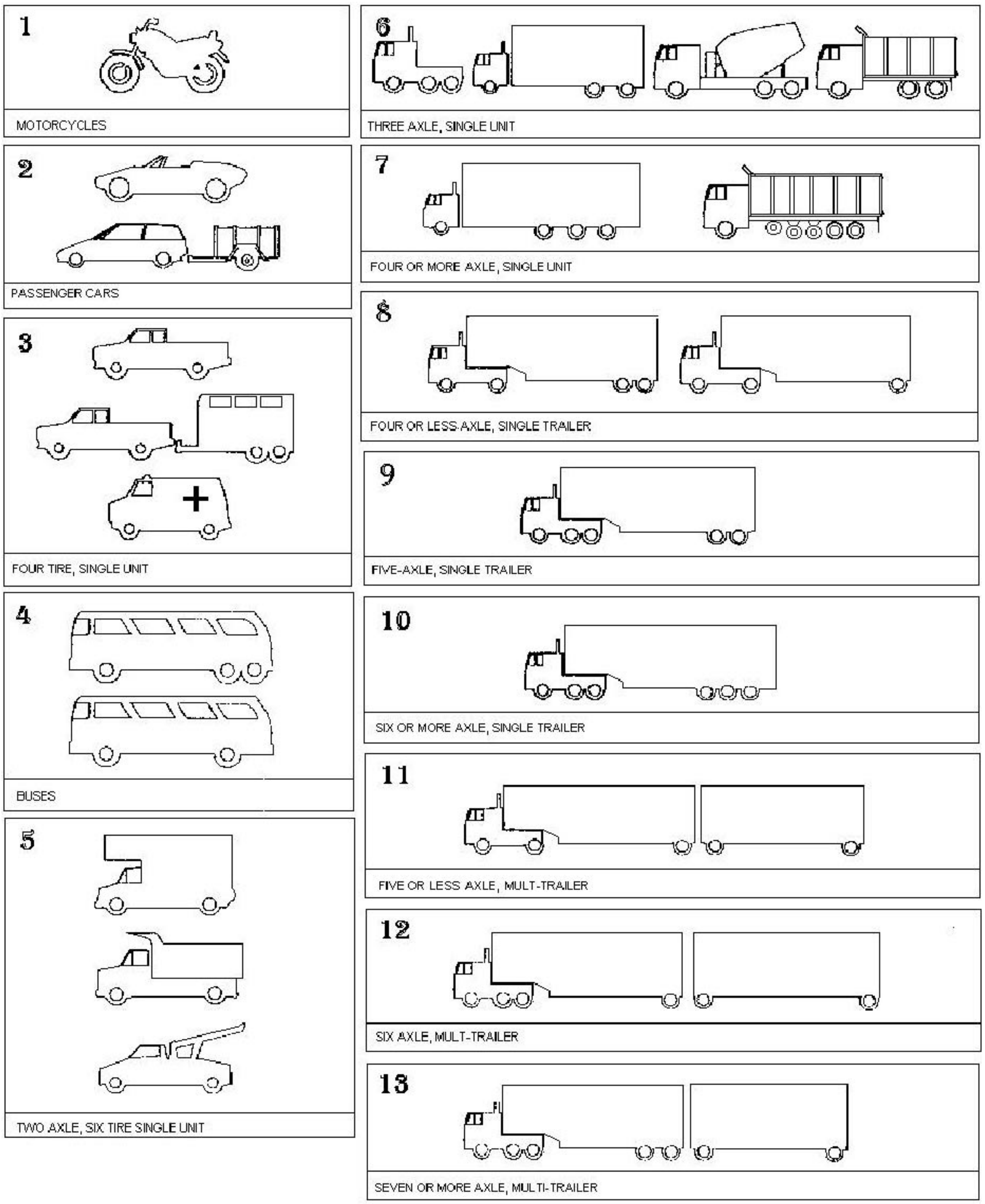
Schmoyer, R. and Hu, P.S. "Analysis of Vehicle Classification and Truck Weight Data of The New England States: Is Data Sharing A Good Idea?" Oak Ridge, TN: Oak Ridge National Laboratory, 1997.

Smith, B.L., Scherer, W.T. and Conklin, J.H. "Exploring Imputation Techniques for Missing Data in Transportation Management Systems". Washington, DC: Transportation Research Board, 82th Annual Meeting Compendium of Papers CD-ROM, 2003.

Zhi, X., Shalaby, A., Middleton, D. and Clayton, A. "Evaluation of Weigh-In-Motion in Manitoba". Canadian Journal of Civil Engineering, 26, 1999, 655-666.

Zhong, M., Lingras, P.J., and Sharma, S.C. "Updating Missing Values of Traffic Counts: Factor Approaches, Time Series Analysis versus Genetically Designed Regression and Neural Network Models". Halifax, NS, Canada: Saint Mary's University, 2002

APPENDIX A. FHWA VEHICLE CLASSIFICATION SCHEME



APPENDIX B. DATA USED FOR PATTERN ANALYSIS

Date	TotalEsal
07-01-99	230.1
07-02-99	243
07-03-99	102.2
07-04-99	68.6
07-05-99	132.5
07-06-99	320
07-07-99	291.3
07-08-99	268.7
07-09-99	314.5
07-10-99	74.9
07-11-99	51.9
07-12-99	276.6
07-13-99	239.1
07-14-99	251.7
07-15-99	252.5
07-16-99	242.6
07-17-99	64.2
07-18-99	44.6
07-19-99	203.7
07-20-99	202.2
07-21-99	259.7
07-22-99	267.4
07-23-99	278.5
07-24-99	73.2
07-25-99	65.4
07-26-99	274.7
07-27-99	199.1
07-28-99	229.8
07-29-99	291.8
07-30-99	228.4
07-31-99	65.2
08-01-99	43.3
08-02-99	213.5
08-03-99	183.1
08-04-99	247.3
08-05-99	205
08-06-99	261.7

08-07-99	78.2
08-08-99	39.7
08-09-99	206.5
08-10-99	255.7
08-11-99	253.7
08-12-99	149.3
08-13-99	225.6
08-14-99	59.8
08-15-99	61.6
08-16-99	216.8
08-17-99	228.6
08-18-99	165.7
08-19-99	158.8
08-20-99	209
08-21-99	54.6
08-22-99	40.4
08-23-99	171.2
08-24-99	153.5
08-25-99	182.3
08-26-99	162
08-27-99	216.3
08-28-99	50.1
08-29-99	42.3
08-30-99	179.5
08-31-99	211.2
09-01-99	218.2
09-02-99	279.4
09-03-99	254.3
09-04-99	64.6
09-05-99	32.5
09-06-99	53.3
09-07-99	245
09-08-99	246.2
09-09-99	247.1
09-10-99	269.5
09-11-99	97.6
09-12-99	72.8
09-13-99	208.4
09-14-99	217.7
09-15-99	210.6
09-16-99	184.1
09-17-99	190.6

09-18-99	70.7
09-19-99	57.6
09-20-99	219.5
09-21-99	193.8
09-22-99	50
09-23-99	243.1
09-24-99	209.3
09-25-99	65.3
09-26-99	54.6
09-27-99	246.2
09-28-99	208
09-29-99	211.3
09-30-99	253
10-01-99	214.5
10-02-99	66.8
10-03-99	28.9
10-04-99	188.5
10-05-99	256.9
10-06-99	190.4
10-07-99	150.3
10-08-99	215.3
10-09-99	41.7
10-10-99	37.1
10-11-99	222.2
10-12-99	217.5
10-13-99	211.1
10-14-99	215.1
10-15-99	199.9
10-16-99	61.6
10-17-99	22.6
10-18-99	205.1
10-19-99	207.2
10-20-99	186.5
10-21-99	228.5
10-22-99	199.3
10-23-99	65.5
10-24-99	40.8
10-25-99	240
10-26-99	180.5
10-27-99	199.4
10-28-99	280.2
10-29-99	264.5

10-30-99	119.3
10-31-99	38.1
11-01-99	245.2
11-02-99	266.5
11-03-99	267.5
11-04-99	257.9
11-05-99	234.7
11-06-99	58.4
11-07-99	32.7
11-08-99	222.9
11-09-99	208.3
11-10-99	260.8
11-11-99	164.3
11-12-99	186.7
11-13-99	62.2
11-14-99	33.6
11-15-99	204.2
11-16-99	183.6
11-17-99	161.6
11-18-99	206.1
11-19-99	191.6
11-20-99	55.7
11-21-99	33.8
11-22-99	237.8
11-23-99	224
11-24-99	164.3
11-25-99	31.9
11-26-99	85
11-27-99	43.7
11-28-99	38.3
11-29-99	193.6
11-30-99	224.7
12-01-99	210
12-02-99	279.5
12-03-99	203
12-04-99	81
12-05-99	52
12-06-99	241.5
12-07-99	232
12-08-99	224
12-09-99	221.4
12-10-99	195

12-11-99	50.3
12-12-99	51.2
12-13-99	181.3
12-14-99	216
12-15-99	281.5
12-16-99	246
12-17-99	201.2
12-18-99	68.9
12-19-99	39.1
12-20-99	245.5
12-21-99	225.2
12-22-99	207.6
12-23-99	159.7
12-24-99	34.8
12-25-99	3.3
12-26-99	18.4
12-27-99	171.2
12-28-99	204.2
12-29-99	263.2
12-30-99	145.1
12-31-99	51.3
01-01-00	15.7
01-02-00	39.3
01-03-00	172.2
01-04-00	187.8
01-05-00	264.5
01-06-00	285.5
01-07-00	223.1
01-08-00	87.4
01-09-00	48.9
01-10-00	272
01-11-00	243
01-12-00	184
01-13-00	214.5
01-14-00	239.5
01-15-00	82.7
01-16-00	54.6
01-17-00	160.9
01-18-00	177.8
01-19-00	215.4
01-20-00	122.5
01-21-00	105.6

01-22-00	45.1
01-23-00	42.6
01-24-00	204.8
01-25-00	247.6
01-26-00	222.6
01-27-00	245.6
01-28-00	260.5
01-29-00	68.2
01-30-00	45
01-31-00	298.1
02-01-00	245.7
02-02-00	293.8
02-03-00	258.7
02-04-00	223.6
02-05-00	91
02-06-00	46.7
02-07-00	255.5
02-08-00	259.8
02-09-00	353.3
02-10-00	199.8
02-11-00	181.9
02-12-00	72.3
02-13-00	38.8
02-14-00	211.7
02-15-00	177.7
02-16-00	267.2
02-17-00	208.4
02-18-00	154.8
02-19-00	54.1
02-20-00	60.4
02-21-00	240.9
02-22-00	235.7
02-23-00	263.2
02-24-00	250.5
02-25-00	227.9
02-26-00	62.8
02-27-00	57.9
02-28-00	245.1
02-29-00	207.6
03-01-00	242.6
03-02-00	221
03-03-00	286.1

03-04-00	72.9
03-05-00	50.4
03-06-00	238.9
03-07-00	321.8
03-08-00	317.9
03-09-00	271.3
03-10-00	227.7
03-11-00	67.2
03-12-00	58.1
03-13-00	240.3
03-14-00	278.4
03-15-00	281.5
03-16-00	202.5
03-17-00	177.2
03-18-00	52.8
03-19-00	45.1
03-20-00	176.8
03-21-00	231.6
03-22-00	225.1
03-23-00	232
03-24-00	187.5
03-25-00	64.5
03-26-00	63.5
03-27-00	228.7
03-28-00	203.7
03-29-00	258.7
03-30-00	268.3
03-31-00	252.8
04-01-00	74.6
04-02-00	27.7
04-03-00	246.2
04-04-00	208.7
04-05-00	242.6
04-06-00	280.7
04-07-00	185.9
04-08-00	58.7
04-09-00	61.7
04-10-00	186.3
04-11-00	219
04-12-00	240.3
04-13-00	231.6
04-14-00	277.7

04-15-00	83.4
04-16-00	53.7
04-17-00	165.4
04-18-00	219.7
04-19-00	188.5
04-20-00	207.3
04-21-00	114.9
04-22-00	37
04-23-00	24.2
04-24-00	193.7
04-25-00	245.1
04-26-00	291.2
04-27-00	271.3
04-28-00	228.5
04-29-00	81.7
04-30-00	41.6
05-01-00	245.7
05-02-00	250.3
05-03-00	255.1
05-04-00	278.7
05-05-00	241.6
05-06-00	96.9
05-07-00	67.2
05-08-00	283.8
05-09-00	255.4
05-10-00	240.1
05-11-00	229.7
05-12-00	276.5
05-13-00	62.7
05-14-00	39.9
05-15-00	273.1
05-16-00	199.5
05-17-00	217.9
05-18-00	238.3
05-19-00	170
05-20-00	42.7
05-21-00	43.2
05-22-00	268.3
05-23-00	252
05-24-00	311.6
05-25-00	291.1
05-26-00	212.4

05-27-00	41.7
05-28-00	15.5
05-29-00	41.3
05-30-00	229
05-31-00	239.6
06-01-00	287.7
06-02-00	244
06-03-00	94
06-04-00	31.9
06-05-00	150.9
06-06-00	273.7
06-07-00	343.9
06-08-00	342
06-09-00	343
06-10-00	85.8
06-11-00	58
06-12-00	214
06-13-00	236.2
06-14-00	242.9
06-15-00	298.2
06-16-00	276.4
06-17-00	77.6
06-18-00	51.6
06-19-00	254.3
06-20-00	210.2
06-21-00	331
06-22-00	323.9
06-23-00	277
06-24-00	71.4
06-25-00	68.1
06-26-00	245.7
06-27-00	238
06-28-00	264.9
06-29-00	307.2
06-30-00	280.6

APPENDIX C. DATA USED FOR IMPUTATION METHODS

Date	TotalESAL
08-01-00	336.3
08-02-00	301
08-03-00	245.3
08-04-00	198.7
08-05-00	63.4
08-06-00	50.8
08-07-00	256.4
08-08-00	289.5
08-09-00	263.9
08-10-00	289.6
08-11-00	220.6
08-12-00	43.6
08-13-00	48.2
08-14-00	251.6
08-15-00	282.8
08-16-00	252.7
08-17-00	211
08-18-00	208.4
08-19-00	58.5
08-20-	51.1

00	
08-21-00	209.1
08-22-00	304.3
08-23-00	241.3
08-24-00	262.6
08-25-00	242.1
08-26-00	68.3
08-27-00	48.1
08-28-00	204.1
08-29-00	259.2
08-30-00	263.7
08-31-00	243.8
09-01-00	225.5
09-02-00	50.9
09-03-00	26.3
09-04-00	42.2
09-05-00	205.7
09-06-00	245.7
09-07-00	284
09-08-00	276.2
09-09-00	88.1
09-10-00	44.8
09-11-	287.4

00	
09-12-00	275.8
09-13-00	333.9
09-14-00	213.6
09-15-00	233.5
09-16-00	71.5
09-17-00	59.9
09-18-00	239.6
09-19-00	276.5
09-20-00	273.4
09-21-00	301.4
09-22-00	216.1
09-23-00	70.3
09-24-00	40.8
09-25-00	182.4
09-26-00	243.1
09-27-00	303.9
09-28-00	274.5
09-29-00	214.4
09-30-00	94.1
10-01-00	62.3
10-02-00	310.5
10-03-	276.5

00	
10-04-00	214.4
10-05-00	191.7
10-06-00	191.6
10-07-00	79.9
10-08-00	46.2
10-09-00	219.7
10-10-00	268.1
10-11-00	234.5
10-12-00	276.1
10-13-00	299
10-14-00	82.5
10-15-00	51.4
10-16-00	220.4
10-17-00	260.3
10-18-00	262.4
10-19-00	281.2
10-20-00	276.2
10-21-00	95.2
10-22-00	42.6
10-23-00	278.2
10-24-00	238.5
10-25-	265

00	
10-26-00	244
10-27-00	287.6
10-28-00	79
10-29-00	55.2
10-30-00	212
10-31-00	218.7