Minnesota Department of Transportation
Office of Transportation Data Forecast and Analysis

# ATR-TDA2: User Manual

Mar 25, 2014, Original Publication
July 23, 2021, revised

By
Dr. Take M. Kwon
**University of Minnesota, Duluth**

# Contents

# 1. Introduction

   This manual describes a software packages called ATR-TDA2, developed by Dr. Taek Kwon at the University of Minnesota Duluth (UMD) for MnDOT uses. This software package is distributed in an MSI formatted file (software installer format used by Windows) for installation on any Windows PCs. For software developments, a version of Microsoft Visual Studio with .Net Framework 4.5 was used.

   One of the important aspects of using the ATR-TDA2 software package is understanding of the file structure used by the software. A specific directory tree with specific directory names must be manually created before using the software. The directory tree and names should exactly follow Figure 1 without change of the letter case neither the spelling. First, the root directory named "traffic" is created in this example. The user can choose any name for this root directory but it must be specified in the Settings parameters window. Inside the root directory, two directories should be created, which are **defines** and **processed**. Inside the defines or processed directory, an ATR directory per each should be created.  If you have already installed Bulldog software packages, traffic, defines, and processed are already created and you can simply add ATR directories in defines and processed.
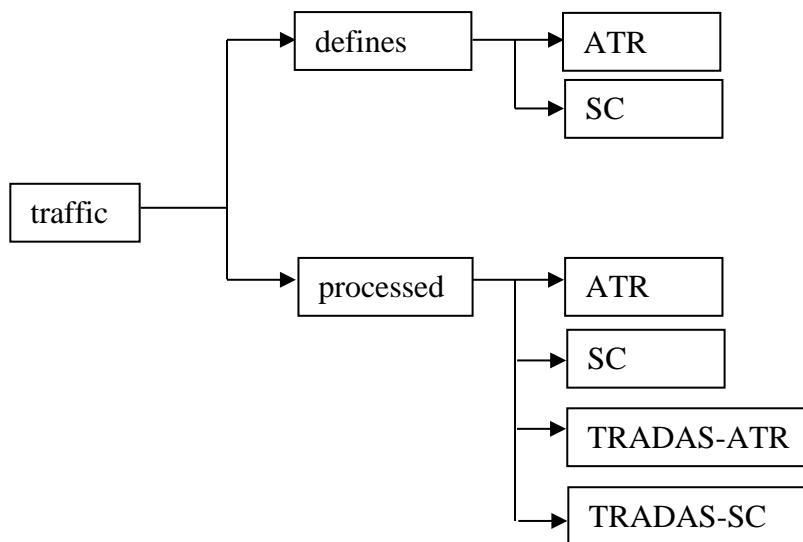
Figure 1: The "traffic" directory structure

After creating the directory tree as shown in Figure 1, the next step is to copy station definition files to the corresponding directories. To run the ATR-TDA2 program, ***the user must first select an ATR station defines file using the Settings/Parameters menu.*** For example, the defines files in the directory may include multiple versions similarly to the following example which shows three:

**traffic\defines\ATR\**
   **ATRDets20071213.txt**
   **ATRDef20080317.txt**
   **Test_ATRDef20080317.txt**

The user can press the "Browse and Select …" button from the Set Required Parameters window (reached by the Settings/Parameters menu) and then select the defines file to be used.

# 2. ATR-TDA2

**2.1 Brief Implementation Description**

Automatic Traffic Recorder (ATR) data is a continuous-count traffic data recorded for every hour, 24/7, all year around, non-stop. The ATR-TDA2 software computes ATR data using the traffic data available from the Regional Traffic Management Center (RTMC). For generating continuous-count data from RTMC data source, the basic algorithm imputes missing data based on spatial and temporal inferences of traffic data relations. The equivalent spatial relation is created by defining primary, secondary, and tertiary detector sets, based on an equivalent traffic flow. Among the three sets of detector data, the program chooses the data set with least missing for computing the ATR data. Temporal relation is created based on the trend that the same day-of-week volume patterns in neighboring weeks are similar. The mathematical description and the detailed algorithm of the imputation process can be found from the final report of the original project [3].

The implementation of the software is described without mathematical descriptions, but focusing on procedural steps. The following list shows the steps of the final ATR data computation from the 30 second detector data available from RTMC.

1. Load the station defines file specified in Settings.
2. Read in the Last Run Ending date from the Settings.
3. Generate an array of consecutive dates consisting of one week, starting from Monday and ending on Sunday since the Last Run Ending date.
4. Check if the traffic data for the computing dates in the array are available. If data for the entire week is not available, give an error message and abort the computation.
5. Read in the 30 second data for primary, secondary, and tertiary detector sets.
6. Compute the missing percentage of each detector set and choose the set with least missing. Spatial imputation.
7. Impute the 30 second detector data for only randomly missing data patterns using the Non-normal Bayesian Linear Regression (NBLR) algorithm, i.e., temporal imputation.
8. Compute the 30 second station data using the imputed detector data.
9. Convert the 30 second station data into 5 minute station data.
10. Impute the 5 minute station data using NBLR algorithm only for the randomly missing patterns.
11. Pack the imputed 5 minute data into hourly data.
12. For hourly missing data, impute them using the historic imputation algorithm if the option for historic imputation was selected in the Settings.

13. Output the data in an ASCII file format defined by the Mn/DOT Office of TDA.

The historic imputation algorithm applied is briefly described. When a large amount of data is missing such as many hours or the whole day, the hourly data is copied from the historic database which is regularly updated whenever a date with no missing data is available. The historic database does not use holidays nor near holidays for the update of the database. Imputation based on the historic database of the detector data is referred to as historic imputation.

Historic data of each hour per day of the week is created using the following relation. Let $d(n)$ be a good data of date n, then the historic data is recursively updated using the following exponential moving average:

$$d(n+1) = d(n)*0.5 + d(n-1)*0.5$$

and the data $d(n+1)$ is kept in the database. This approach assigns a higher weight on the most recent data since $d(n-1)$ was recursively produced using past data. The weight for the averaging is exponentially decreased as the dates get old.

**2.2 How to Use the ATR-TDA2 Software**

**Preparation**

In order to run ATR-TDA2, an ATR station defines file must exist and be selected. If an ATR defines file was not selected or not present, all processing buttons are automatically disabled so that the user cannot run the ATR-TDA2 program. To select an ATR defines file, press the **Settings\Parameters** menu which should bring up the "Parameter Settings" window as shown in Figure 3.

The **Traffic Data Root Directory** is the root node, "**traffic**" directory in Figure 1 and should be entered using the **Browse Folder** button in order avoid spelling mistakes. The textbox below the root directory shows the directory that is accessed by ATR-TDA2, derived from the traffic root directory. The check mark "Perform historic imputation based past good data" should be checked if historic imputation is desired to be used.

The textbox under "ATR Station Definition File" should have the user selected ATR station defines file. It should be set using the "Browse and ..." button provided.

When the ATR-TDA2 produces ATR data from RTMC traffic data, it can simultaneously produce two formats: the legacy input format for a SAS program and the FHWA volume format for TRADAS. The user is responsible to select the **FHWA Export Folder** using the "Browse Folder" button. The ATR-TDA2 will then save all FHWA volume files to that directory.

The setting parameters must then be saved using the **Save** button, otherwise they will be lost.
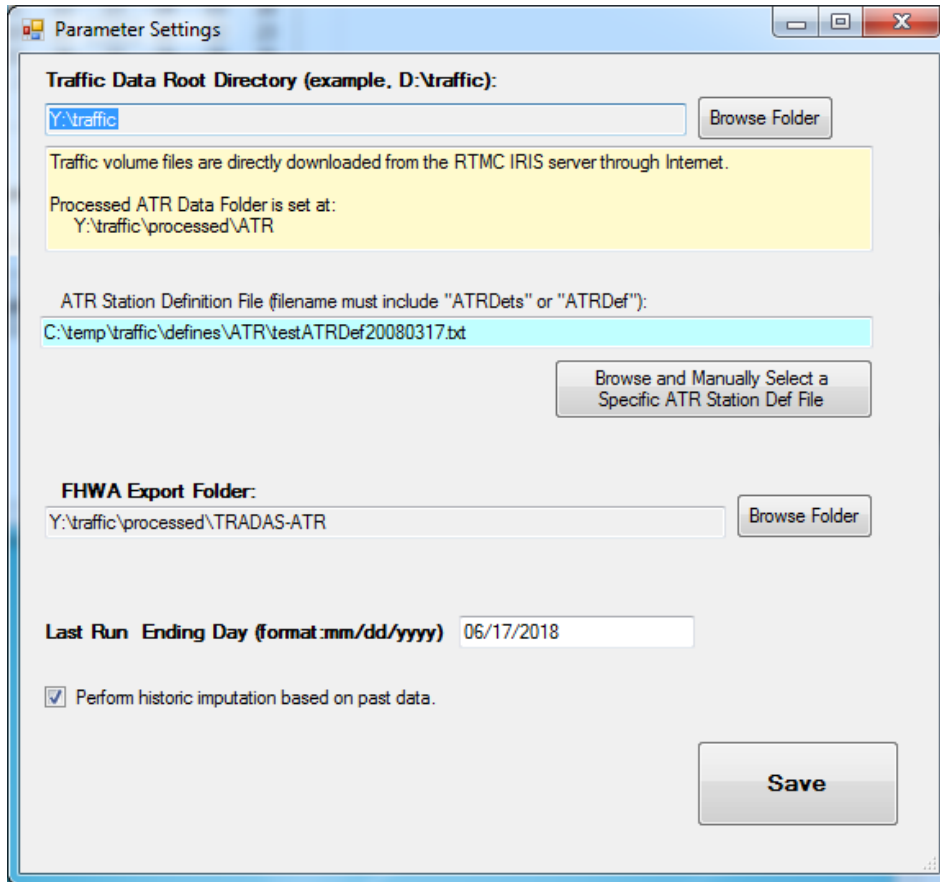
**Figure 2: ATR-TDA Parameter Settings window**

### Runs

Figure 3 shows the main window of the ATR-TDA2 software. There are two options of computing the ATR files, Manual Run or Auto Run. The **Manual Run** is used to run one week data at a time. To use this function, click any day of the week from the calendar. The software will then automatically determine one week (starting from Monday and ending on Sunday) that includes the date the user selected, and will produce the ATR data for the week. The second option is the **Auto Run**. This option computes the ATR files from the week after the Last Ending Date to the most recent week. This **Auto Run** function can be executed using Windows Scheduler. Configuration of this option is described later.

When the check mark "**Generate FHWA VOL file when ATR files are created**" is checked, the ATR-TDA2 produces both the regular ATR formatted data as well as the FHWA volume formatted data, which are then saved in the respective directory. The "**Generate FHWA Vol file from Existing ATR Files**" button is used when a FHWA formatted volume file need to be produced from ATR files. Clicking this button opens a folder dialog in the ATR data directory from which the user can select multiple ATR files for conversion.
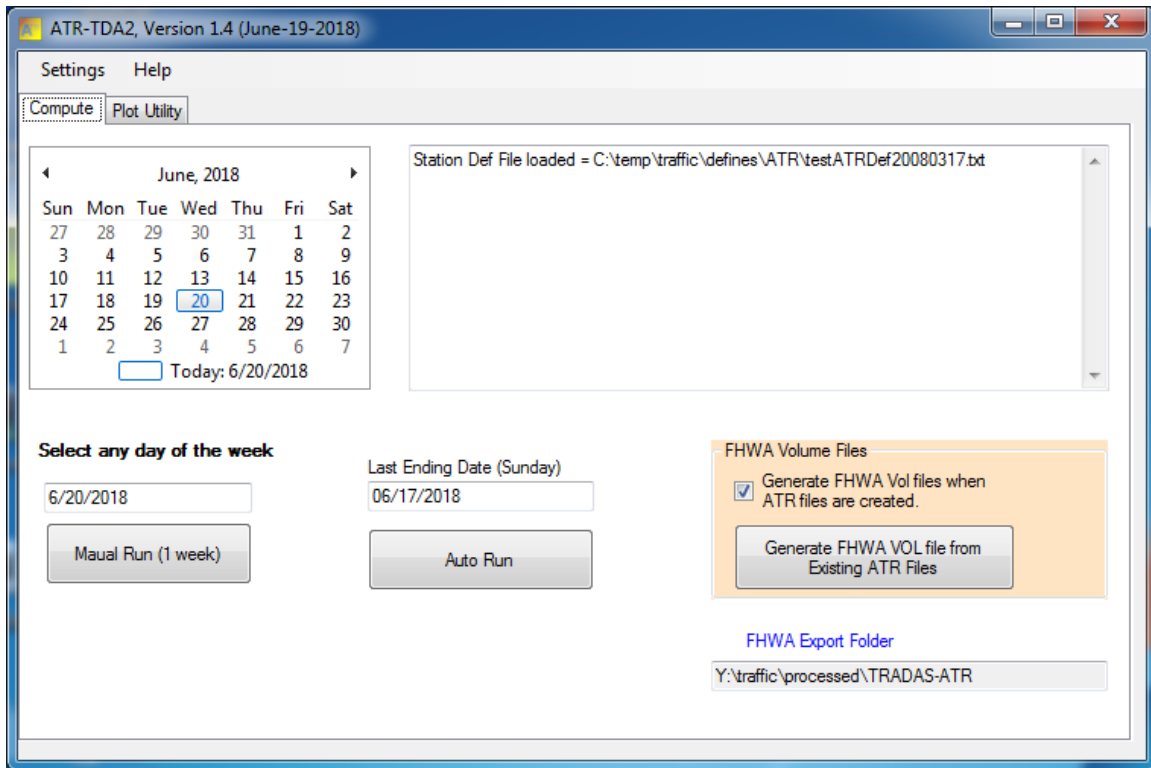
Figure 3: ATR-TDA2 computation tab

## Auto-Run Setting using Windows Task Scheduler

In order to automatically run the ATR-TDA2 program at the scheduled time every day, use the following setup steps for Windows Task Scheduler.

1. Open Windows Task Scheduler by **Control Panel/Scheduled Tasks**.
2. Double click on **Add Scheduled Task.**
3. Click the **Browse** button and browse to select **C:\Program Files\Bulldog\ATR-TDA\ATR-TDA2.exe**.
4. Select the **Every Day** option and set the Start time as desired. (you can also set Once a Week option)
5. Enter User Name and Password
6. Click **Finish**.
7. In the Action, select "Start a program" and select "C:\Program Files\Bulldog\ATR-TDA\ATR-TDA2.exe" for the "Program/Script". In the "Add arguments (optional)" textbox, enter **"-auto"**.

## 2.3 Plot Utility for ATR-TDA2

The second tab named "Plot Utility" (shown in Figure 4) provides visualization of the ATR data generated. To use this utility, browse and load any ATR data file produced by the ATR-TDA2 program. Next, select ATR stations and dates, and then press one of the three buttons available for three different types of outputs. An example of "Hourly Plot" is shown in Figure 5.
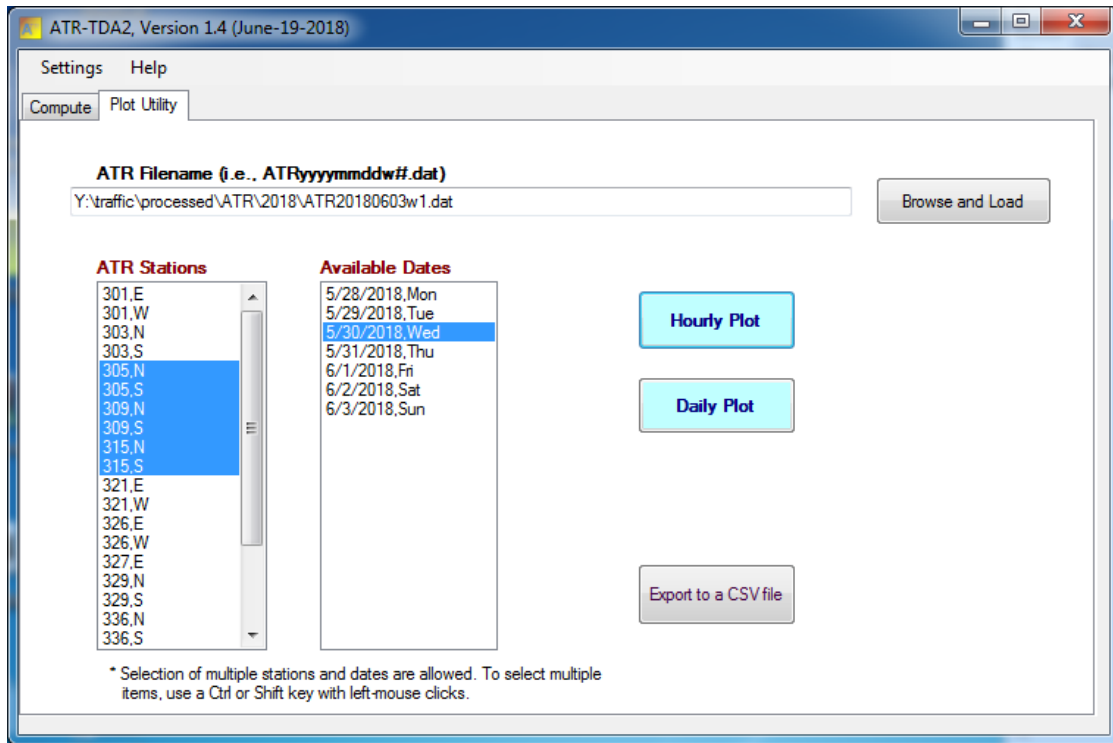


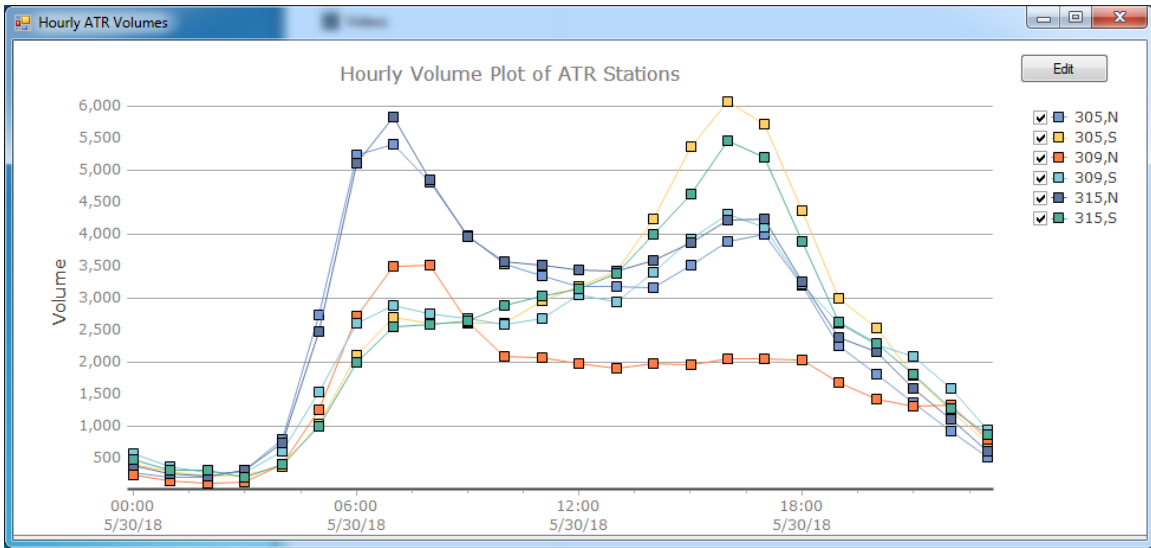**Figure 4: ATR-TDA2 Plot Utility tab**

**Figure 5: Example hourly plot**

The daily plot shows daily volumes of stations selected for one week using a bar graph but it can be changed into a line graph using the **Edit** button. An example is shown in Figure 6.
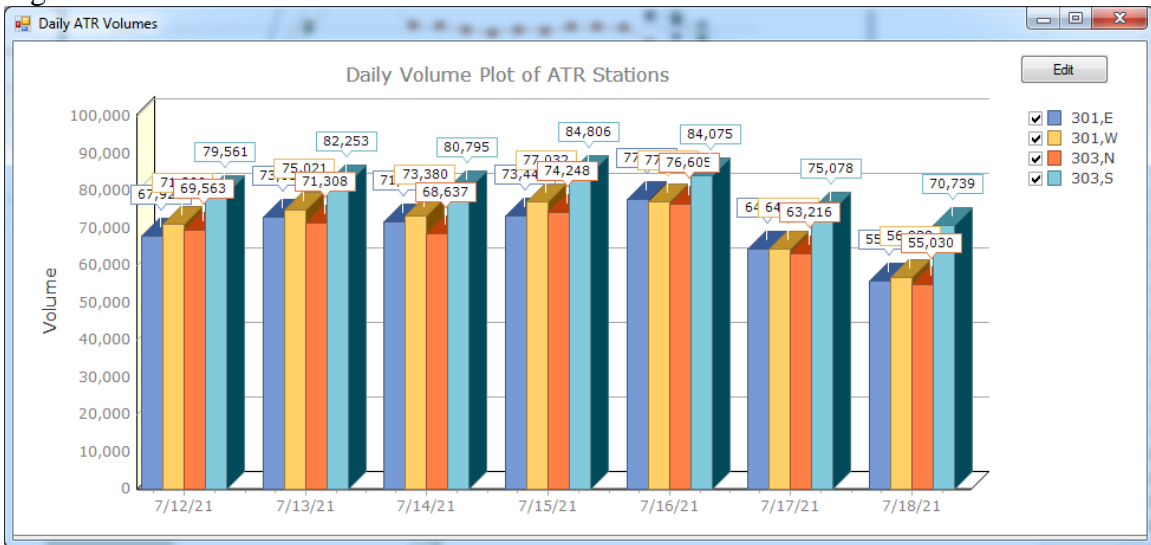


**FIGURE 6: Example daily plot**

The data plotted can be retrieved using Excel by pressing the "Export to a CSV file" button which should open up data in Excel, from which data may be saved.

**2.4 ATR Data and Log File Format**

ATR outputs produced by the ATR-TDA2 program are two text files: ATR data file and log file. The filenames conform to the format where the date is the ending date of the week, i.e.,

**ATRyyyymmddw1.dat**
**ATRyyyymmddw1.log**

In the filename, "w1" indicates that the file contains only one week of ATR data and the format is described next.

**Format of ATRyyyymmddw1.dat:**

All characters in the file are ASCII characters. One day data of a station in one direction occupies two rows: 12 hours per each row corresponding to AM in the first row and PM in the second row of the day. The format of a single line is summarized in Table 1.

**Table 1: Field Description of ATR Data File**

| Digit Position | Number of Digits | Description |
|---|---|---|
| 1 | 1 | Always 2. |
| 2 | 1 | AM=1, PM=2 |
| 3-4 | 2 | Month, 01-12 |
| 5-6 | 2 | Day of the month, 01-31 |
| 7-8 | 2 | Last two digit of the year |
| 9 | 1 | Day of the Week: Sun=1, Mon=2, Tue=3, Wed=4, Thu=5, Fri=6, Sat=7 |
| 10-12 | 3 | Station ID |
| 13 | 1 | Lane direction of the station, E,W,S,N,R |
| 14-73 | 60 | A Set of five digits represents the hourly volume. Sixty digits (12*5=60) are consecutively concatenated in the order representing hours $1^{st}$ to $12^{th}$ depending on AM or PM. |

## Interpretation of ATR data  by an example

Below was taken from top four rows of an actual ATR file, ATR20000131W1.dat.

210131002301E00662004980030900235002760089703106058400577204091038880 4217
220131002301E04678048380567206988807124065760500203348029820332602179 01497
210131002301W00631004260030000324005830230105530068960692805005044130 4565
220131002301W04565047570541505826066410684704897029360252802314018480 1073

Interpretation of the first line:
210131002301E00662004980030900235002760089703106058400577204091038880 4217

| Digit Position | Value | Meaning |
|---|---|---|
| 1 | 2 | ATR data, always 2 |
| 2 | 1 | AM |
| 3-4 | 01 | January |
| 5-6 | 31 | 31$^{st}$ day |
| 7-8 | 00 | Year 2000 |
| 9 | 2 | Day of the Week: Monday |
| 10-12 | 301 | Station ID = 301 |
| 13 | E | Lane Direction, East |
| 14-73 | 00662  … | A set of five digits representing the hourly volume. The data consists twelve five digits. |

## ATR Log-File Data Format:

The log file consists of two sections. The first section shows the statistics of missing detectors in the original traffic data and the missing percent of the station. Each set of these statistics is specified for Primary (P), Secondary (S), and Tertiary (T). Let's take an example.

> Inspecting missing det files and missing-data (MD) on Tuesday, August 04, 2009
301-3:: P: None, MD=.0%: S: None, MD=.0%: T: None, MD=.0%
301-7:: P: None, MD=.0%: S: None, MD=.0%: T: None, MD=.0%
303-1:: P: None, MD=.0%: S: None, MD=.0%: T: None, MD=.0%
303-5:: P: None, MD=.0%: S: None, MD=.0%: T: None, MD=.0%
305-1:: P: None, MD=.1%: S: None, MD=.0%: T: 0,MD=100.0%
305-5:: P: None, MD=.0%: S: None, MD=.0%: T: 0,MD=100.0%
309-1:: P: 341,342,343,MD=100.0%: S: None, MD=.0%: T: 335,336,337,338,-339,-340,MD=100.0%
309-5:: P: 178,179,189,MD=100.0%: S: 95,MD=25.0%: T: -181,-182,183,184,185,186,MD=100.0%
315-1:: P: 494,495,496,MD=100.0%: S: 266,267,525,1038,MD=100.0%: T: 156,268,269,270,533,MD=100.0%
315-5:: P: 256,257,1003,MD=100.0%: S: 1006,1007,1008,MD=100.0%: T: 110,254,255,1002,MD=100.0%

**Line-1:** 301-3:: P: None, MD=.0%: S: None, MD=.0%: T: None, MD=.0%
**Interpretation:** The station ID is 301 and the direction is 3. Primary (P), secondary (S), tertiary (T) had no missing detectors. "MD=.0%" means that the missing data percentage is 0%. This also means no imputation is needed.

**Line-7:** 309-1:: P: 341,342,343,MD=100.0%: S: None, MD=.0%: T: 335,336,337,338,-339,-340,MD=100.0%
**Interpretation:** The station ID is 309 and the direction is 1. The missing Primary detector IDs from the traffic file are 341, 342,343 and the missing data percentage is 100%. The secondary detectors are 100% available from the traffic file. The tertiary detectors are missing 100% from the traffic file, and the detector IDs are 335,336,337,338,-339, and -340. The detector IDs with a negative sign indicate subtraction of the detector volume from the total station volume. In this case, the ATR-TDA used the secondary detector set.

In the second section, the log data shows imputation details. The first line shows the station ID and direction, total volume of the day, and the percentage of imputation adjusted. The remaining part shows the hourly volume of 24 hours. For each hourly volume, it shows which detector set was used (P, S, or T), what was the volume before imputation (raw data), missing percent of raw data, and the amount of volume adjusted by imputation. The hourly volume follows the format described below:

[P, S, T, or B and volume before imputation]:[missing percent]:[the amount adjusted]

The letter "B" indicates that temporal and/or historic imputation occurred.

Let's take an example.

```
327-3  dailyVol=11427  ImpAdj=33.85%
P3:.0:0 P1:.0:0 P0:.0:0 P1:.0:0 P7:.0:0 P70:.0:0
P645:.0:0 P2057:.0:0 P1838:.0:0 P862:.0:0 P345:.0:0 P224:.0:0
B243:2.2:-57 B82:66.7:-4 B56:66.7:320 B228:66.7:413 B254:66.7:830 B263:66.7:987
B190:66.7:541 B117:66.7:147 B27:66.7:171 B31:66.7:269 B8:66.7:169 B7:66.7:82
```

**Line 1:** 327-3  dailyVol=11427  ImpAdj=33.85%
Station ID is 327-3; the daily total volume was 11,427; and imputation volume adjustment was 33.85%

**Line 2:** P3:.0:0 P1:.0:0 P0:.0:0 P1:.0:0 P7:.0:0 P70:.0:0

|              | Hour 0 | Hour 1 | Hour 2 | Hour 3 | Hour 4 | Hour 5 |
|--------------|--------|--------|--------|--------|--------|--------|
| Selected Set | P      | P      | P      | P      | P      | P      |
| Raw Volume   | 3      | 1      | 0      | 1      | 7      | 70     |
| Missing %    | 0      | 0      | 0      | 0      | 0      | 0      |
| Adjusted     | 0      | 0      | 0      | 0      | 0      | 0      |

**Line 3**: P645:.0:0 P2057:.0:0 P1838:.0:0 P862:.0:0 P345:.0:0 P224:.0:0

|  | Hour 6 | Hour 7 | Hour 8 | Hour 9 | Hour 10 | Hour 11 |
|---|---|---|---|---|---|---|
| Selected Set | P | P | P | P | P | P |
| Raw Volume | 645 | 2057 | 1838 | 862 | 345 | 224 |
| Missing % | 0 | 0 | 0 | 0 | 0 | 0 |
| Adjusted | 0 | 0 | 0 | 0 | 0 | 0 |

**Line 4**: B243:2.2:-57 B82:66.7:-4 B56:66.7:320 B228:66.7:413 B254:66.7:830 B263:66.7:987

|  | Hour 12 | Hour 13 | Hour 14 | Hour 15 | Hour 16 | Hour 17 |
|---|---|---|---|---|---|---|
| Selected Set | B | B | B | B | B | B |
| Raw Volume | 243 | 82 | 56 | 228 | 254 | 263 |
| Missing % | 2.2 | 66.7 | 66.7 | 66.7 | 66.7 | 66.7 |
| Adjusted | -57 | -4 | 320 | 413 | 820 | 987 |

**Line 5**: B190:66.7:541 B117:66.7:147 B27:66.7:171 B31:66.7:269 B8:66.7:169 B7:66.7:82

|  | Hour 18 | Hour 19 | Hour 20 | Hour 21 | Hour 22 | Hour 23 |
|---|---|---|---|---|---|---|
| Selected Set | B | B | B | B | B | B |
| Raw Volume | 190 | 117 | 27 | 31 | 8 | 7 |
| Missing % | 66.7 | 66.7 | 66.7 | 66.7 | 66.7 | 66.7 |
| Adjusted | 541 | 147 | 171 | 269 | 169 | 82 |

In summary, the hourly volumes for the above 5 lines of log data indicate:

| Hour | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Before Imp | 3 | 1 | 0 | 1 | 7 | 70 | 645 | 2057 | 1838 | 862 | 345 | 224 |
| Missing % | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| After Imp | 3 | 1 | 0 | 1 | 7 | 70 | 645 | 2057 | 1838 | 862 | 345 | 224 |

| Hour | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Before Imp | 243 | 82 | 56 | 228 | 254 | 263 | 190 | 117 | 27 | 31 | 8 | 7 |
| Missing % | 2.2 | 67 | 67 | 67 | 67 | 67 | 67 | 67 | 67 | 67 | 67 | 67 |
| After Imp | 186 | 78 | 376 | 641 | 1074 | 1250 | 731 | 264 | 198 | 300 | 177 | 89 |

From this data, we can learn that two thirds of detectors did not work during the 12nd – 23rd hours. After imputation, the afternoon peak hour volumes were increased to a normal range.

**FHWA Volume Format**

It contains hourly data for each station for the period defined, typically one year. It follows the standard format defined by Traffic Monitoring Guide, May 1, 2001. There are some default values set, which are described including variables in the following table.

| Column(s) | Description |
| --- | --- |
| 1 | Record Type="3", Traffic volume record |
| 2-3 | FIPS State Code="27", Minnesota |
| 4-5 | Functional Classification Code="12" |
| 6-11 | Station Identification=SC Sequence number in 6 digites |
| 12 | Direction of Travel Code=SC Station Direction |
| 13 | Lane of Travel="0" , combined lanes |
| 14-15 | Year of Data=yy |
| 16-17 | Month of Data=MM |
| 18-19 | Day of Data=dd |
| 20 | Day of Week (1=Sun, 2=Mon, 3=Tue, 4=Wed, 5=Thr, 6=Fri, 7=Sat) |
| 21-25, …, 136-140 | Traffic Volume Counted Field=(5digit hourly vol) * 24 |
| 141 | Restrictions="0", no restrictions |

# 3. Concluding Remarks

The mathematical procedures applied for imputation was not described in this manual. If anyone wishes to explore the mathematical details, the two books [1,2]and the research report [3] in the Reference section are recommended.

The ATR-TDA2 software package will be updated regularly, typically triggered by the following events (1) error was found, (2) operating system (OS) versions have been changed, (3) RTMC changed their data format, or (4) after a major upgrade. Presently, the OS supported by the software are Windows 7 and 10 and it requires .Net Framework 4.5. The latest versions of the software are distributed through the web site:

http://www.d.umn.edu/~tkwon/Download/mndotDownload.htm

# 4. References

[1] Rubin, Donald B., *Multiple Imputation For Non-Response in Surveys*, Wiley Series in Probability and Mathematical Statistics, 1987.

[2] Little, R. J. A. and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley Series in Probability and Mathematical Statistics, 1987.

[3] T.M. Kwon, *TMC Traffic Data Automation for Mn/DOT's Traffic Monitoring Program*, Minnesota Department of Transportation, Report No. MN-RC-02004-29, July, 2004.