SmoothData (version 4.0.5)

Revision History
10 December 2010 Created by Volpe Center

The SmoothData program is used to smooth diurnal distribution curves prior to input to ConvertTrips in TRANSIMS. It applies several iterations of a moving average technique to minimize the impact of random time-of-day spikes.

SmoothData is a console-based program that runs in a command window on either Windows or Linux. The command syntax is:

```
SmoothData [-flag] [control_file]
```

The control_file is the file name of an ASCII file that contains the control strings expected by the program. The control_file is optional. If a file name is not provided, the program will prompt the user to enter a file name. The flag parameters are also optional. Any combination of the following flag parameters can be included on the command line:

-Q[uiet] = execute without screen messages
-H[elp] = show program syntax and control keys
-K[eyCheck] = list unrecognized control file keys
-P[ause] = pause before exiting
-N[oPause] = never pause before exiting
-B[atch] = execute in batch processing mode

The program automatically creates a printout file based on the control_file name. If the file name includes an extension, the extension is removed and ".prn" is added. The printout file will be created in the current working directory and will overwrite an existing file with the same name.

Control File Examples

EXAMPLE 1 CREATE SMOOTHED TIME OF DAY DISTRIBUTION

```
INPUT_DATA_FILE_1             TripTimeRough.txt
INPUT_DATA_FORMAT_1           TAB_DELIMITED
OUTPUT_DATA_FILE_1            TripTimeSmooth.txt
OUTPUT_DATA_FORMAT_1          TAB_DELIMITED

SMOOTH_FIELD_NUMBER           3
SMOOTH_GROUP_SIZE             3
SMOOTH_TIME_INCREMENT         900
PERCENT_MOVED_FORWARD         20.0
PERCENT_MOVED_BACKWARD        20.0
NUMBER_OF_ITERATIONS          10
CIRCULAR_GROUP_FLAG           TRUE
```

The keys define a 10-iteration (NUMBER_OF_ITERATIONS) moving average involving three (SMOOTH_GROUP_SIZE) 15-minute time periods (SMOOTH_TIME_INCREMENT) with 20 percent forward (PERCENT_MOVED_FORWARD) and backward shifts (PERCENT_MOVED_BACKWARD)on the third column (SMOOTH_FIELD_NUMBER) of the TripTimeRough.txt file.

Control File Parameters

Control parameters are defined using a control key followed by a string or number. The control parameters can be specified in any order. If a given key is defined more than once, the last instance of the key is used. Note that comment lines or extraneous keys can be included in the file. They will be ignored by the program.

The keys recognized by the **SmoothData** program are listed below. These keys can be defined in a variety of different ways to perform different tasks.

A number of output examples are presented in graphs later in this document. Unless otherwise specified, they use the parameters as stated in Example 1, above.

Required Keys

INPUT_DATA_FILE

The name of the input file, typically a diurnal distribution file. An excerpt from such a file appears below:

START	END	SHARE
0	0.25	0.001
0.25	0.5	0.001
0.5	0.75	0.0011
0.75	1	0.001
1	1.25	0.0009
1.25	1.5	0.0008
…	…	…
22.5	22.75	0.0073
22.75	23	0.0062
23	23.25	0.0072
23.25	23.5	0.0045
23.5	23.75	0.0042
23.75	24	0.004

OUTPUT_DATA_FILE

The name of the output file that contains the smoothed time distribution.

NUMBER_OF_ITERATIONS

The Number_of_Iterations defines how many times the moving average process

is applied to the full set of time periods. The range is 1 to 25. If NUMBER_OF_ITERATIONS is not supplied, the default is 0, and then the SMOOTH_DATA program does nothing. Typically, a value of 10 is used.

Figure 1 shows the impact of the NUMBER_OF_ITERATIONS parameter, showing the difference in outputs between a parameter of 3 and a parameter of 10. The higher parameter produces more smoothing.

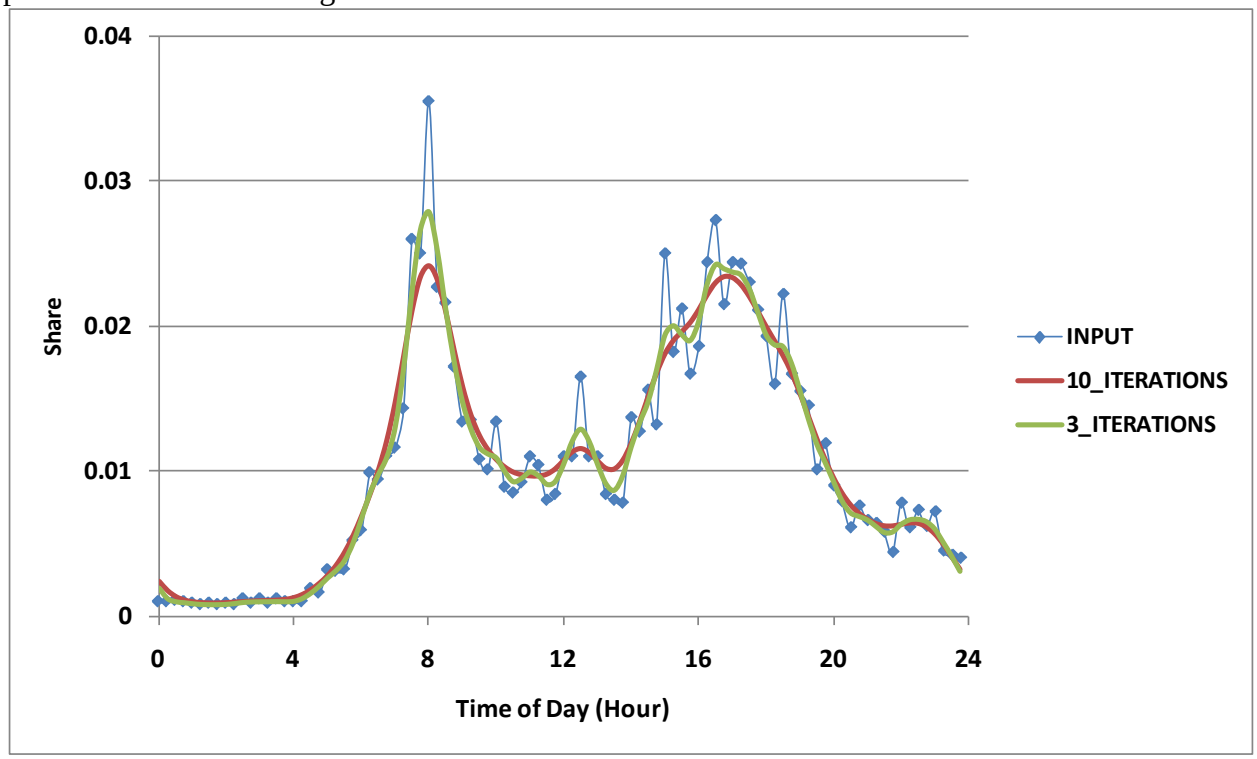

Figure 1 Impact of Number of Iterations

SMOOTH_FIELD_NUMBER

The number of the column in the input file that is being smoothed. This is typically field number 3 in a diurnal distribution file.

Optional Keys

CIRCULAR_GROUP_FLAG

The CIRCULAR_GROUP_FLAG indicates whether smoothing should span the midnight boundary. Table 1 illustrates the impact. It shows a 15-minute diurnal distribution with much higher values just before midnight than just after midnight. When the parameter is set to TRUE (the default), the smoothing spans the midnight boundary. When it is FALSE, there can be an abrupt jump at the midnight boundary.

Table 1 Impact of CIRCULAR_GROUP_FLAG

START TIME	*INPUT*	*OUTPUT if TRUE*	*OUTPUT if FALSE*
23	0.0072	0.005668	0.005787
23.25	0.0045	0.00496	0.005288
23.5	0.0042	0.004103	0.004837
23.75	0.004	0.003203	0.004571
0	0.001	0.002381	0.001013
0.25	0.001	0.001741	0.001007
0.5	0.0011	0.001319	0.000991
0.75	0.001	0.001082	0.000963
1	0.0009	0.000962	0.000928

REPLICATE_FIELDS_FLAG

An input file might contain fields in addition to Start, End and Share. If the Replicate_Fields_Flag is set to TRUE, these added fields are replicated in the output. The default is FALSE.

SMOOTH_GROUP_SIZE

Smooth_Group_Size is the number of time periods included in a moving average smoothing process. The default is 3, which is normally used for 15 minute time periods (i.e., +1 and -1 period). Possible values are 3, 5, 7, 9. Higher values result in more smoothing. Figure 2 is a plot of 15-minute diurnal shares. It shows the different impacts of SMOOTH_GROUP_SIZE = 3 (the default), and SMOOTH_GROUP_SIZE = 9.

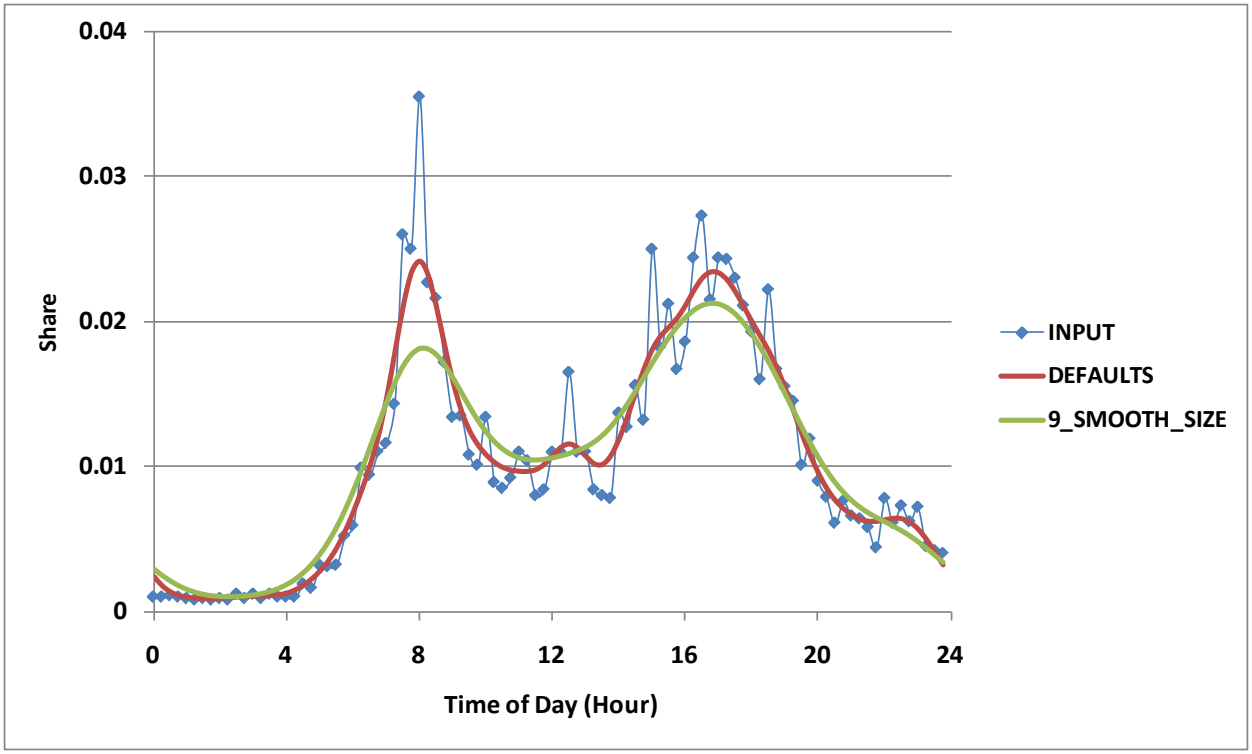

Figure 2 Impact of Smooth Group Size

PERCENT_MOVED_FORWARD

The Percent_Moved_Forward defines the percentage of the current period volume that is added to the next time period. It defaults to 20 percent. The range is 0..(100 – 50 / [smooth group size])

PERCENT_MOVED_BACKWARD

The Percent_Moved_Backward defines the percentage of the current period volume that is added to the previous time period. It defaults to 20 percent. The range is 0..(100 – 50 / [smooth group size])

Figure 3 shows the impact of reducing these parameters from 20% to 5%. With the reduction to 5%, there is less smoothing. Figure 4 shows the impact of different parameters for Percent_Moved_Forward and Percent_Moved_Backward. In this case, Percent_Moved_Forward = 40 and Percent_Moved_Backward =0. The result is a shift to the right.

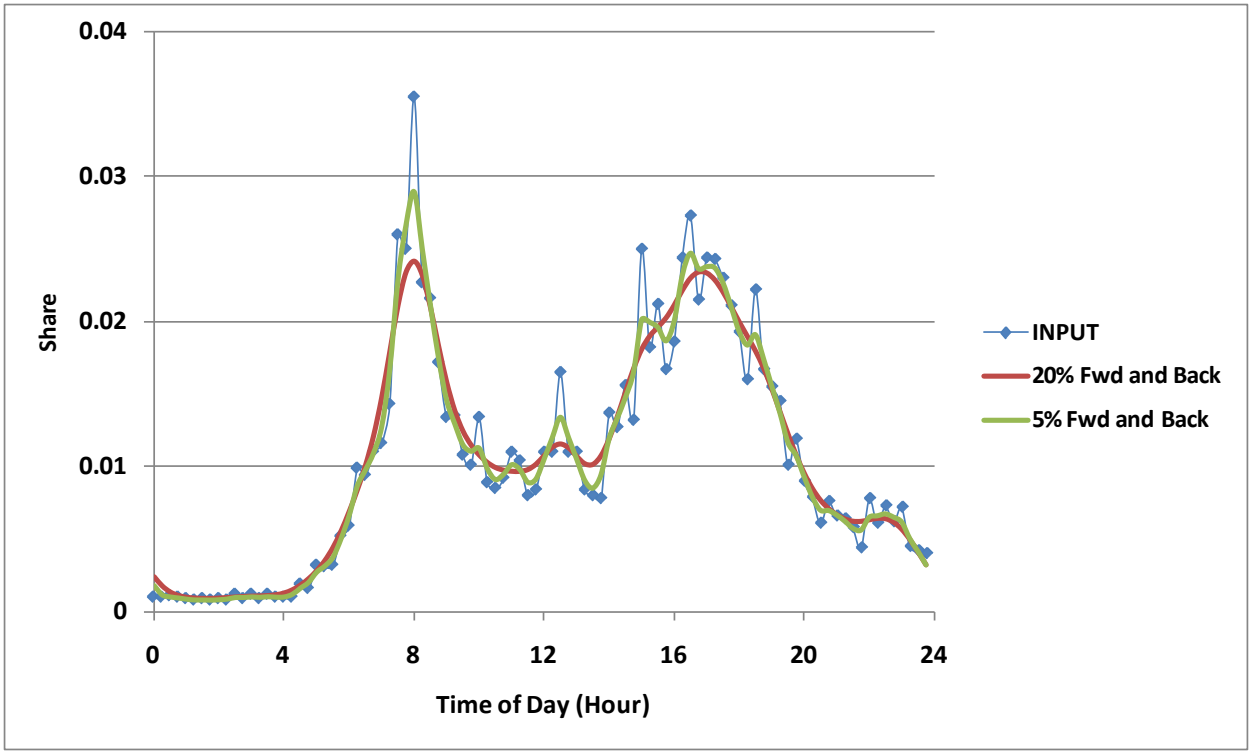

Figure 3 Impact of Reducing Percent Forward and Backward

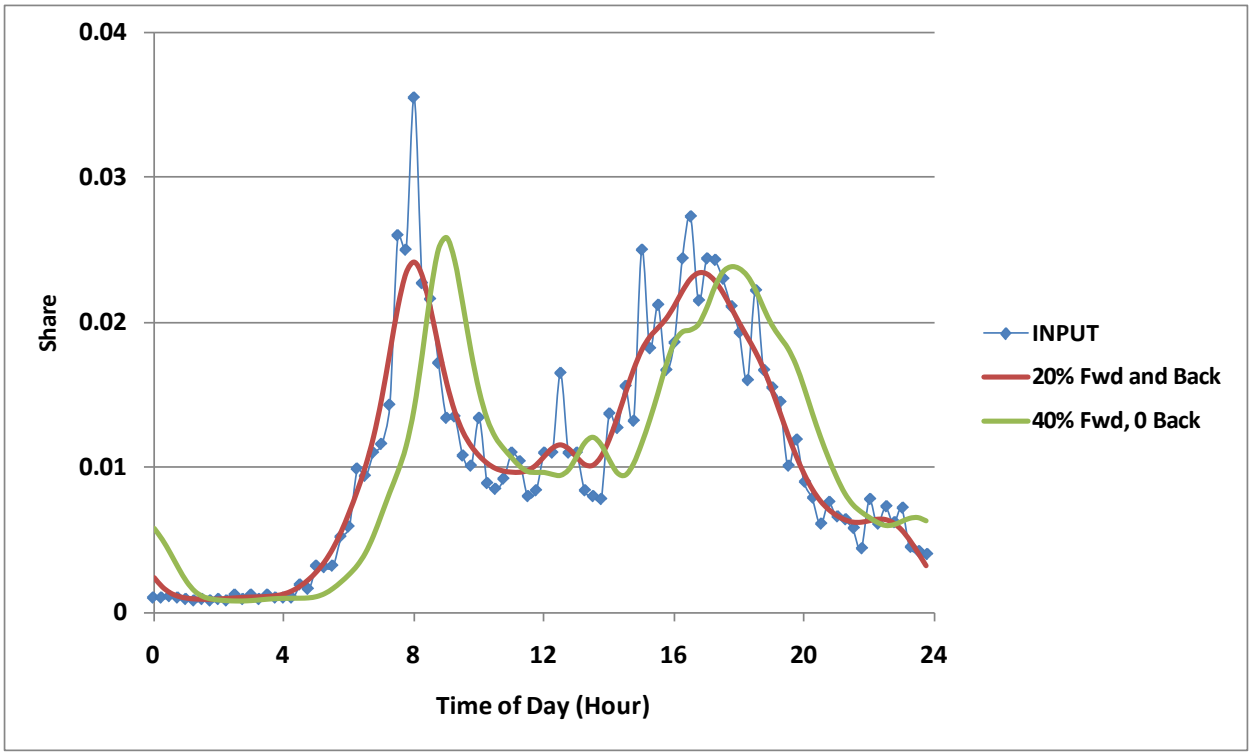

Figure 4 Impact of 40% Forward, nothing Backward

SMOOTH_TIME_INCREMENT

The number of seconds in each diurnal bucket of the output file. The default is to match the bucket size on the input file. The value should be equal to or less than the number of seconds in each bucket of the input file.

MAX_WARNING_EXIT_FLAG

If the maximum number of warning messages is exceeded, this flag directs the program in what to do. If the flag is TRUE (the default), the program is terminated with an error message about the warning messages. If the flag is FALSE, the program continues execution, but no additional warning messages are sent to the screen or written to the printout file. The warning message counter continues to count the messages and reports the total at the end of the execution.

MAX_WARNING_MESSAGES

When the program generates a warning message, a counter is incremented and the total number of warning messages is reported and a warning return coded (2) is set at the end of the execution. By default the program prints up to 100,000 warning messages to the print-out file. If more than 100,000 warning messages are sent, the program stops printing additional messages to the file or terminates the program with an error message.

MAX_WARNING_EXIT_FLAG

This parameter enables the user to modify the default warning limit.

PROJECT_DIRECTORY

The project directory key is not required. If it is specified, it is added to all non-network file names required by the program. If it is not specified, all non-network file names should fully specify the file path.

REPORT_FILE

The report file name is optional. If a file name is not provided, the program automatically creates a report file name based on the input control file name plus the partition number. The report file will overwrite an existing file with the same name if the Report Flag key is False or not specified.

REPORT_FLAG

The report flag key is optional. If it is specified as Yes or True, the report file or default printout file will be opened in 'Append' mode rather than 'Create' mode. This permits the user to consolidate the output of several programs into a single report file.

TITLE

Any text string can be used on this line. This text is printed on the top of each output page.

DEFAULT_FILE_FORMAT

Default format for files other than network files. Default is VERSION3. Other possible values include BINARY, FIXED_COLUMN, COMMA_DELIMITED, SPACE_DELIMITED, TAB_DELIMITED, CSV_DELIMITED, DBASE, LANL and SQLITE3.

DISTRIBUTION_FILE

Undocumented!

DISTRIBUTION_FORMAT

Undocumented!

INPUT_DATA_FORMAT

Format of the input file. Default is VERSION3. Other possible values include BINARY, FIXED_COLUMN, COMMA_DELIMITED, SPACE_DELIMITED, TAB_DELIMITED, CSV_DELIMITED, DBASE, LANL and SQLITE3.

INPUT_TIME_FORMAT

The time of day format defines how the start and end times are presented in the input file. The default format will display values in hours. The format options include HOURS, SECONDS, 24_HOUR_CLOCK, and 12_HOUR_CLOCK.

OUTPUT_DATA_FORMAT

Format of the output file. Default is VERSION3. Other possible values include BINARY, FIXED_COLUMN, COMMA_DELIMITED, SPACE_DELIMITED, TAB_DELIMITED, CSV_DELIMITED, DBASE, LANL and SQLITE3.

OUTPUT_TIME_FORMAT

The time of day format defines how the start and end times are presented in the input file. The default format will display values in hours. The format options include HOURS, SECONDS, 24_HOUR_CLOCK, and 12_HOUR_CLOCK.

Sample Printout

A sample printout file generated by the **SmoothData** program is shown below. Each printout is an ASCII text file with a maximum of 95 characters per line and 65 lines per page. The file can be viewed or printed using a variety of text editors. For best results in a word processor, use a 10-point Courier font and 0.5 inch margins on all sides.

```
*****************************************
|                                       |
|        SmoothData - Version 4.0.5     |
|   Copyright (c) 2009 by AECOM Consult |
|         Mon Dec 06 12:20:26 2010      |
|                                       |
*****************************************

Control File = SmoothData.ctl
Report_File  = SmoothData.prn (Create)

Moving Average Smoothing Procedure

Input Data File #1 = TripTimeRough.txt
Input Data File #1 Format = TAB_DELIMITED
Output Data File #1 = TripTimeSmooth.txt
Output Data File #1 Format = TAB_DELIMITED
Input Time Format = HOURS
Output Time Format = HOURS
Smooth Field Number = 3
Smooth Time Increment = 900 seconds
Smooth Group Size = 3
Percent Distributed Forward = 20.0 percent
Percent Distributed Backward = 20.0 percent
Number of Iterations = 10
Circular Group Flag = TRUE
Replicate Fields Flag = FALSE

Number of Input Data File #1 Records = 96
Number of Output Data File #1 Records = 96

Mon Dec 06 12:20:26 2010 -- Process Complete (0:00:00)
```