APPENDIX B

ADVANCE

Advanced Driver and Vehicle Advisory Navigation Concept

Traffic Related Functions
Evaluation Report (2 of 7)
Documents # 8460.00

CONTAINS:

Base Data and Static Profile Evaluation Report	**-- Document # 8460-1.00**
• Data Screening Evaluation Report	**-- Document # 8460-02.02**
Quality of Probe Reports Evaluation Report	**-- Document # 8460-03.01**
***Travel* Time Prediction and Performance of Probe and Detector Data Evaluation Report**	**-- Document # 8460-04.00**
Detector Travel Time Conversion and Fusion of Probe and Detector Data Evaluation Report	**-- Document # 8460-05.00**
Frequency of Probe Reports Evaluation Report	**-- Document # 8460-06.00**
Relationships among Travel Times Evaluation Report	**-- Document # 8460-07.02**

Prepared by
University of Illinois-Chicago
Urban Transportation Center

DISCLAIMER

ADVANCE Evaluation

Data Screening

Stanislaw Berka, Aaron Sheffey, and Ashish Sen

URBAN TRANSPORTATION CENTER
University of Illinois at Chicago
1033 West Van Buren Street
Suite 700 South
Chicago, Illinois 60607

March, 1996

Executive Summary

A key component of the Traffic Related Functions (TRF) subsystem of the ADVANCE Project is Data Screening (DS). Evaluating travel time and congested-distance data collected from probe (vehicle) reports, as well as in-pavement loop detector reports which contain volume and occupancy data, DS identifies potentially incorrect information. The DS module evaluates both sets of data by assessing the reasonableness of each then processing the overall consistency of the information within the set. This process determines the accuracy of the information gathered from the probe and detector reports. If the information is deemed to be inconsistent or incorrect, it is removed from the data set.

The DS algorithm was first tested using simulated data. For both probe and detector data, limits for given links and time periods were constructed. Three test figures, two of which were designed to be flagged by DS and one meant to pass DS, were then run through the algorithm. The results of these tests were then evaluated to see whether those simulated reports designed to be eliminated by the DS algorithm were in fact flagged, and whether those designed to pass, actually did. For the simulated data, the DS algorithm performed as designed.

The results of using actual field data were cross-checked against manually-recorded measurements and recorded incident data. DS for actual data led to an overall probe data consistency success rate of greater than 99% and a 92% success rate for detector data consistency. Probe data reports which failed DS without providing an incident flag indicated some potential inconsistencies in the MNA's formulation of the original data set. It was concluded that the DS algorithm was effective in identifying rare occasions when MNA's malfunctioned and its use should be continued.

Contents

1 Introduction

Data screening is part of the Data Fusion subcomponent of the Traffic Related Functions of the ADVANCE system and runs on-line on the TIC computer. It is performed to identify those probe and detector reports which appear to correspond to unusual traffic conditions or probe reporting errors. Identified reports which could have an adverse effect on the results of travel time predictions are marked. Two phases of data screening are distinguished:

1. reasonableness for a single datum and

2. consistency checking of a pair of data.

As evaluated in data screening, all data reports consist of two data items. In the case of probe reports, these items are travel time (t_p) and congested distance (l_c). In the case of detector reports, these items are volume (v) and occupancy (o). Each report is tested for the reasonableness of each datum, and for the mutual consistency of a pair of data.

Datum x, which is a measurement of travel time, congested distance, volume, or occupancy for a given link at a certain time (see Section 2.3), is considered reasonable if it meets the following condition:

$$x_{min} \leq x \leq x_{max}, \tag{1}$$

where x_{min} and x_{max} are appropriate cut-offs. For example, for a signalized link, the travel time t_p passes the reasonableness test if it fulfills the following condition:

$$\frac{l}{f_v \cdot v_m} \leq t_p \leq \frac{3600 \cdot l}{l_Q \cdot S \cdot (\frac{g}{c})_{min}} + r_{max}. \tag{2}$$

[Note: To avoid duplication, not all tests used in data screening will be cited here. Comprehensive descriptions can be found in Berka, Tian, and Tarko (1995), which is available on the web (http://beijing.dis.anl.gov/ADVANCE).] Symbols in the above expression denote fixed parameters defined by the data screening algorithm and are defined as follows:

f_v	=	adjustment factor for random variations of travel speed
$(\frac{g}{c})_{min}$	=	Minimum green-to-cycle ratio (-)
l	=	length of a base segment (m)
l_Q	=	effective length of vehicle in queue (m)
r_{max}	=	maximum length of red signal (sec)
S	=	saturation flow, which is lane capacity if entire signal cycle is effectively green (veh/h)
t_p	=	individual probe travel time along a link (sec)
v_m	=	posted speed limit (m/s).

A pair of data (x, y) is considered consistent it if meets the following condition:

$$y_{min}(x) \leq y \leq y_{max}(x). \tag{3}$$

Where $y_{min}(x)$ and $y_{max}(x)$ depend on characteristic x. For example, for a signalized link, the travel time t_p is considered to be consistent with congested distance l_c if it meets the following condition:

$$\frac{l}{t_p} \geq \frac{l}{a_{L1} \cdot \frac{N_l \cdot l_c}{C_{lg}} + a_{L2} \cdot \frac{l - l_c}{v_m} + a_{L3} \cdot (c-g)} - k_a \cdot \sigma_{aL} \tag{4}$$

and

$$\frac{l}{t_p} \leq \frac{l}{a_{L1} \cdot \frac{N_l \cdot l_c}{C_{lg}} + a_{L2} \cdot \frac{l - l_c}{v_m} + a_{L3} \cdot (c-g)} + k_a \cdot \sigma_{aL}, \tag{5}$$

where

a_{L1}	=	parameter (= 470) in the probe report consistency relationship for an arterial link outflow controlled with a traffic signal
a_{L2}	=	parameter (= 1.3) in the probe report consistency relationship for an arterial link outflow controlled with a traffic signal
a_{L3}	=	parameter (= 0.5) in the probe report consistency relationship for an arterial link outflow controlled with a traffic signal
c	=	signal cycle (sec)
C_{lg}	=	lane group capacity (veh/hr)
g	=	green signal (sec)
k_a	=	standard deviation multiplier (= 3) for arterial probe data relationship (-)
l	=	length of a base segment (meter)
l_c	=	congested distance which is the distance traversed at a speed below the critical value (meter)
N_l	=	number of lanes in a group (-)
t_p	=	individual probe travel time along a link (sec)
σ_{aL}	=	standard deviation for probe report consistency relationship and link with an outflow controlled with a traffic signal (meter/sec).

Notice that unlike the single datum case, the cut-offs are based on reported data. The description of the data screening algorithm can be found in Berka, Tian and Tarko (1995) and the calibrated parameters in Berka and Tian (1995).

2 Evaluation Procedure

2.1 Overview

The evaluation procedure consists of several steps. First, the attribute database, needed in subsequent steps, is checked to make sure it contains all necessary information. Second, an evaluation using simulated data is performed to determine the veracity of the implementation. Third, an evaluation using actual field data is done. Fourth, an analysis of both the reports which pass data screening, as well as those which fail data screening, is performed.

2.2 Verification of Attribute Database

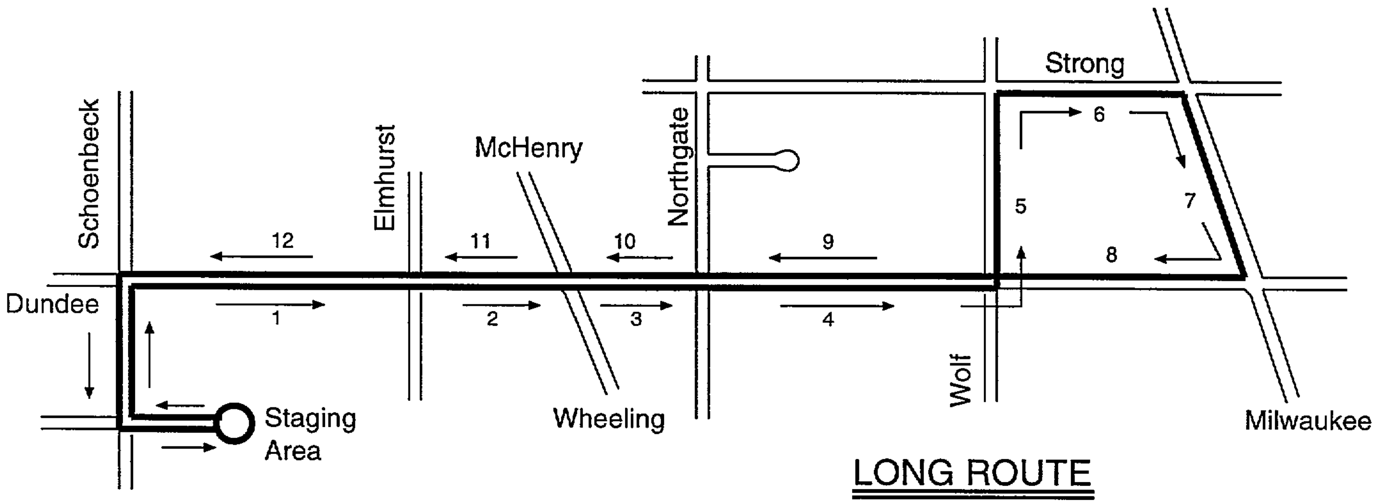
The Attribute Database contains data elements required by the data screening algorithm. It includes elements of the ANR (ADVANCE Network Representation), an ASCII file which combines information from the CATS file, field observer logs, and the Motorola Interchange File (MIF) file to describe a given ADVANCE test area. This information contains traffic related characteristics of the transportation network, including approach geometry, speed limits, and type of control. All flow characteristics (volume, capacity, green times, and cycles) are derived from NFM (Network Flow Model) results.

2.3 Evaluation Using Simulated Data

Actual parameters of the TRF-DF algorithm were obtained from TIC. Representative links for specific categories, such as those comprising four-way stops and left-turns, were selected from the evaluation route (see Figure 1) to serve as test subjects. For each selected link, the limits x_{min}, x_{max} and y_{min}, y_{max} were calculated for specific data elements. These limits were defined by the screening conditions for each link (see Section l), and were calculated in such a way that for each screening condition, one report would violate the lower bound (x_{min}) one would violate the upper bound (x_{max}), and one would satisfy the condition.

As mentioned above, probe and detector data comprise different data elements. To pass the screening, a data element for a given report would have to fall within its respective limits of reasonableness so that x_{min} < data element < x_{max}. For probe reports, these data elements are travel time and congested distance; for detector reports, they are volume and occupancy. Using the same procedure, these data elements are then screened in pairs for consistency (travel time vs. congested distance for probe reports, and volume vs. occupancy for detector reports). In sum, there are 6 screening procedures which can be performed for information from a given link. For example, for (2), representing a ‘reasonableness’ condition, the limits of travel time for the probe report from link 1 shown Figure 1 were x_{min} = 35.7 sec and x_{max} = 7998.7 sec (it is clear that the upper limit will hardly ever be reached; it can be adjusted using some educated guessing). Three reports were then generated with travel times of 30, 3000, and 8500 seconds. Of the three generated reports, only one, with a t_p of 3000 seconds, was designed to satisfy this condition.

The data screening algorithm was performed for such simulated data, generating output indicating which data passed and which failed the reasonableness and consistency tests. Results were compared with the test design; that is, whether reports which were generated to pass in fact passed, and whether those generated to fail actually failed.



ADVANCE LINK	ID
88cb2b - 88cae7	1
88cae7 - 891a48	2
891a48 - 8c689	3 short
891a48 - 88cad8	3 long
88cad8 - 8c9b3	4
8c9b3 - 88c906	5
88c906 - 88c9a8	6
88c9a8 - 8cad2	7
8cad2 - 8cad8	8
8cad8 - 91a48	9
91a48 - 8cae7	10
8cae7 - 8cb2b	11
8cb2b - 8d1f1	12

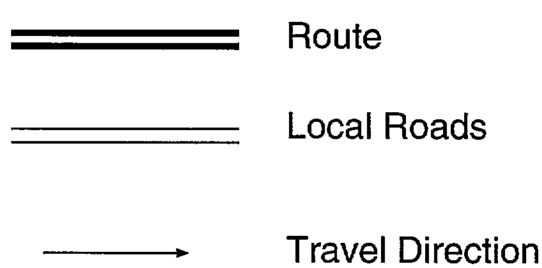


Figure 1: Test Route

2.4 Evaluation Using Actual Data

The critical issue facing evaluation of the actual data is a lack of information as to which reports are in fact correct, and which are not correct. Two limited sources can be utilized for this determination. The first source is the manually recorded measurements performed by observers in the field (using stopwatches), done as part of the Quality of Probe Reports evaluation task. The primary limitation of this source is observer error, the impact of which is unknown. In addition, this source lacks information which could be used to validate the detector reports.

The second source which provides data related to the correctness of the reports is the presence or absence of an incident for a given link during a specified time interval, for both probe and detector data. Because incidents can cause cars to remain motionless for extended periods, we can expect that the presence of an incident may cause some reports to be incorrect. On the other hand, the data recorded during an incident may in fact be correct. While data screening is not designed for use with data generated in the presence of an incident, the lack of available information forces the use of incident data. After the data screening is performed, those reports from the original data sets which have incidents are then compared with those reports which were screened. While not exact, this comparison can act as a tentative benchmark of the correctness of the screening process.

To verify the performance of the data screening algorithm, the actual probe and detector data for selected links were processed off-line by the data screening algorithm. The results from this screening were compared with the incident data derived from the drivers' logs. If there was an incident marked by the driver for some link and some 5-minute time interval, all detector and probe links for this link and this time interval were assigned the same incident flag. [Note: For detector data, only those links with detectors (i.e., links 1, 7, and 11) were used.] This process resulted in each report being assigned to one of the following cases:

1. report failed, incident present,
2. report failed, no incident present,
3. report passed, incident present, and
4. report passed, no incident present.

A report is considered to have failed data screening if any one of the three screening conditions (two reasonableness tests and one consistency test) failed. Otherwise, the report is considered to have passed screening. However, data screening results are considered subject to the limitations of both data sources mentioned above.

2.5 Analysis of Failed Reports

Link reports which failed the reasonableness and/or consistency tests should correspond to either an incident or a probe reporting malfunction. With that premise in mind, failed reports were analyzed to determine which reports were obviously erroneous, and which should have passed the screening. A comparison of reports which should have passed the screening to those which did in fact fail was then developed.

2.6 Analysis of Passed Reports

As with those reports which failed data screening, reports which passed data screening were checked for possibly erroneous reports, including those reports marked with incidents. As mentioned above, an incident can be an indication of an unusual traffic condition. However, the effect of that incident may be quite unexpected and the corresponding reports themselves may be similar to non-incident reports for congested conditions. Due to a lack of precise criteria to determine incorrect reports, this analysis was somewhat limited.

3 Simulated Data Generation, Processing and Analysis

Links selected for particular tests using the simulated data are presented in the following table. Link numbers used in this table conform to those numbers both on the map and in the table in Figure 1. An 'X' indicates that a report was generated for that particular link to test that particular screening condition. Because some reasonableness and consistency formulae are identical for different categories of links, not all procedures were performed for all links in this step.

Link	Travel Time Reason.	Congested Distance Reason.	Probe Report Consist.	Volume Reason.	Occupancy Reason.	Detector Report Consist.
4	X	X	X			
6	X		X			
7				X	X	X
8	X					

Links in the above table cover the following categories.

1. Link controlled by traffic lights:
   (a) link 4 - left-turn movement
   (b) link 8 - through movement
   (c) link 7 - right-turn movement
2. Link controlled by STOP sign: link 6

A total of twenty-six (26) reports were generated: eighteen (18) probe reports and eight (8) detector reports. These data were then processed by the stand-alone data screening code, a copy of the original code prepared by the EECS group and executed on the TIC computer. The output contains information about the result of each test performed, along with the input data. An example of the output from this code follows:

```
6 19 95   19 45 12    8e3c4    8e4e6   (tt=46.00,cd=0.00)   RTT=1   RCD=l   CProbe=l
6 19 95   19 45 12    92f99    8e60b   (tt=l.98,cd=0.00)    RTT=O  RCD=l   CProbe=0
6 19 95   19 45 12    912ba    a62cl   (tt=43.00,cd=43.00)  RTT=1   RCD=l   CProbe=l
6 19 95   19 45 13   88e738    8ee53   (tt=40.00,cd=0.00)   RTT=l   RCD=1   CProbe=l
6 19 95   19 45 13    8cae7    91036   (v=65,o=4)          RVol=1  ROccp=l CDet=1
6 19 95   19 45 13   88cc8a    Qd42f   (v=67,o=3)          RVol=1 ROccp=l  CDet=O
6 19 95   19 45 18    8cb24    8cb20   (v=40,o=2)          RVol=1  ROccp=l CDet=l
6 IQ 95   19 45 18   888079  88cb24   (v=47,o=3)          RVol=l  ROccp=l CDet=l
```

In the above output, one line represents the screening results for one report. The fields in the output are as follows: 1 and 2 give date and time, respectively; 3 and 4 define the link using hex identifiers; 5 presents actual data reported (tt = travel time, cd = congested distance, v =volume, o = occupancy); and 6-8 show the results of the tests performed for the report, 0 indicating a failed test, 1 a passed test. The abbreviations in fields 6-8 are as follows: RTT is travel time reasonableness, RCD is congested distance reasonableness, CProbe is probe report consistency, RVol is volume reasonableness, ROccp is occupancy reasonableness, and CDet is detector report consistency.

4 Actual Data Generation, Processing, and Analysis

Like the evaluation using simulated data, the evaluation utilizing actual probe data also used links covered by the Test Route as shown in Figure 1. As noted above, in the case of the detector data, only links equipped with a detector were used (links 1, 7, and 11).

For the analysis, the MNA and detector report log files for June 19-22 and Jul 17-20 were retrieved from the Argonne National Laboratory WWW home page, reduced, reformatted appropriately, and processed by the stand-alone Data-Screening module mentioned above. In the case of the detector data, the data aggregated over all detectors within a detector station were used. These data were in the same format as those received on-line by the data screening algorithm. Output data were matched with incident data obtained from the drivers' logs. For each 5-minute interval starting from 12:OO pm, if there was an incident flag for a link, all probe and detector reports for this link and interval were assigned the same incident flag.

5 Results

5.1 Simulated Data

All simulated reports generated as described in Section 2.3 were processed by the data screening algorithm. In other words, testing was done to see whether those reports generated to fail the testing conditions would fail and whether those generated to pass the condition would pass. Because it needed to be determined if the data screening process was working in line with expectations, this first step needed to be concluded before analysis of the actual data could be initiated. Thus, this round of testing acted as the verification of the correctness of the implementation of the data screening algorithm.

5.2 Actual Probe Data

The results of the analysis of the actual probe data are presented in the following table.

	Passed Screening	Failed Screening	Total
No incidents	14704	102	14806
Incidents	357	8	365
Total	15061	110	15171

The presence of incidents in the above table corresponds to appropriate entries in log files kept by the drivers. The following types of incidents were recorded by the drivers: left-turn blocks traffic, through blocks traffic, train, construction, accident, emergency vehicle, and rain/weather. The large number of incidents in the log files indicates that most incidents were minor.

If there had been no relation between screening and the presence of incidents, but the total number of incidents and data-screening failures had been the same, we would have obtained the following table (values have been rounded to whole numbers to make the interpretation easier).

	Passed Screening	Failed Screening	Total
No incidents	14699	107	14806
Incidents	362	3	365
Total	15061	110	15171

Assuming the independence of the passage/failure of the individual reports from each other, results were analyzed using a chi-square test. The value of the test statistic is 11.56, which is significant at a 1 per cent level (critical point: 6.64). Therefore, we can reject the null hypothesis that the correlation of passage/failure and the presence of an incident is due to chance. However, the size of the chi-square statistic is essentially due to the fact that data screening trapped 8 incidents correctly, when, if screening were to be totally ineffective, we would have seen a number closer to 2.6.

One accident and seven (7) train blockages constitute the eight (8) reports which failed the screening test and had incidents. Of the 357 reports which passed data screening, 191 reports noted incidents related to accidents and train blockages. The impact of these incidents was clearly not great enough to be flagged by the screening procedure.

The 102 reports where the screening test failed but where no incidents had been found include the following.

1. 26 reports failed the travel time reasonableness test. All of them correspond to travel times over 55 mph, some of them over 100 mph. These reports are apparently due to a probe reporting malfunction.

2. 41 reports failed the reasonableness of congested distance test and indeed correspond to congested distances longer than the link length. Like the previous items, these reports are also most likely due to a probe reporting malfunction:

3. 28 reports failed the consistency test. Due to the high variance of the traffic data, this level of test failure (approx. 0.2%) is considered acceptable even if all of these reports are correct. This acceptance is due to the fact that the data screening algorithm is designed to accept the majority, but not necessarily all, of the correct reports.

5.3 Actual Detector Data

The results of the analysis of the detector data are presented in the following table.

	Passed Screening	Failed Screening	Total
No incidents	1794	152	1946
Incidents	7	0	7
Total	1801	152	1953

The presence of incidents in this table was as in the log files kept by the drivers. The relatively high percent of failed reports (about 8%) may be due to the random factors in the measurement and the random nature of traffic. This observation suggests that some data screening conditions for detectors may be too restrictive. However, this statement assumes that all non-incident data are expected to be correct, and should not fail screening. As noted above, we lack the criteria needed to validate this assumption.

Using a similar methodology to that in Section 5.2, and assuming that no relationship exists between screening and incidents, the table of expectations would have slightly different values (values have been rounded to whole numbers to make the interpretation easier).

	Passed Screening	Failed Screening	Total
No incidents	1795	151	1946
Incidents	6	1	7
Total	1801	152	1953

No formal test is needed to show that this table and the previous one have nearly identical values, indicating that no relationship exists between the results of screening and incidents.

6 Conclusions

The first part of the evaluation process using the simulated data indicates that the data screening algorithm is implemented as designed.

The analysis of the results of the data screening for the actual data leads to the following conclusions. As mentioned in Section 5.2, most of the time, when a report failed data-screening and no incident was present, we apparently had a probe reporting malfunction. Also as noted there, some reports which failed data screening had unrealistically high speeds or congested distances larger than the link lengths,

Thus, we reach the conclusion that data screening is rather effective in finding probe reporting malfunctions. This use was itself the purpose of data screening, as prompted by the early-stage technical quality of the MNA's. As such, it did a reasonably good job. While our assessment of the efficacy of the data screening procedure remains somewhat mixed (especially for the detector data), we find the procedure should continue to be used, largely because most of the probe observations which fail the test should indeed be removed from the data base. While it does not do a perfect job, use of the data screening process does improve the data base.

The additional benefit from the evaluation of the data screening algorithm is the verification of the algorithm's implementation (as performed in the stage of the task using simulated data). The error of the implementation revealed by the screening process is described in the Appendix together with suggestions for fine-tuning the algorithm.

References

Berka, S., X. Tian and A. Tarko (1995) Data Fusion Algorithm for ADVANCE Release 2.0, ADVANCE Working Paper Series, No. 48, Urban Transportation Center, University of Illinois, Chicago.

Berka, S. and X. Tian (1995) Changes in Parameters of Data Fusion Subcomponent, Technical Memo, Urban Transportation Center, University of Illinois, Chicago.

Appendix I: Corrections and Suggested Changes

1. According to the design, the volume used in the testing conditions should be a lane volume and not the total approach volume. Division of the volume by number of mid-block lanes from the Attribute Database should be added.

2. The testing criteria for screening detector data consistency could be relaxed.

3. The testing criterion for the maximum probe travel time could be made more restrictive.