Evaluation of Roadside Surveys for Freight Transport Data at Border Crossings

Wirach Hirun [#1]

[#] Department of Civil and Environmental Engineering,
Kasetsart University Chalermprakiat Sakon Nakhon Province Campus,
59 Moo 1 Chiangkear subdistrict, Muang district, Sakon Nakhon province, Thailand.
[1] wirach@hotmail.com

Abstract — **Roadside freight surveys are one of the main types of survey used to observe road freight transport. These surveys require many staff, resulting in high costs, limited survey periods, and variable sampling rates. As a result, data may be incomplete and/or unreliable. This paper presents an examination of roadside survey data of freight transport at a border crossing in Thailand to determine the best day in the week for gathering data and the appropriate sampling rate. The data were collected at the border between Thailand and Laos at Friendship Bridge Number Three in the north-eastern past of Thailand using the face-to-face interview method for all trucks at the outbound border of Thailand during open hours of the border for four weeks. The objective of this analysis is to determine the best day for survey data and the optimum sampling rate. The results indicate that Wednesday's survey data can represent all data better than the other days of the week, both in terms of the number of trucks passing the survey point and the captured weight of freight. Of the remaining days, the best representative day is Thursday. A low sampling rate (20%) has the ability to provide an estimated total weight of freight passing through the survey point which comes close to the actual weight. However, the estimated freight weight is an aggregate from many origins and thus cannot provide the details of transport data. The freight weight classified by freight origin analysis revealed that many origins were disappearing from the captured data of low sampling rate. To solve this problem, the survey needs a high sampling rate for low volume traffic, perhaps as high as 90%.**

Keywords: Roadside surveys, Freight, Origin destination, Freight transport, Border crossing transport

I. INTRODUCTION

Roadside freight surveys are one of the main types of survey used to observe road freight transport, and to collect shipment data by face-to-face interviewing vehicle drivers at points along the transportation route. They are typically used to capture data about the characteristics of the current trip: specifically, the trip origin and destination, goods carried (and their characteristics), and vehicle type [1]. These surveys normally involve working with police or the appropriate law enforcement agency to pull over moving vehicles/drivers and interview them at the roadside about their current trip. They also can be conducted at off-road locations such as weigh stations or inspection areas. The major disadvantages of the survey are the expense of high staffing requirements and the need for other agency involvement (e.g., law enforcement) [2]. On the other hand, the advantages of roadside surveys are high response rates and information about trip purpose, goods carried, origin/destination, weight of goods, and route of transport.

Due to the heavy staffing requirement and the resulting high cost of these surveys, survey periods are limited. A number of studies have done one-day surveys in selected weeks. Furthermore, the sampling rates of the one-day data have ranged from 5-80 percent of truck traffic. For this reason, the survey may capture only a portion of the data and be unreliable. Moreover, the best day of the week for a survey, differences between one-day data and all-week data, and the relationship between errors in the data and the sampling rate have not been explored. This paper presents an examination of the roadside survey data of a freight transport crossing at a border in Thailand. Freight weight, number of trucks, and number of freight origins will be analysed. Then the best day of the week for gathering data and appropriate sampling rate will be explored.

II. LITERATURE REVIEW

The roadside survey method was used to collect traffic data from many studies, but there were differences in survey periods and sampling methods.

The Georgia Department of Transportation (GDOT) conducted a series of roadside origin-destination surveys at eight weigh stations along interstate highways with high truck volumes. A total of 3,636 trucks were sampled, which represented between 4.0% and 15.8% of the daily truck volumes [3].

The Strategic Freight Transportation Analysis (SFTA) project collected the data during a four-week period for each season (spring (April 2002), summer (July 2002), fall (October 2002), winter (January 2003)) to overcome seasonal variation. They suggested that data would be collected for 24 hours and would be collected on the Wednesday of each week to avoid unusual traffic flow patterns at the beginning and end of the week.

They interviewed 60%-80% of the trucks passing through the station during its open hours at lower volume sites while the sites with a higher volume saw between 5%-20% of the total trucks being surveyed [4].

The 2007 Interprovincial Roadside Truck Survey in the National Capital Region (NCR) suggested that roadside intercept surveys be conducted at a minimum every season (four times a year) ideally, over a one-week period for 24 hours per day at each site to account for seasonal variation, day-of-the-week variation, and time-of-day variation, and they strongly recommended that intercept surveys be conducted at a minimum of one weekday and one weekend day to account for day-of-the-week variations as truck flows are typically lower on the weekends than on weekdays. However, due to budget limitations, they collected data only on Wednesdays. The survey collected 1,410 samples of interprovincial heavy truck trips at Macdonald-Cartier Bridge and Chaudiere Bridge, which represent 38% of all truck traffic passing through the survey site, during a 24-hour survey period [5].

The Washington State Freight Truck Origin and Destination Study was designed to provide a profile of state-wide truck movements during each season. A total of 25 sites were selected for 24-hour roadside interviews within as short of a time frame possible during each season. They collected data on Wednesdays to obtain median traffic patterns rather than exceptionally heavy Monday or Friday flows. A systematic sampling strategy was developed for the survey: one out of every 10 trucks on heavy traffic routes, one out of every five trucks on medium traffic routes, and one out of every two commercial vehicles on low volume routes. A total of approximately 7,000 truck drivers were interviewed during each of the four seasonal surveys [6].

The Edmonton Roadside Truck Survey project launched a roadside survey to conduct truck trip data at 14 separate locations around the city over 14 days between Tuesday to Friday and between 9 a.m. and 3 p.m. A total of 2,294 trucks were surveyed [7].

Tennessee Metropolitan Planning Organizations (MPO) launched a study to collect truck origin-destination (O-D) survey data at 13 roadside weigh stations and inspection stations in Tennessee. The data were collected from December 7 through December 23, 2009 at various hours throughout the day (7 a.m. to 7 p.m.) and night (7:00 p.m. to 7:00 a.m.) and a total of 2,312 trucks were interviewed [8].

The Tulare County Association of Governments (TCAG) sponsored an origin and destination truck study in Tulare County. Two rest areas and three truck stops were selected for data collection. They suggested that the best time to conduct the in-person survey in the summer season was from 6 a.m. to 6 p.m. on a Monday and a Tuesday as those periods were observed to have the highest truck volumes along the selected route in Tulare County. The second in-person truck survey, representing the fall season, was conducted from 7 a.m. to 3 p.m. on five days in October. They gathered 329 completed surveys with a response rate of roughly 14% for the summer season survey period and gathered 417 completed surveys with a response rate of roughly 35% for the fall period [9].

The New Jersey Department of Transportation commissioned a face-to-face interview survey of large trucks traveling along US Route 206 in Hillsborough, Somerset County, New Jersey, and along US Route 202/NJ Route 31 in West Amwell, Hunterdon County, New Jersey. A one-day survey from 7 a.m. to 5 p.m. was conducted on US Route 206. A total of 93 northbound US Route 206 tractor-trailer trucks were stopped and surveyed and 4 trucks bypassed the site and were not surveyed, while 193 southbound US Route 206 tractor-trailer trucks were stopped and surveyed and 22 trucks were not surveyed. The survey of US Route 202/NJ Route 31 was conducted on November 8th, 2007 from 6:45 a.m. to 4:45 p.m. and the captured percentage of hourly truck volumes in the survey ranged from 37% to 82% [10].

Southern California Association of Government launched a project to study transportation data at the California – Baja California border region. Truck intercepts were conducted at the Calexico East POE on both sides of the border, capturing northbound and southbound flows of goods transported by truck. Intercept surveys for northbound trips were performed between 8:25 a.m. and 5:48 p.m., while surveys were collected between 7:45 a.m. and 7:22 p.m. on southbound trips [11].

III. METHODOLOGY

A. Data Collection

The border crossing between Thailand and Laos at Friendship Bridge Number Three in the north-eastern part of Thailand is called a new trade route since it’s the shortest route to Laos, Vietnam, and China. It was selected for data collection. The data were collected using the face-to-face interview method for all trucks at the outbound border of Thailand during the open hours of the border. The interview survey gathered vehicle type, freight weight, freight type, origin, destination, and time of interview. The data were collected by four trained surveyors during October 2014.

B. Data Analysis

Data from field surveys served as inputs for analysis. The data analysis process was divided into three parts: data reduction, examining for day of survey, and examining for sampling rate. For the data reduction part, the four-week data was reduced to one-week data and one-day data. The data of each record was numbered for next part analysis.

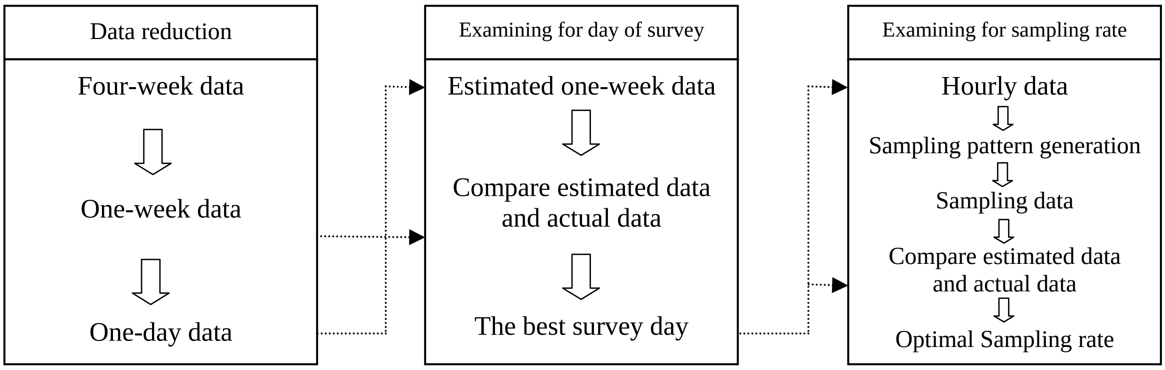

Fig.1. Study method

To calculate estimated one-week data, the one-day data was expanded to represent one-week data by means of an expansion factor, which is multiplying by 7. For examining the sampling rate part, the data of the best survey day was reduced to hourly data and the expansion factor for expanding hourly data to all-day data was calculated as follows:

$$Expansion\_Factor = \frac{Volume}{OD\_sample}$$

where Volume is total number of trucks passing through survey point

OD_sample is number of selected interviewed trucks

To explore the best day for survey, a general comparison of estimated data against actual data which included truck volume and freight weight was initiated. The Root Mean Square Error (RMSE) was employed for statistics testing as given below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - e_i)^2} \qquad \text{For } i = 1 \text{ to } n$$

where e_i is the estimated value

a_i is the actual value from survey

n is the number of data points

To explore optimal sampling rate, the estimated data was compared to actual data. This task was divided into two steps. The first step was comparison between freight weight of sampling data and actual data. The second step was origin destination matrix comparison. In this step, the all-day trip matrix and sample matrix of sampling rates of 3, 5, 10, 20, 30, 40, 50, 60, 70, 80 and 90 percent were constructed. Mean Absolute Percent Error (MAPE) was employed to consider the difference between sampling rate and actual data as given below.

$$MAPE = \frac{1}{n}\left(\sum_{i=1}^{n}\left|\frac{a_i - e_i}{a_i}\right|\right)x100 \qquad \text{For } i = 1 \text{ to } n$$

where e_i is the estimated value in each cell

a_i is the actual value in each cell

n is the number of data points

IV. Results

A. Data description

The details from four weeks of collected data for freight transport at the border crossing are summarized in Table 1. Total truck volume passing from Thailand to Laos is 2,107 trucks within the four weeks of the survey period. Truck volumes in weeks 1, 2 and 4 are slightly different while the volume in week 2 is lower than the others. Likewise, freight weights in weeks 1, 2 and 4 are not quite different while the weight of freight in week 2 is lower than the others. The total weight of freight passing the border is 38,663.70 tons. The freight passing from Thailand to Laos which was captured in the four-week survey period was from 21 origins (province) in Thailand. The data of each week captured at the origin is less than 15 origins, which is quite lower than the four-week data.

TABLE I. Description of Survey Data

Variable	Week	Data
Total Truck Volume (Veh.)	Week 1	549
	Week 2	460
	Week 3	561
	Week 4	537
	All weeks	2,107
Total Weight of Freight (Tons)	Week 1	10,313.09
	Week 2	8,267.42
	Week 3	10,545.15
	Week 4	9,538.042
	All weeks	38,663.70
Number of Origin	Week 1	9
	Week 2	13
	Week 3	14
	Week 4	13
	All weeks	21

B. Examining for effect of day of survey

1) Examining for effect of day of survey and estimated truck volume

Comparisons of truck volume and day of survey are shown in Figure 2. Truck volume on Monday is much higher than the others while Sunday shows the lowest volume. The truck volumes of Wednesday, Thursday and Friday are slightly different. The Root Mean Square Errors of estimated one-week truck volume for all days are shown in Table 2. Wednesday achieves the lowest value, followed by Thursday. For this reason, Wednesday provides the data which can represent all data better than any other day of the week.

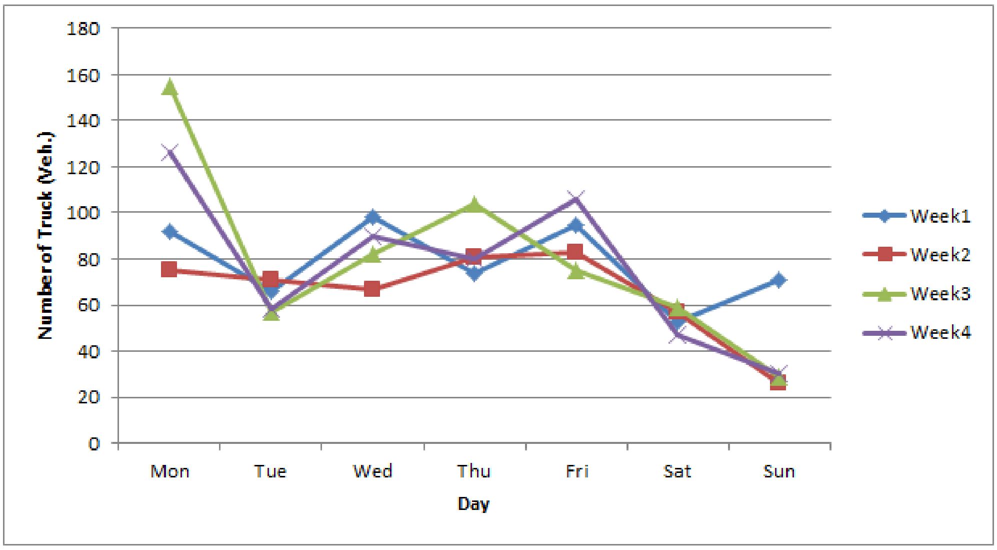

Fig. 2. Truck volume versus day of survey

TABLE II. Day of Survey and Root Mean Square Error of Estimated Truck Volume

Day of Survey	Root Mean Square Error
Monday	149,714.92
Tuesday	16,143.20
Wednesday	10,333.39
Thursday	15,083.76
Friday	23,255.12
Saturday	29,022.75
Sunday	91,978.19

2) Examining for day of survey and freight weight

Comparisons of freight weight and day of survey are shown in Figure 3. The captured freight weight on Monday is much higher than the others while Sunday shows the lowest volume. The freight weights of Wednesday, Thursday and Friday are slightly different. The Root Mean Square Errors of estimated one-week freight weight for all days are shown in Table 3. Wednesday achieves the lowest value followed by Thursday. For this reason, Wednesday provides the data which can represent all data better than the other days of the week. Thus Wednesday data is prepared for the next step analysis.

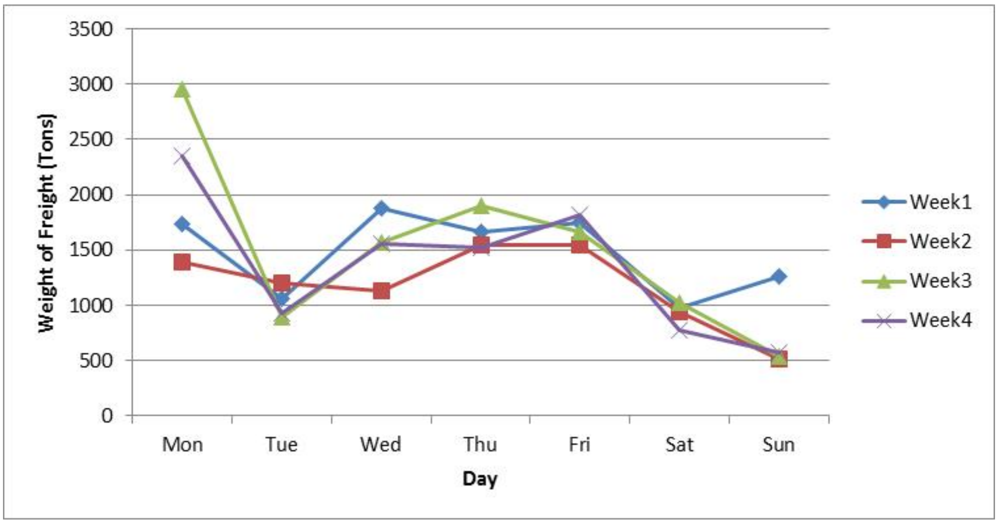

Fig. 3. Weight of freight versus day of survey

TABLE III. Day of Survey and Root Mean Square Error of Estimated Freight Weight

Day of Survey	Root Mean Square Error
Monday	56,663,266.48
Tuesday	10,948,783.12
Wednesday	4,062,695.37
Thursday	5,131,979.13
Friday	6,322,312.62
Saturday	11,856,702.79
Sunday	29,967,730.95

C. Examining of Sampling Rate

Comparisons between estimated freight weight of sampling data and actual data are shown in Figure 4. There is more fluctuation in the estimated freight weight which was captured at the survey point with a small sampling rate (less than 20%). The estimated data was close to actual data with a sampling rate of more than 50%. Based on the Mean Absolute Percent Error in Table 4, a sampling rate of 20% achieves an error rate of approximately 5% and the errors slightly decrease when sampling rates significantly increase. For this reason, a sampling rate of 20% (one out of every five trucks) is suitable for collecting data with an error rate of approximately 5%. The lower sampling rate results in reducing staff and cost for collecting data. However, the estimated freight is the weight of freight that passed the survey point, aggregated from many origins. Thus, this weight cannot provide details of transport data or produce origin destination matrices.

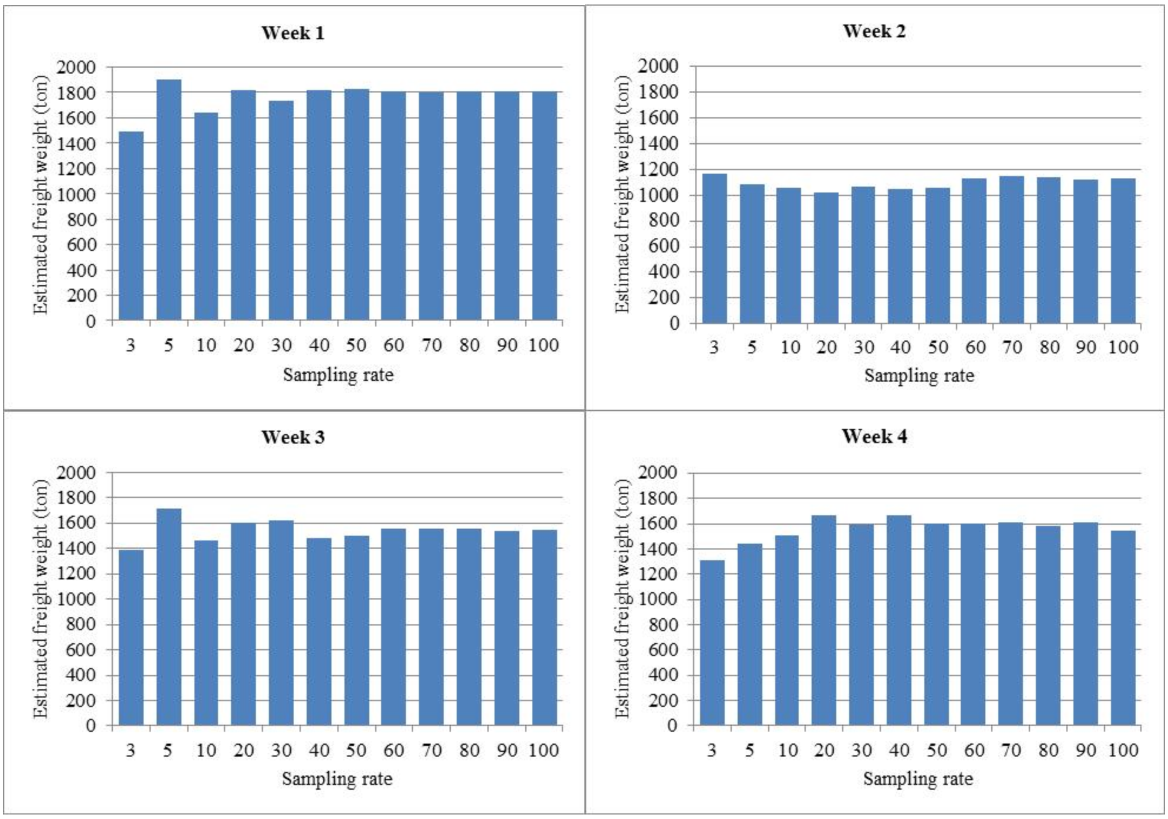

Fig. 4 Estimated freight weight versus sampling rate

The freight weight classified by freight origin is shown in Figure 5. The colour of each cell represents the difference of estimated freight weight and actual freight weight. A green colour is positive (actual data more than estimated data), red colour means negative (estimated data more than actual data), and blue colour is data that was not captured by the survey. Many origins disappeared from the captured data in each week as shown in the picture, especially week 1 and week 2 data, while the total estimated freight weight was not significantly different. Lower sampling rates led to more disappearances in freight origin.

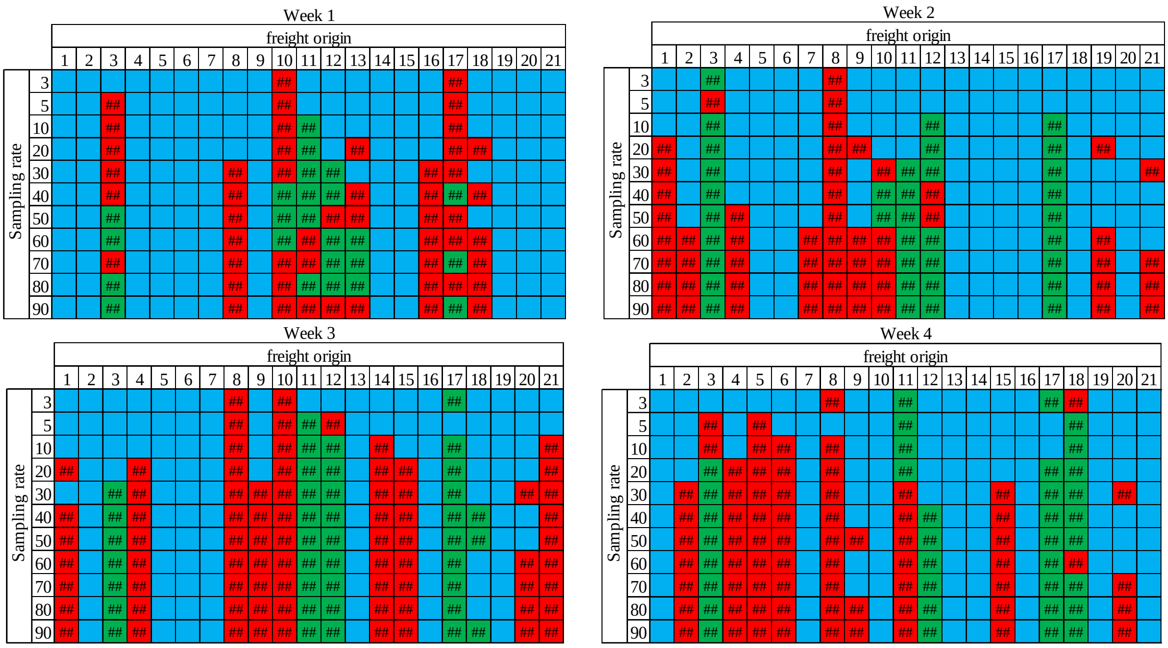

Fig. 5 freight origin versus sampling rate

The Mean Absolute Percent Error in Table 4 shows that more error occurs in all sampling rates, especially low sampling rates (less than 20%). Since the Mean Absolute Percent Error was calculated by subtracting the sample matrix from the actual data matrix, all origins in the actual matrix had the sample matrix subtracted, even those with empty cells in the sample matrix. In cases of low sampling rates, many origins disappear from captured data, leading to many empty cells and a high Mean Absolute Percent Error. This data will thus be missing if expanded data is used to produce origin destination matrices.

TABLE IV. Mean Absolute Percent Error of Data and Sampling Rate

Sampling rate	**Mean Absolute Percent Error**	
	Freight weight	**Freight weight by origin**
3	11.47	130.52
5	6.56	151.20
10	5.60	133.72
20	5.15	108.66
30	4.28	89.53
40	4.87	65.61
50	3.49	48.76
60	1.17	38.44
70	1.88	30.75
80	1.22	19.64
90	1.44	11.39

V. DISCUSSION

Examining the effects of survey days revealed that Wednesday samples provide the best sample data, as suggested by many studies [4] [5] [6], because Wednesday obtains median traffic patterns rather than exceptionally heavy Monday or Friday flows. Of the remaining days, the best alternative is Thursday. For examining sampling rate, the results indicated that the expansion factor was successful in producing estimated total weight of freight passing through the survey point which is close to actual data even with a low sampling rate (20%). In contrast, the freight weight classified by freight origin analysis revealed that a high Mean Absolute Percent Error occurred for all sampling rates. With a low sampling rate, many origins disappear from the captured data, making the data unreliable for producing origin destination matrices. The solution to this problem is a high sampling rate for low traffic volume, perhaps as much as 90%. In surveys done with low traffic volume, freight origin data may be limited.

VI. CONCLUSION

Roadside freight surveys are one of the main types of survey used to observe road freight transport. Due to the heavy staffing requirement and the resulting high cost of these surveys, survey periods are limited. Furthermore, sampling rates have widely varied among the existing research. For this reason, a survey may capture partial data and be unreliable. This paper presents an examination of roadside survey data of freight transport at a border crossing in Thailand to explore the best day of the week for gathering data and the appropriate sampling rate.

The data was collected at a border crossing between Thailand and Laos at Friendship Bridge Number Three in the north-eastern past of Thailand using the face-to-face interview method for all trucks at the outbound border of Thailand during the border's open hours. The analysis focuses on exploring the best day of survey data and the optimum sampling rate.

The results indicated that Wednesday's survey data can represent all data better than the other days of the week, both in terms of the number of trucks passing the survey point and the captured weight of freight. Of the remaining days, the best representative day is Thursday. A low sampling rate (20%) has the ability to provide an estimated total weight of freight passing through the survey point which comes close to the actual weight. However, the estimated freight weight is an aggregate from many origins and thus cannot provide the details of transport data. The freight weight classified by freight origin analysis revealed that many origins were disappearing from the captured data of low sampling rate. To solve this problem, the survey needs a high sampling rate for low volume traffic, perhaps as high as 90%.

REFERENCES

[1] Allen, J. and M. Browne, “Review of Survey Techniques Used in Urban Freight Studies,” University of Westminster, London, U.K., 2008.
[2] Jessup, E., K.L. Casavant, and C.T. Lawson, “Truck Trip Data Collection Methods,” Oregon Department of Transportation and Federal Highway Administration, Salem, 2004.
[3] National Cooperative Highway Research Program, “Freight Transportation Surveys: A Synthesis of Highway Practice”, Transportation Research Board, Washington D.C., 2011.
[4] Prozzi, J., Wong C. and R. Harrison, “Texas Truck Data Collection Guidebook,” Centre for Transportation Research, The University of Texas at Austin, Texas, 2004. [www.utexas.edu/research/ctr/pdf_reports/0_4713_P4.pdf]
[5] TRANS Committee, “2007 Interprovincial Roadside Truck Survey Summary of Results”, 2011.[http://www.ncr-trans-rcn.ca/wp-content/uploads/2013/03/InterprovincialRoadsideTruckSurvey-finalreport.pdf]
[6] William R. G. and K.L. Casavant, “Washington State Freight Truck Origin and Destination Study: Methods, Procedures, and Data Dictionary”, Washington State University, WA, 1994.[http://ewits.wsu.edu/reports/research/err3.pdf]
[7] The City of Edmonton, “Roadside Truck Survey”, Transportation Planning, City of Edmonton, Alberta, Canada. [www.edmonton.ca]
[8] Tennessee Department of Transportation, “Tennessee Roadside Truck Origin-Destination Surveys”, Tennessee, 2010.[http://www.tdot.state.tn.us/longrange/freightplan/Task5-DR2-TNDOT-Roadside-Truck-Surveys.pdf]
[9] Tulare County Association of Governments, “SR 99/SR 198 Gateways Truck Origin and Destination Study”, Idaho, 2015.[http://www. http://tularecog.org/wp-content/uploads/2015/07/SR99-SR198-Gateways-Survey-Results-2015.pdf]
[10] New Jersey Department of Transportation, “Large Truck Origin/Destination Survey”, New Jersey, 2009.[http://www.nj.gov/transportation/freight/plan/pdf/TruckSurvey09.pdf]
[11] Southern California Association of Government, “Goods Movement Border Crossing Study and Analysis”, 2012.[http://www.freightworks.org/DocumentLibrary/]