AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

CTRF

Transportation: Emerging Realities

Les Transports: realités en puissance

VOLUME 2

Canadian Transportation Research Forum
Proceedings of the 32nd Annual Conference
Toronto, Ontario May 25-28, 1997

Le Groupe de Recherches sur les Transports au Canada
Actes de la 32ième conférence annuelle
Toronto, Ontario 25 au 28 mai, 1997

AN EXPERIENCE IN MODELLING COMMERCIAL VEHICLE SURVEYS

Robert J. Tardif
Ministry of Transportation, Ontario *

Don L. McLeish
Department of Statistics & Actuarial Science
University of Waterloo

ABSTRACT

The Ministry of Transportation of Ontario conducted an inter-city roadside survey of commercial vehicles over a period of 167 days in 1995 at a cost of approximately $600,000. In total, 31,860 surveys were conducted at 68 sites throughout Ontario. Through a combination of; weighting each axle, observations and interviews, information was collected within five major areas; vehicle, trip, commodity, carrier and driver.

On completion of previous commercial vehicle surveys, a series of expansion factors were created using the counts collected at the time of survey. These factors merely expanded the survey sample to represent the total truck population observed at each survey station for the same period. Additional factors were then generated using broader time periods, daily, seasonal and annual. These were generally created in isolation of a host of potential problems, primarily related to bias in sampling, over counting in route assignment and without regard to trip travel time on the highway network. The data was analyzed using crude expansion factors. As a result, many of the estimates, even percentages, were inaccurate.

Realizing that flaws existed in the previous survey expansion methodology, the ministry embarked on a review of the survey design and analysis methodologies. The primary purpose of this paper is to explain how the survey expansion process was improved by adding additional trip routing questions and identifying the stations passed for each trip.

* The views expressed are those of the author and not necessarily those of the ministry.

Once additional routing information was collected, four Data Enhancement factors were created, they are described in greater detail within. The first two factors, accounting for station specific hour and daily truck population, are consistent with previous expansion methods. The remaining factors, remove sampling bias and over counting which hold the key to matching the count data for all survey locations involved, especially when all surveys are used.

The factors were evaluated within our modelling framework, where the results proved that no degree of error exists in the Data Enhancement process. However, deficiencies with the model were highlighted. The model is now the weak link in the process with an overall accuracy rate of 92% for representation of truck volumes at each station. The focus now turns from calibrating the Data Enhancement process to calibration of the model for future applications.

INTRODUCTION

In an effort to update and increase our knowledge of the trucking industry, the Ministry of Transportation of Ontario completed a roadside survey of commercial vehicles. The information collected from roadside surveys is viewed as being the most comprehensive and reliable method to obtain provincial wide information on trucking activity trip behaviour. These data are required for a variety of purposes, primarily for policy and planning initiatives. In total, 31,860 surveys were completed over 167 days in 1995. Prior to '95, as part of this survey program, a similar survey was conducted in 1993 and another of similar scale in 1988. In '95 surveys were primarily completed at permanent Truck Inspection Stations (97%) with the remaining 3% completed at temporary sites at strategic highway locations.

To initiate the survey, trucks were alerted by flashing lights and requested at random to enter the inspection station. Each of the axles were weighed and visual observations of the vehicle characteristics recorded. The driver was then signalled to bring the truck into the compound where the verbal portion of the survey took place; this was completed in approximately six minutes. Equipped with a PenDOS computer, the interviewer asked the driver a series of questions related to characteristics of driver, employment, trip, commodity and vehicle.

Of particular interest for this paper are the details surrounding the trip characteristics. The drivers were asked a series of detailed trip routing information, including: vehicle base, commodity origin, trip origin, border crossing on entry to Ontario, last stop prior to taking part in the survey, survey location, next stop immediately after the survey, exit border crossing, trip destination and commodity destination. The computer system used for the

survey contained an extensive database of placenames within North America. Along with the placename the corresponding longitude and latitude coordinates and a predefined zone system were simultaneously added to the survey record. If the trip was not a line haul (direct trip without pick-ups or drop-offs on route), and was of the peddle-run variety, the first seven stops were added.

This wealth of geographical data provides both challenges and opportunities. This origin-destination (OD) data is essential to accurately represent the survey sample trip routing within the model and most importantly to facilitate a revised Data Enhancement process. The term "Data Enhancement" refers to a series of measures which minimize survey bias, increase statistical reliability and expand the sample.

Given the cost of this massive effort to conduct the surveys and counts, it seemed appropriate to maximize our investment and increase the reliability of the data while extending the application of its use. The inspiration to develop a trucking model was spurred from the requirement to evaluate this relatively untested "Data Enhancement" process.

This Data Enhancement process was activated in the initial planning stages and carries over to the data expansion and model calibration stages. Thus, it was tied to the survey program from start to end. Initially, an assessment of the potential survey sites was undertaken. The selection criteria for optimal survey locations was based on the strategic level of truck information required, historical data characterising truck movements and daily truck count data. On average, seven 24-hour periods of detailed classification counts were collected for each survey site. Similar to the survey collection, detailed classification counts were collected throughout the summer survey period. As much as possible, we attempted to balance sample size for surveys and hours of counts for weekdays, weekends, nights and days.

On completion of field data collection of surveys and counts, a thorough review of the data was initiated. This review was not limited to merely reviewing and correcting obvious errors. In terms of the trip and commodity OD data, PC*MILER® routing software was used to verify the survey data and generate additional details to be added to the survey record. This detail includes: highways and local roads used (identifying the jurisdiction) and the sections used (from intersection entered to exited). Trip distances were then calculated for distance travelled in Ontario and outside of Ontario. The next and perhaps the most useful piece of data constructed was an inventory of the survey stations passed along the entire route. This inventory consists of a string of station

numbers in which this truck was not necessarily surveyed at, but rather the potential existed for the truck to be surveyed. This inference enables development of a trip relationship with all possible stations passed along the trip route. The code in the inventory is referred to as "SSS", the directional station number.

This paper deals with the advantages afforded by the Data Enhancement factors. The base model developed is more a mechanism to graphically present, test and validate results of the Data Enhancement process. Often in our applications, a model is developed only to emulate a provincial perspective from limited amount of OD data.

Counts were applied to the sample, first on an hourly basis then beyond, to day of survey, average day of week over the survey period, weekly, seasonal and annual. For analysis purposes, if only one survey site was involved this approach would be sufficient. However, this is not the case and an alternative process is required.

The goal in this survey was to provide an accurate measurements of truck movements through each of the 68 directional survey sites. In previous truck OD surveys, data were collected from a number of locations and a conventional model used to simulate the movements. Since the basic expansion factors were developed for hour, day and year for each station to represent the truck population at each site, it was considered logical that this would represent province wide commercial vehicle flows. This assumption was found to be false, as illustrated in Figures 1 and 2. Further, the duration of a truck trip on the network was never accounted for. In the past, the only time value used to expand the survey sample was the hourly count at the specific station in which the survey was collected and counted. In this exercise, the duration of the trip on the network and the relationship to the truck passing each of the other survey stations along the entire route is factored in. This is achieved by constructing a travel time and distance matrix for each time iteration. This hourly offset identifies the impact of the trip for each hour previous to and after the survey was conducted for the trip duration.

The two figures below best illustrate the scope of the benefit awarded by the Data Enhancement process. Figure 1, contains the desired flows at each survey station resulting from the Data Enhancement process and Figure 2 dramatically demonstrates the effects of double counting within the survey sample without the Data Enhancement process applied. It is obvious in Figure 2 that the actual truck volumes are misrepresented at all survey sites and that the order of

magnitude varies throughout the province. The surveys may represent trucking activity at each station but when all OD surveys are combined a province wide perspective becomes misrepresented. Figure 2 includes survey bias and over-counting. Previously, this was considered to be an accurate 24 hour snap-shot of provincial trucking activity, obviously it was not.
Figure 1

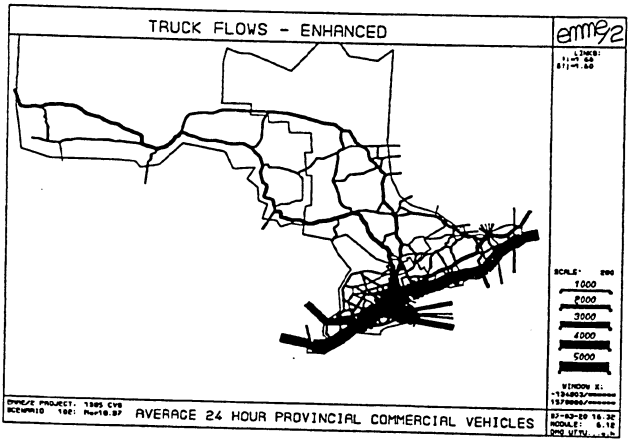
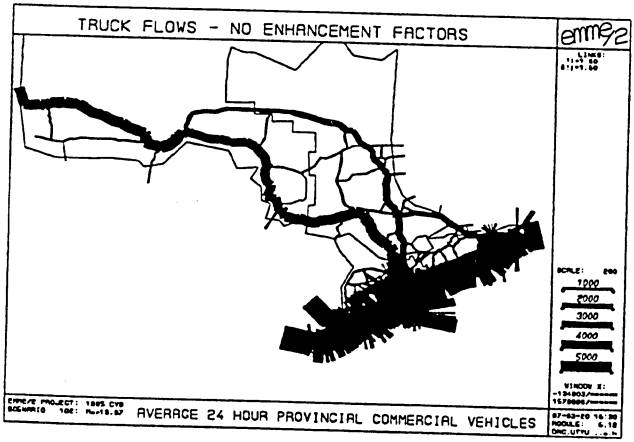


Figure 2



DATA ENHANCEMENT FACTORS

The Data Enhancement process involves the creation of four factors to eliminate bias in sampling and deal with over counting. The desired result is achieved by applying combinations of these factors. Applying sums of the factors to any of the data collected for analysis purposes adjusts the sample proportionally to the truck population. For all applications, including analysis of truck flows in modelling, the factors achieve an assignment value equal to the traffic count at each survey station.

Definition of DE, the Data Expansion Factor

We begin by determining the total truck volume at each station in a week. Let $V(ss,h)$ be the (estimated) truck traffic volume at station/direction combination ss and hour of week (HOW) h where $h = 0, 1, \ldots 167$ and $h = 0$ denotes Sunday, 12:01 a.m. - 1:00 a.m., $h = 167$ denotes Saturday 11:01 p.m. - 12:00. In an ideal world the predefined count schedule would have been carried out flawlessly. Unfortunately this was not the case. On review of the classification counts collected, we refer to these as ground counts, it became obvious that for relatively short time periods, automatic count systems failed or provided unrealistic data at some stations for various hours of the week. In order to overcome these deficiencies, a count model was used to identify errors through statistical analysis and to estimate traffic counts at station-hour combinations where no counts were taken.

In addition to loss of count data during the actual survey period unforseen and uncontrollable circumstances forced cancellations in the preplanned schedule. The causes were varied, ranging from sudden changes in staff schedules, adverse weather, prolonged power outages from lightening strikes at remote sites, and as the summer wore on the exhausted survey crews required additional days off.

The count model approximated the expected counts at all station-hour combinations. Counts at a given station were approximated by regressing on all counts at all stations in the same region, and then smoothing the residuals. Summing over all hours of the week, we obtain total weekly truck traffic volume

$$V(ss)=V(ss,0) + V(ss,1) + \ldots + V(ss,167).$$

Data Expansion Factor: DEF

The purpose of the hourly station data expansion factor DE is to expand all surveys at a given station and given hour of the week to all expected truck traffic at that station/direction/HOW combination. The data expansion factor

for a survey taken at station *ss* and HOW *h* is the ratio of the expected truck volume *V(ss,h)* at that station-HOW to the number of surveys collected at that station-HOW. Thus, if *NS (ss,h)* = total number of surveys collected at station *ss*, in
HOW = *h*, we define the data expansion factor for a record at (*ss,h*) to be

$$DE(ss,h) = V(ss,h) / N(ss,h).$$

In cases where the traffic counts were unavailable or unreliable, there was an additional possible source of information about counts available from the surveys themselves. In particular, if there are 10 surveys in a given HOW at a given station-direction, then it is reasonable to assume that the expected truck traffic volume at that station is greater than or equal to 10 (even though actual count data on some other week or estimated counts may have been less than 10). Thus, we ensure that all factors are greater than or equal to 1 (since the average volume of traffic is assumed at least equal to the number of surveys). Therefore, replace DE above by:

$$DE = min(1,DE) \text{ and } V(ss,h), NS(ss,h))$$

Definition of The Station Sampling Fraction

Let *V(ss)* be the total weekly truck volume at station *ss*. We define the weekly sampling fraction as the *proportion of weekly traffic sampled.* We take as an example one of the stations with the least surveying, station *ss* = 134. We recorded the following number of surveys by hour of week:

Hour of Week	Number of Surveys	Expected Truck Count
106	3	13
107	1	13
111	3	13
112	1	12
113	1	11

Adding the expected truck count over all 168 hours of the week we obtain the weekly count 1,582. These nine surveys need to be expanded over the whole weekly traffic. The weekly station sampling fraction is the ratio

$$SF(134) = 9/1582 = .005689$$

In general:

$$SF(ss) = \frac{\sum_{h=0}^{167} N(ss,h)}{V(ss)}$$

When a full week of surveys was not obtained, resulting in only a partial weeks coverage, an adjustment factor was created and applied to expand the surveyed hours to all 168 HOW. For example, at station 134, the factor DE expands to the surveyed hour, but there were only 5 of the 168 possible HOW in which surveys were taken. Thus, to expand to the weekly traffic, we need to multiply by the ratio of the expected traffic counts for the whole week to those in the hours surveyed. To expand to an average day, we must divide this again by 7, the number of days in a week.

$$DEF = 1582/7*(13+13+13+12+11) = 1582/7*62=25.516/7=3.645$$

Definition of OCF, The Overcount Correction Factor

The purpose of this factor is to adjust for the bias caused by the probability that a given trip could have been sampled at two or more locations *when records at all survey stations are combined.* Thus, although a given truck may be surveyed only once, if it passes several survey stations, *trips of this type* have a higher probability of appearing in the sample than do trips that pass a single station. Suppose, for example, a trip passes two stations, number 105 and 167, with sampling fraction at station 105 of 0.066 and at station 167 of 0.064. Then roughly .066/(.066+.064) =50.8% of these surveyed trip types will be surveyed at station 105, and 49.2% at station 167. For a survey conducted at station 105 with these properties, 50.8% is the OCF. Moreover, a trip type through two or more survey stations has a higher probability of being sampled in the survey as a whole than does one that only passes a single survey station. The OCF is defined for a given survey at station *ss*. Assume this trip type passes stations *sspass* on this trip. Then we define OCF by

$$OCF = \frac{SF(ss)}{\sum_{i \text{ in } sspass} SF(i)}$$

Notice that this factor only depends on the relative sampling fractions at the various stations, not on their magnitude. In other words, it is unaffected by computing sampling fractions on daily, weekly, monthly periods, or indeed computing them for the whole survey period.

Example: Record number 26814. Survey conducted at station 105. *sspass* = 105 167. There were 18 surveys conducted at this station in *HOW* = 69. The expected truck count at *HOW* = 69 is 147, so *DE* = 147/18 = 8.17. The sampling fraction at station 105 is 0.066 and at station 167 is 0.064. Therefore

$$OCF = 0.066/(.066 + .064) = 0.5077$$

Definition of OCFR, The Regional Overcount Correction Factor

The purpose of this factor is to adjust for the bias caused by the possibility that a given trip could have been sampled at two or more locations *when records within the corresponding region only are combined.* The justification for the factor is similar to that for OCF, only it assumes that for information at a given station say in region 2, we only wish to combine records surveyed in the same region 2.

The regions were defined by *ss* number as follows:

1. Southwest	102 103 109 111 112 117 142 148 149 150 151 158 165 166
2. Central	101 104 105 118 119 131 132 137 139 140 161 162 167
3. Eastern	100 108 115 116 120 121 122 123 125 126 135 136 138 141 163 164
4. Northern	106 107 110 124 127 128 133 134 143 144 147 152 153 159 160
5. Northwest	113 114 129 130 145 146 154 155 156 167 113 114 129 130 145 146

For example, for a station *ss* in the central region

$$OCFR = \frac{SF(ss)}{\sum_{i \text{ in sspass and } i \text{ in central}} SF(i)}$$

Example: Record number 8, station *ss* = 101, *sspass* = 138 121 101 132 149 151. This station is in the central region. Those stations in *sspass* that are also in the central region are stations 101 and 132. Their sampling fractions are respectively, 0.0406 and 0.0421. Therefore

$$OCFR = .0406/(.0406 + .0421) = .4909.$$

Summary of Factors

DE: used to expand all records at a given location/direction/hour to all traffic passing that station in the hour

DEF: expands all traffic in all hours surveyed to average daily traffic (multiply by 7 to get average weekly traffic)

OCF: overcount correction factor. Use when combining all surveys (in all regions) that have a given property P and have the given *ss* number anywhere in sspass field.

OCFR: overcount correction factor. Use when combining all surveys in a single region that have a given property P.

Truly, this survey was a departure from our common OD survey methodology. This extraordinary detail related to the routing opens many opportunities to accurately eliminate survey bias and network wide trip overcounting. Detail characteristics include, total trips, specific commodity flows, vehicle characteristics, driver characteristics, and just about any of the approximately 200 variables. As a result of the Data Enhancement process all trips represent a 24 hour period and most importantly, are systematically calibrated at each and every survey station.

DATA ENHANCEMENT VERIFICATION\MODEL CALIBRATION

The resulting OD matrix total after applying the Data Enhancement process, inclusive of the OCF stage expands the sample to represent approximately 51,000 truck trips for a 24 hour period. Once this matrix is assigned through the model onto the provincial road network with external centriod connectors, the combined volume at each station amounts to 91,000 trips. This is due to trips that pass more than one survey station and on the other hand, total trips are reduced as a result of the coarse zonal definitions and resulting inter-zonal trips.

As discussed earlier, the factors related to the Data Enhancement process are attached to every record. Using a combination of factors for the desired time reference any combination of flows can be realistically represented. The same process can be tested within the model. To verify this statement, provided in Table 1, are model results using Select Link analysis for each survey station and results generated from the survey data. In both cases the Data Enhancement factors have been applied.

Initially, the trip matrix was prepared through a typical approach, with each OD pair summed and weighted by the data enhancement factors. This matrix was then assigned to the provincial network as free-flows. Assignment results were compared against the 24 hour truck volumes at each station, border crossings and transportation boundary. For ease of repeating calibration procedures, survey stations, border crossings and transportation boundaries were tagged as screen lines. A screen line is defined as a series of points on a road network forming a cordon or net around a particular area of interest. For analysis purposes, this cordon is often consistent with existing political boundaries or other similar set of features.

Select Link analysis was performed to verify the routings at all screen lines identified. In addition, assignment volumes were checked against actual volumes. In general, initial model results were within 18%, but this was not satisfactory. When these system wide results were displayed graphically and compared to the border crossing movements, it became obvious that changes in the matrix preparation were necessary. In order to accurately represent border crossings and station flows, it was decided to prepare the matrix as if in two stages. In the first stage, a matrix was constructed from only intra-provincial flows. This was accomplished by coding Ontario zones between 1 and 60 and using only OD surveys with both origin and destination zones between 1 and 60. This leaves out any trips which, originate outside Ontario, and\or have destinations outside Ontario. For these trips, all border crossings on entry and exit were given unique zone numbers as gateways. Numbers ranged from 69 to 97. Regardless of differing origin zone numbers, as would be the case with a trip from a Western Province or California, but they share the same border crossing, the trip origin was coded as the crossing. This is an excellent method since a number of potential border crossing may exist for like OD pairs and now the trip matrix exactly replicates the each surveys routing, regardless of the assignment technique. Storing this revised OD zoning in a new field permits us to build matrices specifically for the model.

This matrix was then assigned in the same manner as in the first attempt. Verification of results was completed from screen line reports for the three types of screen lines. Overall, model results were improved to within 8% of ground counts or within 92% of the actual daily truck volumes for 68 sites throughout Ontario. Meanwhile, testing the Data Enhancement process based on the new model zone structure, but not running through the model, continues to reach the level of accuracy in matching ground counts at each survey station with 100%.

Unique to this methodology, from the survey database possesses the ability to construct a trip matrix for any station or combination of stations. The result of the Data Enhancement process ("Enhanced Survey"), will produce total flows identical to the 24 hour truck traffic counts ("Ground Counts") for each station. To verify these results I present the following table:

Table 1 - Comparison Between Counts\Data Enhancement\Model

STATION NAME	DIRECTION	GROUND COUNTS	ENHANCED SURVEY	ENHANCED MODEL	% DIF D\E	REAL DIFFERENCE
FORT ERIE	West	1,418	1,418	1,310	1.08	108
WINONA	East	4,100	4,100	4,495	0.91	(395)
FRUITLAND	West	3,934	3,934	4,628	0.85	(694)
OAKVILLE	East & West	11,254	11,254	10,969	1.03	285
THAMESFORD	East	432	432	52	8.31	380
CHATHAM	West	498	498	253	1.97	245
TILSONBURG	West	328	328	169	1.94	159
MORRISTON	North	934	934	285	3.28	649
FERGUS	South	249	249	238	1.05	11
GLEN TAY	East & West	734	734	938	0.78	(204)
NEW HAMBURG	East	370	370	437	0.85	(67)
VICTORIA	South	719	719	474	1.52	245
GRAVENHURST	South	829	829	649	1.28	180
WASI NORTH	North & South	1,056	1,056	847	1.25	209
NEW LISKEARD	North & South	830	830	573	1.45	257
COCHRANE	West	243	243	217	1.12	26
HEARST	East & West	494	494	368	1.34	126
RED ROCK	East & West	889	889	871	1.02	18
MORTON	South	201	201	160	1.26	41
KEMPTVILLE	North & South	690	690	263	2.62	427
HALEY STATION	East & West	870	870	439	1.98	431
NORTH BAY	West	422	422	103	4.10	319
NORTH SHORE	East & West	1,042	1,042	1,028	1.01	14
HEYDEN/SSM	North & South	507	507	612	0.83	(105)
THUNDER BAY	East & West	520	520	0	0	520
DRYDEN	East & West	957	957	793	1.21	164
WINCHESTER	North & South	278	278	12	23.1	266
DOVER CENTRE	North	279	279	113	2.47	166
ELDORADO	North & South	165	165	23	7.17	142
KEARNS	East & West	423	423	29	14.5	394
TIMMINS	East & West	366	366	255	1.44	111
THUNDER BAY	East & West	1,169	1,169	862	1.36	307
WOODBRIDGE	South	2,754	2,754	2,664	1.03	90
KING	North	2,571	2,571	2,728	0.94	(157)
CURRY HILL	East	2,181	2,181	2,111	1.03	70
LANCASTER	West	2,459	2,459	2,386	1.03	73
GANANOQUE	East & West	5,682	5,682	4,820	1.17	842
WHITBY	East	4,413	4,413	4,258	1.04	155
BOWMANVILLE	West	3,880	3,880	3,571	1.09	309
TRAFALGAR	East &West	12,361	12,361	12,855	0.96	(494)
PUTNAM	East & West	10,409	10,409	9,550	1.09	859
WINDSOR	East & West	5,779	5,779	5,875	0.98	(96)
SARNIA	East & West	3,538	3,538	3,342	1.06	196
BRANTFORD	East & West	4,156	4,156	3,497	1.19	659
CASSELMAN	West	743	743	685	1.08	58
					1.08	

The table above contains stations for which no volume in the Enhanced Model column while the Enhanced Surveys column provides the actual number of trucks that were recorded at that station. This is due to a combination of reasons, ie: competing parallel links where all the volume is assigned to the optimal route or the zone system in the matrix and network are too course thus incapable of replicating the real simulation. The fact remains, the Data Enhancement process provides better results than the model can produce.

However, it should be noted that the Data Enhancement process may provide excellent representation of the base situation, but its ability to simulate future year scenarios does not exist. Nor can changes in the actual road network the day after the survey be accurately reflected. In addition, the network in the model contains a vast number of highways not included in the survey. For these highways and highway sections between survey points the model provides us with a relative understanding of travel. Only with a model can we track the value of goods transported or set priorities for maintenance and capital planning or graphically identify strategic corridors.

Commodity Flow Validation

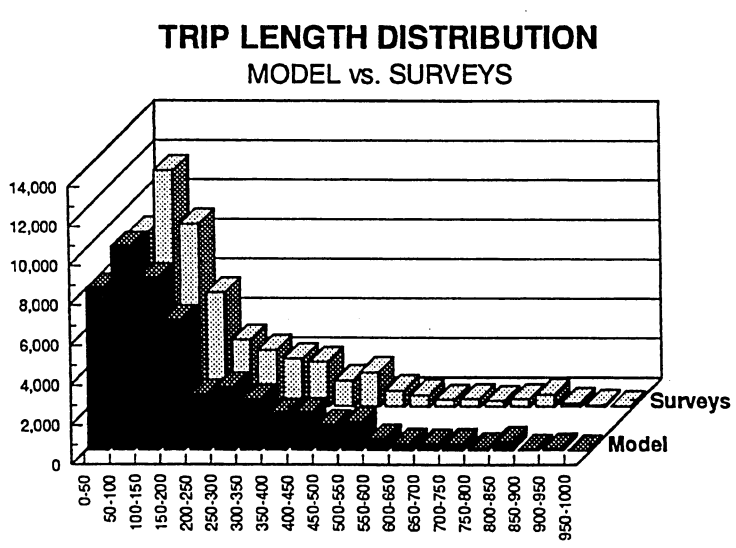
Table 2 - Comparison -Data Enhancement & Model Results - Food Commodity

STATION NAME	DIRECTION	ENHANCED SURVEY	ENHANCED MODEL	% DIF ES\EM	ABSOLUTE DIFFERENCE
FORT ERIE	West	109	92	1.18	17
WINONA	East	511	656	.77	(145)
FRUITLAND	West	635	745	.85	(110)
OAKVILLE	East	937	834	1.12	103
OAKVILLE	West	845	744	1.14	101
DRYDEN	West	103	93	1.11	10
THUNDER BAY N.	West	126	96	1.31	30
WOODBRIDGE	South	279	265	1.05	14
KING	North	372	388	.96	(16)
CURRY HILL	East	393	387	1.02	6
LANCASTER	West	279	274	1.02	5
GANANOQUE	West	487	429	1.14	58
GANANOQUE	East	364	313	1.16	51
WHITBY	East	501	473	1.06	28
BOWMANVILLE	West	884	839	1.05	45
TRAFALGAR	East	1047	1135	.92	(88)
TRAFALGAR	West	903	961	.94	(58)
PUTNAM	West	817	802	1.02	15
PUTNAM	East	729	710	1.03	19
WINDSOR	East	366	421	.87	(55)
WINDSOR	West	314	359	.87	(45)
SARNIA	West	139	165	.84	(26)
SARNIA	East	148	163	.91	(15)
BRANTFORD	East	354	250	1.42	104
BRANTFORD	West	229	213	1.08	16
CASSELMAN	West	112	101	1.11	11
				1.006	

We have established that the Data Enhancement process accurately matches ground counts. The next step, given that we require the model to estimate commodity related activity, is to perform a test with commodity data. The following table provides these results, a comparison of Model results and Survey data for food related products. In preparation of this table, trips carrying food related commodities were filtered into a separate database. Using the data subset, two parallel streams of comparison were undertaken.

The number of trips passing each station was generated from the Survey data with enhancement factors applied are provided in the column "Enhanced Survey". For comparison, a matrix was constructed from the same subset of data, also with the enhancement factors applied and assigned in the model in an "all-or-nothing" assignment, results are provided in column "Enhanced Model".

Another test was conducted to verify travel characteristics related to trip length distribution (TLD). The chart below illustrates and confirms the validity of the Data Enhancement process by checking against the TLD from the model. Given that in general the model results contain a degree of error estimated at 8%, these results are well within that range. This confirms that the model is producing logical results, especially when we consider that peddle trips are included, peddle trips account for 5% of all trips and have numerous stops and illogical routing in relation to the common line-haul OD pairs. In addition, the "all-or-nothing" assignment technique will not perfectly replicate routing diversion of congested urban areas. Even without these two explanations for possible error, the difference generated is less than 1%.



The accuracy of the results provided from the Data Enhancement process are not surprising, nor is the fact that model results are not as reliable. There are a number of explanations for this. The actual survey data contains much more information than can be easily replicated in the model. The model uses only a fraction of the routing information collected to construct the trip matrices.

The nature of the model contains a number of potential errors in terms of: no trip generation model, zonal aggregation problems causing excess intra-zonal trips, precision or simulation of volume delay functions, no distribution model or cell and trip end recalibration, not mode sensitive, auto and truck OD can be combined in one database, not sensitive to trip assignment techniques, DE process precalculates travel time and distance between all OD pairs, adjusting for not only at time of survey flow on network but potential impact on network over time. For all of these reasons how can a model better represent the base situation when the volumes at each survey station are matched.

Benefits of the Data Enhancement Process Are Passed to the Model

- Select link analysis was not conducted on a zonal basis but rather for each survey site and for a series of select links.
- In calibration the matrix was known prior to the assignment stage, double counting was eliminated in the Data Enhancement process.
- Building of trip matrices was not a simple process of summing OD pairs and applying weights
- Border crossings become the entry point regardless of origin zone, this forces all known trips through their respective border crossing rather than allowing an equilibrium assignment or gravity model distort the actual flows
- Data Enhancement factors remove double counting basis for all records at each station.

In conventional models with OD data, the model itself is calibrated as an end product producing an OD matrix resembling flows. Without internal select link analysis only trip ends are calibrated. The information collected in the '95 CVS is not limited to OD and routing information. In fact OD questions relate to only 10% of the survey. In a conventional model the calibration achieved remains within the model and is not automatically related back into the flat data file. Since the Data Enhancement process remains tied to the flat data file, any data element enjoys the same benefits. In addition, each data element becomes tied to the trip. In the conventional model this relationship is very difficult and cumbersome to obtain. In addition, most OD surveys contain only origin and destination.

The additional trip information collected allows creation of multiple trip segment matrices, ie: trip origin to border-crossing, border crossing to survey station. A combination of 12 OD matrices can be constructed while maintaining the flexibility to apply the Data Enhancement factors. The resulting combined matrix more closely replicates each segment of the trip. Assigning this matrix produces a improved association with "U" Trips, those trips which are of shorter duration and not captured in this inter-city survey.

FUTURE SURVEY PROGRAM AND MODEL DEVELOPMENT

The ability exists to identify each stations relative strategic importance while using existing base year information to pre-test the sensitivity of policies and capital expenditures. Thus, the survey program can now be streamlined.

In terms of model development, given that the truck volumes are accurately represented at many screen-lines. It is now desirable to examine economic relationships. These involves developing a model to represent the value-of-goods transported while matching with Import\Export values, GDP and GNP trucking activity indices.

ACKNOWLEDGEMENTS

The authors wish to thank Michele Babin of the Freight and Economic Research Office, MTO for her contributions.

BIBLIOGRAPHY

1. Don L. McLeish. 1994. *Report on The 1993 Pilot Commercial Vehicle Study*

2. Don L. McLeish. 1997. *Data Enhancement Process Documentation For The 1995 Ontario Commercial Vehicle Survey.*