Sketch Model for Estimating Station Level Ridership for LRT

Mausam Duggal, Manager, Regional Planning, Transit and Urban Mobility. WSP|PB
Nemanja Radakovic, Project Associate. WSP|PB
Arun Bhowmick, General Supervisor, City of Edmonton
Sandeep Datla, Senior Engineer, City of Edmonton

Paper prepared for presentation at the "WHAT ARE TRANSIT-SUPPORTIVE ENVIRONMENTS, ANYWAY? INNOVATIVE INDICATORS AND METHODOLOGIES TO EVALUATE TRANSIT SUPPORTIVENESS" Session

Of the 2016 Conference of the
Transportation Association of Canada
Toronto, ON

ABSTRACT

The City of Edmonton aims to undertake an evidence-based approach to prioritize its 2047 unfunded LRT network for implementation. The Regional Travel Model (RTM) developed by the City does not explicitly model LRT as a separate mode and in general, like other regional-scaled models, is limited in its ability to accurately reflect station-level patterns. On the other hand, a Direct Ridership Model (DRM) estimates ridership as a function of the station's environment and transit service attributes. A DRM can be used as complementary to the traditional four-step models like the RTM to provide a better understanding at a local level, which a system-wide analysis cannot provide. It also provides significantly greater flexibility than a RTM for evaluating different land use and service scenarios, and with minimal calibration, is transferable to other major urban areas. In this study, we developed a DRM of LRT stations using data collected for 15 existing stations in the City of Edmonton straddling over two horizon years. Individual DRMs were developed for the a.m. and p.m. peak periods after an assessment of the available data. The results indicate that population, employment, and post-secondary enrollment within a 1000m buffer around each station; along with, bus frequency, distance to CBD, parking spaces, and CBD stations were statistically significant variables. Five methods to expand the peak period station level boarding to daily estimates were also developed to understand the potential range of daily forecasts the City could expect to experience in the future.

INTRODUCTION

In 2015, the City of Edmonton undertook a prioritization study to identify the timing and business case for each of their proposed Light Rail Transit (LRT) lines that are expected to be built by 2047. Like other major urban areas in North America, the City has a strategic four-stage travel demand model (RTM) that it has used for estimating line level ridership. However, strategic models warrant a certain level of generalization on the supply, demand, and behavior-side of the population to result in a tractable solution. Strategic models can also be less flexible, and significantly more time consuming, when it comes to testing multiple permutation and combinations between LRT lines and their individual extents. Recognizing the above challenges, along with certain limitations (absence of LRT as an explicit travel mode) of the current RTM, the City of Edmonton structured the study in two distinct parts. First, a "sketch planning" tool (DRM) was developed for estimating boardings as a function of station-level and transit-service variables for studying prioritization scenarios. Second, running the selected prioritization scenario through the RTM to develop a business case.

This paper focuses on the development and structure of the DRM. The resulting model from this study is expected to be a first step in developing inexpensive ridership estimation tools at the station level. The relatively small dataset that was available for model development presents unique opportunities for future enhancements as data from other cities in Canada is added to the estimation model.

BACKGROUND

An initial literature review highlighted the fact that there were no known cases of DRMs in use in Canada. While in the US, DRMs have emerged as a low cost, quick alternative to traditional four-stage demand models to forecast ridership at the station or corridor level. A comprehensive evaluation of DRMs in the US is beyond the scope of this paper and hence the reader is pointed to the following two publications - "*How to Increase Rail Ridership in Maryland? Direct Ridership Models for Policy Guidance*", and "*Sketch Model to Forecast Heavy-rail Ridership*".

Analysis done so far has shown that DRMs are especially sensitive to three categories of independent variables; built environment characteristics of neighborhoods (population and employment density, distance to the Central Business District (CBD), etc.); socioeconomic factors (household income, auto ownership, etc.); and transit service attributes (feeder bus frequency, parking spaces, station type, etc.). While DRMs are extremely flexible, and can be sensitive in capturing intangible benefits such as the contribution of Transit Oriented Development (TOD) design elements to ridership (Cervero and Murakami 2008), they also present challenges. These challenges arise primarily due to the small sample sizes of observations, which limit the number of explanatory variables that can be used in model estimation along with their specification. Further, DRMs do not account for travel time and cost differentials between auto and transit modes, which are very critical in developing accurate mode choice forecasts. Thus DRMs are complementary to conventional travel demand models and well suited for producing order-of-magnitude estimates for quickly evaluating localized smart growth scenarios.

METHODOLOGY

A multivariate linear regression model was developed to forecast ridership using data collected by the City of Edmonton. The model was estimated using the R programming language, which is an environment for statistical computing and visualization. ArcGIS was used as the primary spatial analysis tool to display and calculate input variables necessary for model estimation.

Initial Data

Figure 1 shows the existing and proposed LRT lines in the City of Edmonton. In 2012, there were a total of 15 stations along the two lines. Given the paucity of data points, it was decided to also add the 2009 station boardings to the dataset, for a total of 30 observations.

The City of Edmonton's RTM was the primary source of demographic (population and employment) and network (number of feeder buses, etc.) information for both 2009 and 2012. A number of other variables such as distance to CBD, parking spaces, identifiers for CBD vs TOD vs University stations, and whether a station was a terminal (end-of-line) were generated after discussions with City staff. The population and employment within a

Figure 1. City of Edmonton's Existing and Proposed LRT Network
Source: (City of Edmonton 2012)

1000 meter buffer from each station was estimated using ArcGIS's spatial analysis tools. Given the proximity of the stations, there was significant overlap between the 1000-meter buffers, which led to a significant overestimation of population and employment within the respective buffers. This was overcome by undertaking the following steps:

- Merging all the buffers using the UNION spatial operation in ArcGIS;
- Tag all the traffic analysis zones (TAZs) from the RTM that intersect the above union buffer;
- Assign the population and employment within each of the intersected TAZs to the nearest station.

Table 1 shows the candidate of variables that were initially considered for data collection.

Table 1. Candidate Variables for the City of Edmonton's Direct Ridership Model

Variable Categories	Variable List
LRT service attributes	• Extent of service • Daily number of LRT services in both directions • Daily hours of service • Daily number of perpendicular feeder buses and bus lines within 100 m of station • Number of transfer opportunities to other higher order transit • Number of parallel and competing mode choices i.e. BRT, LRT
Location and neighborhood attributes (within 800 m of station)	• Population and employment density within 1000 metres of station • Different age and employment stratifications within 1000 metres of station • Street connectivity index (number of intersections divided by number of links) • Mixed land use potential • School and University enrollment
Station/Site attributes	• Park-and-ride lot (0-1) • Number of parking spaces • Whether station is a terminal (end-of-line) stop (0-1) • Whether station is part of a mobility hub and/or shopping mall

DATA TRANSFORMATION

All of the input variables were plotted against the AM and PM peak period boardings, a sample of which is shown in Figure 2. This was done to document any non-linear relationships and understand the potential transformations needed to either the dependent value (boardings) or the independent values (selected variables). The analysis indicated that the boardings, employment, and distance to CBD variables could be transformed depending upon the time period being modeled. Tukey and Mosteller's Bulging Rule, shown in Figure 3, was used as a guideline in transforming a non-linear relationship to a linear relationship. These transformations have been documented in Table 3.

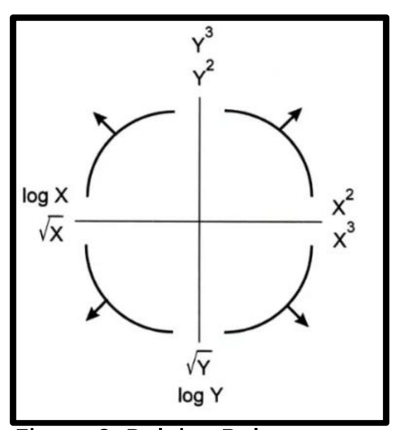

Figure 3. Bulging Rule transformation guidelines
Source: (Charpentier 2014)

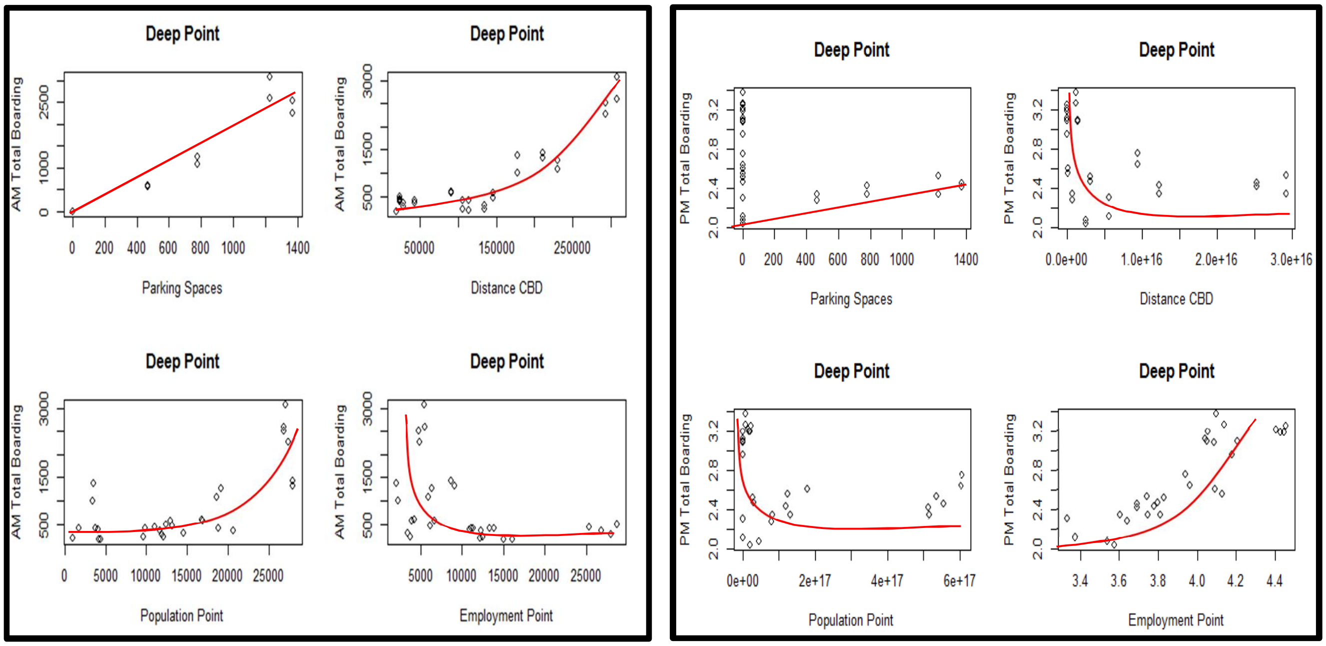

Figure 2. A Sample of Boardings vs Input Variables Scatter Plots – AM and PM Peak Periods

Final Data

After careful evaluation of the available data, observing its transformations, discussions with City staff, and an initial set of model runs, the list of variables was filtered down to those presented in Table 2, along with their descriptive statistics. The initial set of model runs was based on the philosophy of step-wise regression (backward selection process), where we start with all possible variables (from Table 1) in the model, and we iteratively remove variables that are statistically insignificant (small t-stat value). This process continues until all the remaining variables are statistically significant.

Table 2. Descriptive Statistics – Final Variables for the City of Edmonton's Direct Ridership Model

Final Variable	Minimum	Maximum	Average
Population	999	27,280	13,255
Employment	2,149	28,706	10,650
Distance to CBD (kms)*	21,733	300,327	129,714
Post Secondary Enrollment	0	27,350	3,043
AM Bus Frequency	0	119	60
PM Bus Frequency	0	122	61
Parking Spaces	0	1372	256
CBD**	0	1	NA
Non-CBD	0	1	N/A

*distance is calculated from each station zone to all the zones that make up a CBD
**dummy variables

MODEL ESTIMATION RESULTS

The final DRM is shown below in Table 3. Ordinary least squares (OLS) regression in the R programming language was used to estimate the results shown in Table 3. Of note, an attempt was also made to directly estimate total daily boardings as a function of the input variables. This however did not result in a statistically sound model, and was further compounded by the fact that the coefficient signs were opposite to what one would expect. Thus, the subsequent model estimation efforts were focused on the AM and PM peak period DRMs only.

The collection of significant variables in the two models is generally in line with expectations. Bus frequency was significant in both the DRMs, with the other variables playing a role in one or the other time periods. The bus frequency for the PM DRM proved to be significant for the non-CBD stations only.

The PM DRM has a very low constant, which is generally preferred as it represents the unexplained part of a regression model. However the AM DRM was also found to be statistically robust with the correct coefficient signs.

Table 3. Final Multivariate DRM – City of Edmonton

Final Variable	Transformation	Coefficient	T-stat
AM Peak Period Model			
Boardings (Dependent variable)	-	-	-
Constant	-	-336.3	-1.85
Population	Pop ^ 3	2.673*10 ^ -11	1.747 (sig at 0.1)
Distance to CBD (kms)	DCBD ^ 0.5	1.676	2.856
Parking Spaces	-	0.811	4.279
AM Bus Frequency	-	4.792	2.86
PM Peak Period Model			
Boardings (Dependent variable)	Log base 10	-	-
Constant	-	-0.748	-1.298
Post Secondary Enrollment	-	1.858*10 ^ -5	3.709
Employment	Log base 10	0.825	5.457
PM Bus Freq * Non-CBD Station		1.815*10 ^-3	2.043
CBD Station (dummy)		0.303	2.945

Figure 4 shows the modeled versus observed boardings by time period. The left set of graph compares the observed counts to the modeled counts with the coloured dots representing GEH values (a green dot indicates GEH less than 5, orange dot is GEH between 5 and 10 and the red dot is a GEH value higher than 10). It is quite clear that the predictions from the DRMs are strong and positive. The r-square values in both cases are very high at 0.88 and 0.76 for the AM and PM peak period forecasts, respectively.

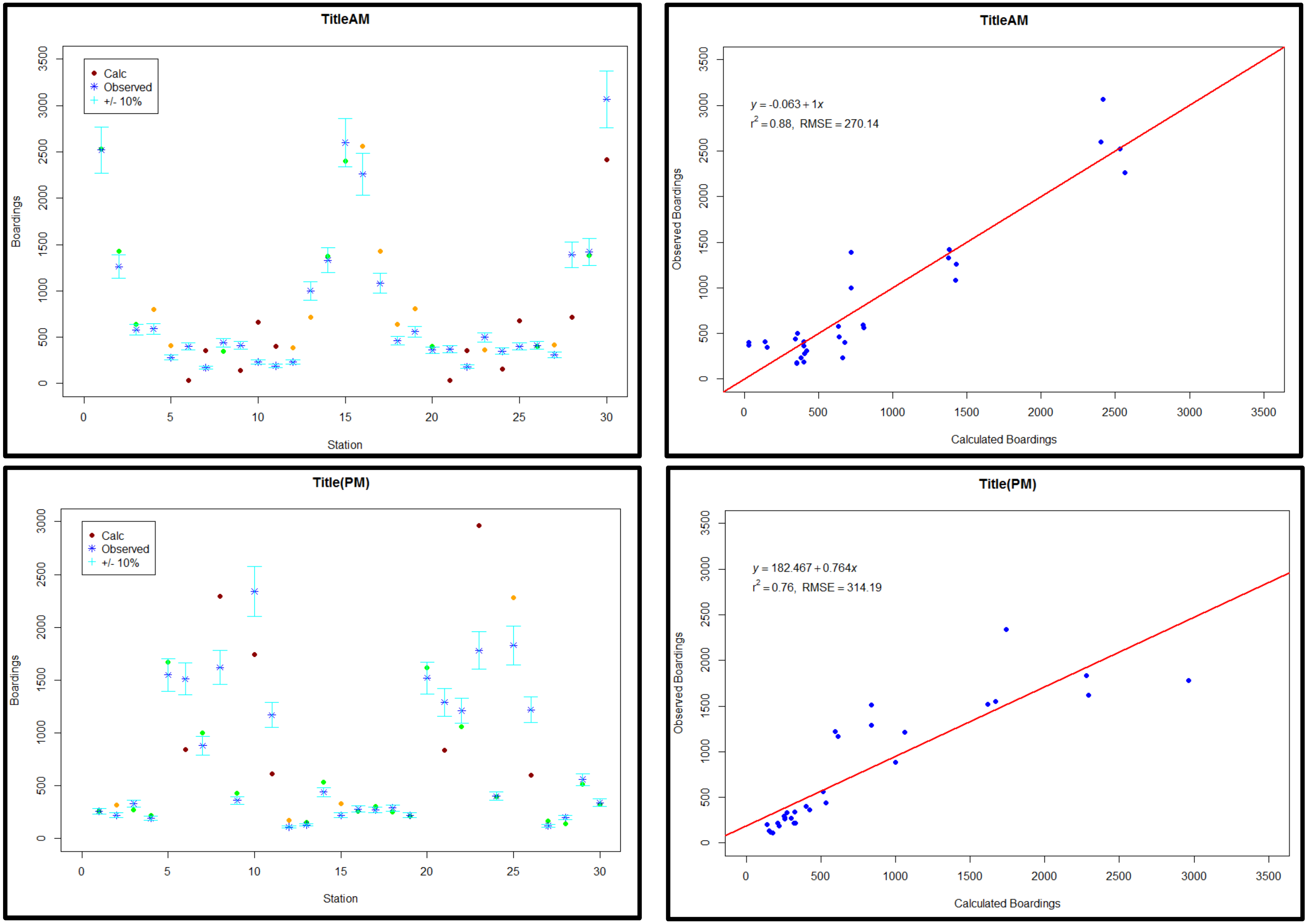

Figure 4. Modeled vs Observed Boardings – AM and PM Peak Period DRMs

DAILY EXPANSION

Due to the difficulties of directly estimate total daily boardings as a function of the input variables, a process for expanding AM and PM peak period boardings to the daily level was developed as part of the study. The details of the expansion procedure are not documented here because it is deemed to be specific to the City of Edmonton and as such does not allow it to be generalized to other cities. However, for the sake of completeness, we note that five expansion procedures were developed by classifying the stations into CBD, non-CBD, TOD, or University stations, which then allowed the team to estimate daily expansion factors, as shown in Table 4.

Table 4. Tested Daily Expansion Factors – City of Edmonton

Scenario	City Zone	Formula (X = AM+PM Boardings)
1	Non-CBD Station	Y = 3.329 * X
	CBD Station	Y = 3.779 * X
	University Stations	Y = 5.702 * X
2	TOD Stations	Y = 3.865 * X
	CBD Stations	Y = 3.779 * X
	Other Stations	Y = 4.043 * X
3	CBD Stations	Y = 3.779 * X
	Other Stations (Excluding University of Alberta Station)	Y = 3.393 * X
	University of Alberta Station	Y = 6.327 * X (Average of the expansion Factors)
4A	CBD + TOD Stations (Including University of Alberta Station)	Y = 3.84 * X
	Other Stations	Y = 4.043 * X
4B	CBD + TOD Stations (Excluding University of Alberta Station)	Y = 3.462 * X
	University of Alberta Stations	Y = 6.327 * X (Average of the expansion Factors)
	Other Stations	Y = 4.043 * X

CONCLUSION

The DRMs developed as part of the Edmonton LRT prioritization project have shown that such sketch planning methods can provide a complementary set of inexpensive methods to predict potential station level ridership. The DRMs cannot completely replace the City of Edmonton's RTM, but it fulfills the role of a flexible and fast tool for scenario evaluation. They should be thought of as a tool that helps filter down the range of observations a RTM must test in detail.

During the course of the development of these DRMs, an exhaustive list of variables was studied and some were found to be significant through the estimation work. However, additional variables can be evaluated and added to improve model fit and strengthen its predictive prowess. More importantly, these DRMs should be considered as a first generation of such tools, which present a unique opportunity to subsequently add data points from other major urban centers in Canada to expand the estimation dataset.

References:

Cervero, Robert, and Jin Murakami. "Rail and Property Development in Hong Kong: Experiences, Impacts, and Extensions." *Urban Studies*, May 2008: 2019-2043.

Charpentier, Arthur. "Tukey and Mosteller's Bulging Rule (and Ladder of Powers)." *R-bloggers.* June 16, 2014. http://www.r-bloggers.com/tukey-and-mostellers-bulging-rule-and-ladder-of-powers/.

City of Edmonton. "Long-Term LRT Expansion." March 2012. http://www.edmonton.ca/transportation/PDF/Long_Term_LRT_Network_Plan_March_2012.pdf.