Factors Impacting Link Travel Speed Reliability: A Case Study at Cairo, Egypt

Akram Sabry
Nile University, Giza, Egypt
Email: akramsabry@hotmail.com

Hoda Talaat
Faculty of Engineering, Cairo University, Giza, Egypt
Email: hoda.talaat@gmail.com

Abstract—Day-to-day and within-day fluctuations in traffic patterns in many traffic systems negatively impact the system reliability. Travel time reliability, which considers the probability that a trip is completed within a given travel-time range, is an important performance measure in travel-related policy formulation. This research is focused on the assessment of factors impacting travel speed reliability. The study investigates the impact of specific factors (namely; land-use, road-class, congestion level, and time period) on travel speed reliability using probe data in Greater Cairo Region, Egypt. The GPS data was collected using a tailored Smart Mobile phone application. Testing results revealed the significant impact of all four factors on travel speed reliability, given appropriate disaggregation of factors' levels. The relationship between travel speed reliability (dependent variable) and identified impacting factors (predictors) is then formalized in a multiple linear regression model. The developed model returned an average root mean square error of 8.8%.

Index Terms—travel reliability, travel speed, linear regression, land use, probe vehicles, traffic survey

I. INTRODUCTION

Travelers appreciate travel time reliability; a consistency or dependability in travel times as measured from day-to-day. Drivers want to know that a trip will take half an hour today, half an hour tomorrow, and so on. As such, travel time reliability plays a crucial role in travelers’ route choice behavior. Measures of travel time reliability represent commuters’ experiences better than a simple average travel time. Most reliability measures compare excessive delay days to those with average delay.

The day-to-day fluctuation in traffic patterns of complex urban roadway networks challenges travel time prediction accuracy. The degree of fluctuation is expected to vary according to a set of spatial/temporal network characteristics (such as time of day and roadway segment class). Understanding and modeling the impact of specific factors in travel time/speed reliability, is hence, crucial for operational as well as planning activities.

This study investigates the impact of specific factors (namely; land-use, road-class, congestion level, and time period) on travel speed reliability. The study aims to develop a systematic model to estimate link travel-speed reliability, given specific temporal/spatial characteristics of the traffic network. The conducted analysis is based on data collected at a case study route in Cairo, Egypt.

II. BACKGROUND

Over the past decade, several research efforts have been conducted to study the impact of travel time reliability on drivers’ route choices. In a study conducted at UCF, Abdel-Aty has investigated the effect of travel time variability on route choices in the Los Angeles area [1]. Results of the conducted field survey indicated that 54% of respondents stated that travel time reliability was either the most or second most important factor for choosing their daily commute routes. In several similar studies, reliability-related attributes have been reported to be among the most considered service attributes in various travel-related decisions [2, 3, 4]. Lam measured values of average travel time and travel time reliability from 1998 data of the actual behavior of commuters on State Route 91 in Orange County, California. The results suggested that travelers are willing to take tolled lanes to save travel time and reduce travel time uncertainty [5]. Rakha et al. questioned the assumption of normality of path travel times. In their study, they analyzed AVI data from San Antonio, concluding that the assumption of normality is inconsistent with field observations and that a lognormal distribution is more appropriate [6].

On the other hand, Litman examined ways that transportation decisions affect land-use patterns. He found that transportation planning decisions have many direct/indirect land-use impacts. In particular, certain transportation planning decisions tend to increase sprawl while others support smart growth [7]. Another study examined traffic and land-use patterns along Lagos-Badagry corridor, in Lagos metropolis. Results showed that residential, institutional and commercial land-uses generated 28.0%, 20.6% and 18.3% of the traffic observed along the corridor, respectively [8].
The significance of travel time reliability as an important measure of traffic systems’ performance is apparent, based on reviewed literature. Factors, such as land-use and traffic congestion are realized to impact travel time reliability. However, such impact is not formally quantified, and hence, this research has been motivated.

III. RESEARCH OVERVIEW

A primary objective of this research is to identify factors affecting travel time reliability and quantify their impact. Four factors have been identified for the undertaken assessment (namely; land-use, road-class, congestion level, and time period). GPS data points for multiple trips, on a specific case study route, have been collected within a period of four month. The test route is composed of a number of roadway segments with different traffic/layout/land-use characteristics. The variability in travel speeds from trip to another was estimated through a conventional reliability measure; Planning Rate Index (PRI), as will be discussed in details in section 6.2.

Data collection has been conducted in two stages; an early and a late data collection stage. The early data collection stage, aimed at providing insightful observations on the appropriate disaggregation schema of investigated factors. The early data collection stage was followed by an experiment design and a late data collection activity. The impact of identified factors on travel speed reliability was evaluated using a set of ANOVA tests. A regression-based model was, then, formalized to enable the estimation/prediction of the PRI of a specific roadway segment, given the values of the impacting factors.

The impact of the road-class factor could be depicted more apparent, based on reviewed literature. Factors, such as land-use and traffic congestion are realized to impact travel time reliability. A significant portion of the study route lies on the NA corridor, which is a primary arterial connecting New Cairo and Nasr city. It also includes the 90th road, a main corridor in New Cairo, and Nasr road which is a main corridor in Nasr City. Other secondary arterials and collector are included with variant land-use types. The study route is divided into 69 roadway links. These links are divided based on a high degree of homogeneity of land-use, road-class, and geometric layout.

V. EARLY DATA COLLECTION AND ANALYSIS

The early data collection consisted of 14 round trips on the study route; 50% in weekdays and 50% during weekends. Speed/location data was collected using a tailored data collection Mobile application. The developed application records continuous GPS-based location/speed data at predefined time intervals. TROPOS agent-oriented software engineering methodology was adopted for the application development [9]. The system was integrated with a GIS platform for data analysis purposes.

The total collected data points over three months of early data collection were 5455 points, after removing the skewed values. The preliminary data analysis was based on the graphical representation of observed speeds categorized based on three factors; land-use (7 levels; Shopping, Residential high-end, Residential medium standard, Residential/Commercial, Governmental/Institutional, Open Space, and Educational), road class (5 levels; Regional Primary Highway, Urban Expressway, Urban Primary, Urban Secondary, and Collector), and time period (2 levels; peak and off peak). Fig. 2 presents the variation in average travel speeds relative to land-use and time of day variations. An apparent impact of land-use in varying average travel speeds could be depicted. Travel speeds declined significantly within the residential-commercial land-use level in both peak and off-peak periods. On the other hand, the difference between the average speeds in different time periods was rather limited. Accordingly, based on data visualization, the land-use factor was believed to have a main stream effect on traffic operations regardless the time of the day.

The impact of the road-class factor could be depicted in Fig. 3. Low travel speeds were observed in urban expressways and collectors. The differences between the

Figure 1. Case Study Route.
average speeds for different time periods were still extremely limited. In an attempt to better understand the impact of the time period factor, a decision was taken to increase the number of levels of this factor to 8 levels; four within day periods (7:00 to 11:00, 11:00 to 15:00, 15:00 to 19:00, and 19:00 to 24:00) and two day-of-the-week levels (weekday, and weekends).

![Figure 3. Early data – Impact of road class and time period on speed variation](image)

VI. LATE DATA COLLECTION AND ANALYSIS

A. Experiment Design and Data Collection

Additional trips were conducted at the late data collection stage based on an orthogonal fractional factorial experiment design. The design was performed using SPSS software. The experimental design considered three factors as stated below (from 1 to 3). An additional factor is added at the analysis phase to reflect the link average congestion level; the Speed Index (SI) (factor 4). The adopted design brings down the number of combinations to 148 combinations (instead of 280 for full-factorial design). The design guarantees that the variables are not correlated and hence ensure the analysis credibility.

- Land-use; 7 levels (Shopping, Residential High Standard (HL), Residential Medium Standard (ML), Residential-Commercial, Government/Institutional, Open space, and Educational).
- Road class; 6 levels (Regional Primary Highway (2), Urban Expressway (3), Urban Primary (4), Urban Secondary (5), Collector (6), and local (7)).
- Time Period; 8 levels (1 to 4 in weekdays, and 5 to 8 in weekends, periods 1 and 5 start at 7:00am, followed by subsequent periods as discussed in section 6)
- Speed Index (average speed divided by free flow speed); 5 levels (level1 0-0.25, level2 0.25-0.5, level3 0.5-0.75, level4 0.75-1, and level5 1-1.25).

B. Data Analysis

The assessment of the level of impact of the identified four factors in travel speed reliability has been undertaken through one-way Analysis of Variance (ANOVA) testing. The Planning Rate Index (PRI) was selected as a measure of travel speed reliability, as per equation 1. Results of the ANOVA tests shows a significant impact of all four factors, given 95% confidence interval. Interpretation of the results could be better captured through the graphical representation in Fig. 4.

\[
PRI = \frac{95\text{th percentile travel speed}}{\text{Free } – \text{Flow Speed}}
\]  

(1)

The impact of traffic congestion (represented by SI) in travel speed reliability could be clearly depicted form Fig. 3. The PRI increases (enhanced reliability) with the increase in SI (decrease in traffic congestion). The effect of the other factors could also be visually realized. For the time period factor, one could realize that enhanced reliability were attained during weekend mornings (time period 5) and during working-days in the off-peak from 19:00 to 24:00 (time period 4). This means that during these periods, travelers are expected to be able to reasonably budget their trip travel time. Alternatively, during periods (1, 2, 3, 6, 7, and 8) travelers are expected to face challenges to arrive to their destination on time.

The road class factor was proven to have significant impact on travel speed reliability; where level 2 of the road class factor (Regional Primary Highway) returns the best performance. Alternatively, levels 3 and 5 of the road class factor (Urban Expressway & Urban Secondary) witnessed lower performance levels with higher degrees of uncertainty. On the other hand, for the land-use factor, enhanced reliability could be depicted in levels 2 and 6 (residential (HL) and open space) while the other levels showing medium and low performance. While the two levels of Land-uses are different, they both have stable conditions with respect to traffic patterns.

![Figure 4. Variation of PRI with impacting factors levels](image)

VII. REGRESSION ANALYSIS

A. Simple Linear Regression

The development of a regression-based estimation model for the PRI was undertaken through two steps. In the first step, a simple linear regression model was developed; having the PRI as the dependent variable and 21 independent variables (one continuous and 20 binary variables). The SI is the only considered continuous independent variable. Land-use, road-class, and time period factors are treated as binary independent variables,
for each factor level. For example, for the shopping mall group, a binary variable was used which takes the value of 1 if the land-use is a shopping mall and 0 if it’s not.

Using the linear regression function of the SPSS, the obtained model showed an Adjusted R-Square of 0.91 and a Standard Error of the Mean (SEM) of 0.12. However, as depicted in Table I, some variables have negative coefficient (e.g. regional primary highway road-class and time Period 5) that are not logically explained. On the other hand, given a level of confidence of 95%, some variables showed low significance (e.g. educational class and time Period 5) that are not logically explained.

Whereas, as depicted in Table I, some variables have negative coefficient (e.g. regional primary highway road-class and time Period 5) that are not logically explained. On the other hand, given a level of confidence of 95%, some variables showed low significance (e.g. educational class and time Period 5) that are not logically explained.

For each factor level, a binary variable was used which takes the value of 1 if the land-use is a shopping mall and 0 if it’s not.

Using the linear regression function of the SPSS, the obtained model showed an Adjusted R-Square of 0.91 and a Standard Error of the Mean (SEM) of 0.12. However, as depicted in Table I, some variables have negative coefficient (e.g. regional primary highway road-class and time Period 5) that are not logically explained.

On the other hand, given a level of confidence of 95%, some variables showed low significance (e.g. educational class and time Period 5) that are not logically explained.

However, as depicted in Table I, some variables have negative coefficient (e.g. regional primary highway road-class and time Period 5) that are not logically explained.

The model performance is evaluated based on the Root Mean Square Error (RMSE); which represents the difference between the observed PRI on these links and estimated ones from category means. The estimated RMSE for all twenty categories was 0.088. All estimated RMSEs were under 0.2. Such results are perceived to be promising, given the simplicity of the model in terms of information requirement as well as computational ones.

### VIII. Conclusions

In this paper the impact of four factors (namely; land-use, road-class, congestion level, and time period) on travel speed reliability has been investigated using probe data on a case study route in Cairo, Egypt. The test route is composed of a number of roadway segments with different traffic/layout/land-use characteristics. The

\[
Y_n = b_0 + b_1X_n
\]

where, \(Y_n\) is the estimate of PRI value, \(X_n\) is the estimate of the SI value for category \(n\), and \(b_0\) and \(b_1\) are the estimated regression coefficients.

### TABLE II. MULTIPLE LINEAR REGRESSION MODEL OUTPUTS

<table>
<thead>
<tr>
<th>Category</th>
<th>Time Period</th>
<th>Road Class</th>
<th>Land Use</th>
<th>(b_1)</th>
<th>(b_0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1.222</td>
<td>0.0287</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1.69</td>
<td>-0.003</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1.13</td>
<td>0.0572</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1.118</td>
<td>0.2105</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.794</td>
<td>-0.018</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1.119</td>
<td>0.2987</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.485</td>
<td>-0.181</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1.261</td>
<td>-0.047</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1.306</td>
<td>0.0637</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1.564</td>
<td>0.0244</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0.85</td>
<td>0.4702</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1.616</td>
<td>-0.024</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0.432</td>
<td>0.7766</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1.147</td>
<td>0.0348</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>-0.61</td>
<td>1.6648</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.918</td>
<td>-0.067</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1.172</td>
<td>0.0755</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>0.78</td>
<td>0.3792</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1.757</td>
<td>-0.012</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1.148</td>
<td>0.3093</td>
</tr>
</tbody>
</table>

The model performance is evaluated based on the Root Mean Square Error (RMSE); which represents the difference between the observed PRI on these links and estimated ones from category means. The estimated average RMSE for all twenty categories was 0.088. All estimated RMSEs were under 0.2. Such results are perceived to be promising, given the simplicity of the model in terms of information requirement as well as computational ones.

### B. Multiple Linear Regression

A clustering algorithm was adopted to reduce the number of variables; K-Mean Clustering algorithm [10]. Levels with similar impact on the PRI, per factor, have been collected into one clusters. Four clusters have been identified for the time period factor, three clusters for the land-use, and three clusters for the road class. Based on the formulated clusters, multiple regression analysis was conducted to develop an estimation/prediction model for the PRI. The developed model is a linear regression one considering the SI as the only independent variable, as per equation 2. Different coefficients have been identified for each combination of clusters (category), as presented in Table II.

\[
Y_n = b_0 + b_1X_n
\]
Planning Rate Index was used as a measure of travel speed reliability. ANOVA results revealed the significant impact of all four variables, given the identified levels of each factor and 95% confidence interval. The significance of the time period factor was influenced by the level of disaggregation of time intervals. A significant impact was only realized when the number of levels of the time period factor was increased from two (peak and off peak) to eight levels (four in weekdays and four in weekends).

A primary objective of this research was to develop a prediction model for travel speed reliability. Two models were developed for this purpose; a simple linear regression model, and a multiple linear regression one. The first model considered 21 independent variables (one regression model, and a multiple linear regression one. were developed for this purpose; a simple linear prediction model for travel speed reliability. Two models each combination of clusters (20 categories). The variable. Different coefficients have been identified for considering the Speed Index as the only independent variable. The developed model is a multiple linear regression one. was adopted to reduce the number of variables. The model revealed an Adjusted R-Square of 0.91. However, issues pertaining to illogical coefficient signs and low coefficient significance questioned the integrity of the model.

In the second model, the K-Mean Clustering algorithm was adopted to reduce the number of variables. The developed model is a multiple linear regression one considering the Speed Index as the only independent variable. Different coefficients have been identified for each combination of clusters (20 categories). The estimated average RMSE for all twenty categories was 0.088 and all estimated RMSEs were under 0.2. Such results are perceived to be promising, given the simplicity of the model in terms of information requirement as well as computational ones. Nonetheless, new set of testing data on a different case study route needs to be adopted for model verification purposes.

ACKNOWLEDGEMENT

This research was conducted at the Intelligent Transportation System Center, Nile University, Egypt. Acknowledgement is due to Dr. Dalia Said, Cairo University, Egypt, for her insightful recommendations during the course of this research.

REFERENCES


Akrab Saby has gained his bachelor degree in Urban Planning from Cairo University, Egypt. In 2004, he gained a diploma in GIS from Cairo University. He earned his MSc degree in Intelligent Transportation Systems (ITS) from Nile University Egypt in 2009. He is a Senior Transport Planner with focused experience in GIS development, including; database design and automation, application and system design, and systems integration. He has participated in several transportation system assessment projects in Egypt and the Gulf area.

Hoda Talaat was born in Cairo in 1978. She received her bachelor and MSc degrees in civil engineering in 2000 and 2003 from Cairo University, Egypt. Dr. Talaat received her PhD in transportation engineering from the University of Toronto in 2008. She is currently an Assistant Professor at Cairo University, Egypt. She is the former Director of the ITS program at Nile University, Egypt. She teaches and conducts research in the area of ITS technologies and deployment. She has participated in several consulting projects related to traffic surveillance technologies, traffic network modeling, and traffic operations assessment. She teaches a number of undergraduate and graduate courses in the general domain of Traffic engineering and in the specific field of ITS, at the University of Toronto, Cairo University, and Nile University.

Dr. Talaat is a former NSERC research scholar and the holder of awards from the Transportation Association of Canada, ITS Canada and the Ontario Government.