DEVELOPMENT OF A REGIONAL NETWORK PERFORMANCE MEASUREMENT MODEL FOR PLANNING APPLICATION BASED ON HIGH-FREQUENCY GPS DATA

Hyunsoo Noh, PhD, Corresponding Author
Pima Association of Governments
1 East Broadway, Suite 401, Tucson, AZ 85701
Tel: 520-792-1093; Email: hnoh@pagregion.com

Aichong Sun, PhD, PE
Pima Association of Governments
1 East Broadway, Suite 401, Tucson, AZ 85701
Tel: 520-792-1093; Email: asun@pagregion.com

Xianbiao Hu, PhD
Metropia Inc.
1790 East River Road, Suite 140, Tucson, AZ 85718
Tel: 520-461-2023; Email: xb.hu@metropia.com

Asif Rehan
Metropia Inc.
1790 East River Road, Suite 140, Tucson, AZ 85718
Tel: 860-786-8032; Email: asif.rehan@metropia.com

Word Count: 5,583 words text + 10 tables/figures x 250 words (each) = 8,083 words

Submitted for Presentation at the 96th TRB Annual Meeting

Revised Submission Date: November 15, 2016

ABSTRACT

Rapidly-evolving geo-referenced information and communication technologies have improved the feasibility of collecting individualized high-fidelity and fine-resolution GPS data. Such GPS data is used by transportation practitioners to calibrate and validate various travel and traffic models, and to create transportation system performance measures in compliance with programs such as the Congestion Management Process (CMP) and Moving Ahead for Progress in the 21st century (MAP-21). Research on developing a systematic and practical model to best use this GPS data, however, is rather limited. This research extends and expands this line of inquiry, introducing a comprehensive procedure containing a series of four models to measure regional transportation network performance with high-frequency GPS data. The series consists of a mode detection model, a map-matching model with sub-models of a network expansion and a shortest path model, an activity filtering model, and a link-performance measure update model. With 56.9 million GPS points collected over a 15-month period in Tucson, Arizona, the procedure generated transportation performance measures including free-flow speed, average speed, and delay. The outcome of this study provides transportation practitioners with a comprehensive and practical solution to the development of transportation network performance measures using high-frequency GPS data.

Keywords: high-frequency GPS trajectory data, network performance measurement, transportation planning, rule-based map-matching.

INTRODUCTION

With the rapid advancement of geo-referenced information and communication technologies, "big data" is now a critical subject in the field of transportation. Through a variety of sources including smart phones, vehicle on-board devices (OBD), and many other GPS devices, travelers' trajectory data can be collected and used by researchers and practitioners to assess and improve the performance of transportation systems. This individualized data may now be used to create performance measures required by federally mandated planning as with the Congestion Management Process (CMP) and Moving Ahead for Progress in the 21st century (MAP-21) in the U.S., for example.

Using the GPS-based trajectory data of travelers has not been an easy undertaking, however, due to a lack of systematic and practical procedures to process and analyze the data. This issue is often compounded by transportation network data containing insufficient detail to delineate a traveler's trajectory using GPS data. For instance, transportation networks used by many Metropolitan Planning Organizations (MPOs) in the U.S. are prepared and maintained for planning and modeling purpose and might prune the local roads due to various considerations, e.g. computation cost, model calibration challenges, and so on. If GPS data is collected from those pruned roads, it creates an incompatibility problem. In addition, for the performance measurement of regional transportation systems, outsourced GPS data is frequently utilized, which is usually pre-processed and optimized in a different network configuration. Matching collected GPS data to the planning or modeling network may necessitate additional efforts to adjust and re-sort GPS points based on the planning network.

This research proposes a comprehensive network performance measure development procedure to address the problem of fitting GPS data to road networks along with other issues related to the accuracy of this data. The proposed procedure consists of a series of models— a mode detection model, a map-matching model, an activity filtering model, and a link-performance update model— to facilitate the use of GPS data. Specifically, the mode detection model provides a simple but effective process for determining the used transportation mode among auto, transit-related, and walk-based travel for GPS data. The map-matching model employs (i) a network expansion technique to approximate the sub-network coverage based on initial map-matching results and (ii) a shortest path sub-model to determine the most likely network path over GPS-points. The activity filtering model detects non-driving activities and removes the associated GPS data to improve the quality of performance measure. Finally, the link-performance update model develops and updates the link performance measures. The proposed procedure by this study has been implemented to estimate link-based free-flow speed, average speed, and traffic delay using the data collected in Tucson, Arizona metropolitan region.

LITERATURE REVIEW

In this study, the scope of our approach could best be framed as post-trip analysis of high-frequency GPS data. Above all, a post-trip analysis is appropriate for the purpose of measuring network performance or processing household surveys (*1*,*2*,*3*). This type of analysis focuses more on behavioral pattern analysis around daily activity and additional quantitative analyses like delay estimation and less on outperformed system performance like "real-time" route provision (*4,5,6,7*). Since collected GPS data has a second-by-second temporal resolution we are interested in approaches to "high-frequency" data (*1,8,9,10,11,12*). These models are distinct from a "low-frequency" approach (*13,14*) that frequently uses advanced techniques like hidden Markov models.

Many different methods with various levels of sophistication exist for detecting transportation mode from high-frequency GPS data. Shen and Stopher (*15*) reviewed travel mode detection methods in travel survey data which considered a number of major factors, including speed (*16*), acceleration/deceleration (*13,16,17*), transit GIS information like timetables, routes, and stops, and advanced modeling techniques such as Fuzzy logic (*4,18*), Bayesian belief network (*19*), or hidden Markov chain (*17*). Among the methodologies, this study is interested in a simple approach using transit stop information to filter for GPS data from automobiles as the "main" mode for further analysis.

The map-matching model could be categorized in different ways (*8*, *20*), which might be reframed as "geometric/topological", "probabilistic" and "advanced" approaches. This study is more interested in the "geometric/topological" approach that uses a deterministic relation by predefined rules between GPS data and network topology (*1,2,3,5,6,8,9,21,22,23,24*). Quddus and Washington (*14*), advocate a topological method with this high-frequency data based on a rule-based methodology, or so-called multiple hypothesis technique (MHT) (*25*). Quddus et al. (*26*) also argued that a good network topology is more important than the resolution of the given network. Their advanced, probabilistic and fuzzy logic, models, improved the map-matching rate marginally against the geometric model, especially on an uncomplicated network in a suburban area. In emphasizing network topology over network detail, a network missing local roadways but with good characteristics such as link direction, curvature and turn prohibition will produce good map-matching results, particularly when used with a topological rule-based approach.

METHODOLOGY OF NETWORK PERFORMANCE MEASURE DEVELOPMENT

The methodology of network performance measure development is based upon the availability of the following data sources:

- High-frequency GPS data that includes GPS location coordination, time stamp, heading and instant speed;
- Transportation network; and,
- Transit system, particularly transit stop/station locations.

The overall process in the methodology is shown in FIGURE 1.

For each trip (i.e. a group of GPS location points), the model starts with a simple, but effective, module to evaluate the transportation mode used by the trip, and then excludes the GPS data surveyed from non-auto modes by "Mode Detection".

The "map-matching" module offers a set of successive sub-models to search for the most likely path compatible with the collected trip. Specifically, within "map-matching" module, GPS points from the observed auto trip will be initially matched onto the links of the prepared transportation network based on a calculated point-to-line distance (*2*). The coverage of matched links is then estimated by that link coverage, defined as the ratio of covered distance by the successive GPS points to the link length. Subsequently, a rule-based filtering process is employed to identify anchor links that have high likelihood of being used by the trip. These anchors links are then utilized by an improved shortest path algorithm to search additional network links to connect them in order to form a complete path for the trip being analyzed. Finally, all GPS points of the trip will be re-matched onto the links in the built path only.

The "activity filtering" module determines the set of GPS points observed from off-road, non-driving activities— e.g., car washing and fueling—and excludes them from further analysis.

The "link-specific performance measure update" estimates link-related performance measures, such as link travel time and speed, based on selected matched GPS points.

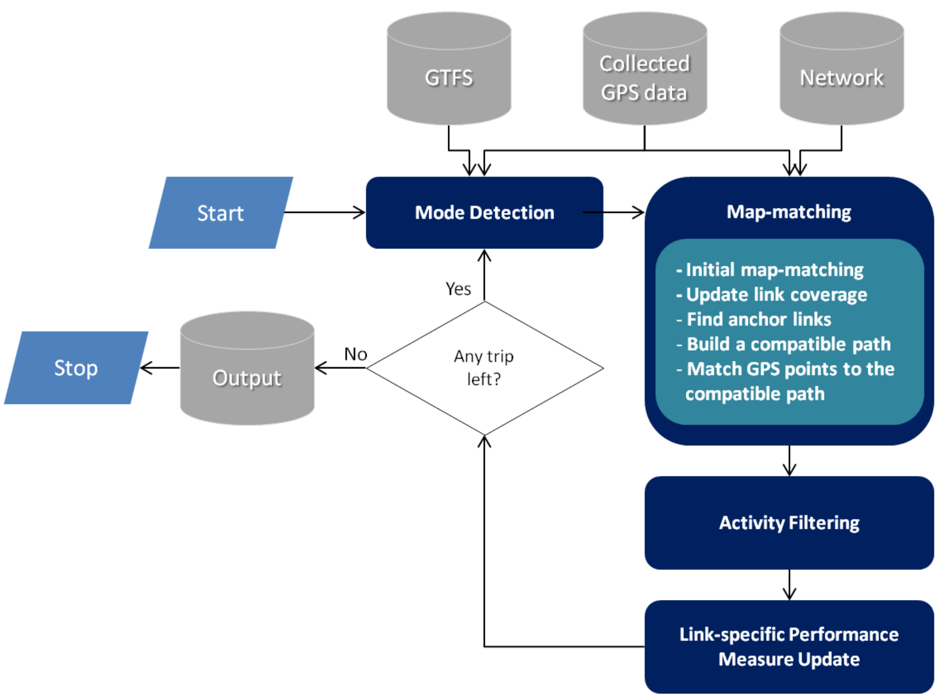


FIGURE 1 Overall Procedure of the Proposed Network Performance Measurement Model

Mode Detection

Specifically, a trip may use any combination of walk, transit and auto modes, such as walk-**auto**-walk, walk-**transit**-walk, walk-auto-**transit**-walk, and walk-**auto**-walk (shopping)-**auto**-walk as well as **walk**-only, where a main mode of a trip is considered in bold. If combined modes have been used, the following rules are applied to identify the main mode:

- Walk cannot become a main mode unless the trip is made entirely by walk mode, which only occurs when none of GPS points of the trip has instant speed exceeding 6 mile/hour (9.7 km/hour), which are for cutting off a obvious walk or slow mode including bike and any non-driving access to a car.
- Transit is the main mode once used, regardless how large the portion of trip is by transit mode. Transit mode usage is determined by a sequence of occurrences of low speed GPS points clustering around the bus stops within close proximity (e.g. 100 ft;30.5 m), supplemented by the direction information of the GPS points. This approach works better with far-side transit stops since stopping at an intersection is not easily confused with stopping at a bus stop.

Map-matching

Initial Map-Matching

Initial map-matching simply measures distances from each GPS point to any network links within 200 feet (61 m) and selects the nearest one to match the GPS point. The distance between a GPS point and a network link is measured as either the perpendicular distance along the pedal line from the GPS point to the link or the shortest distance from the GPS point to either end of the line (*2*,*22*). An example of a mapped network location for each GPS point is shown in FIGURE 2, specified as “Mapped GPS point.”

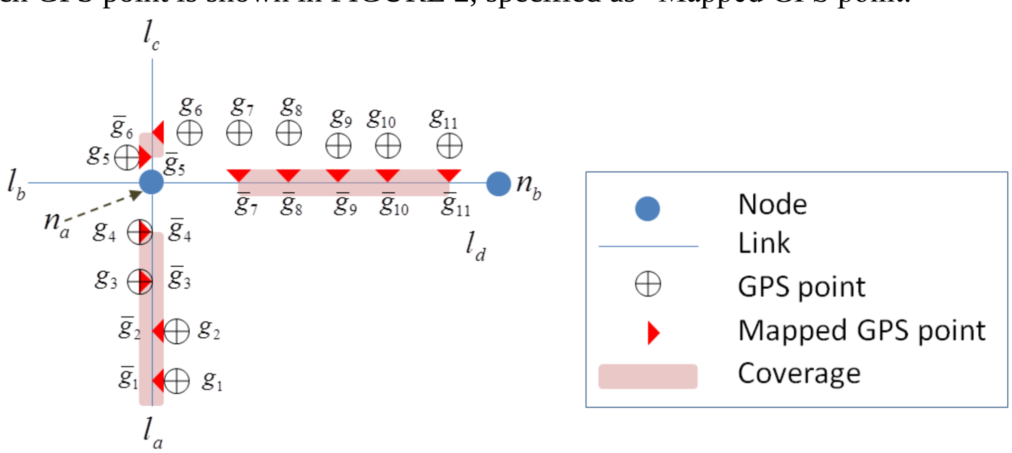


FIGURE 2 Example of Initial Map-matching and Link Coverage Estimation

The initially matched network links may not always form a sufficiently complete and clean path for the analyzed trip due to missing or erroneous GPS data and on occasion to an incompatible match reference network. Given this difficulty with matching, the path building model searches for the most likely trip path based on the initially matched links.

Link Coverage Estimation Model

A weighting function for GPS points may be useful to assess point-to-link map-matching (*2*,*3*,*5*,*9*). This study, alternatively, proposes a "link coverage" concept, especially for assessing a group of GPS points over a link instead of one-by-one assessment. The link coverage is estimated with the initially mapped positions on a given network. Calculation of the link coverage C_l for each initially matched link l is shown in Equation (1) and (2).

$$C_l = \bar{L}_l / len(l) \tag{1}$$

$$\bar{L}_l = len(l) - d(g_u, \bar{g}_{\min_i}) - d(g_d, \bar{g}_{\min_j}) \text{ such that } \bar{g}_{\min_i} = \arg\min_i \, d\left(g_u - \bar{g}_i\right) \text{and}$$

$$\bar{g}_{\min_j} = \arg\min_j \, d\left(g_d - \bar{g}_j\right) \;\; \forall i < j \text{ and } g_i, g_j \in G_l \tag{2}$$

Where,

$\bar{L}_l$: Covered link distance by mapped GPS points,

$len(l)$: Length of link l or $l(u,d)$ with upstream node u and downstream node d,

g_u and g_d: GPS point of upstream u and downstream node d, respectively,

$d(g_i,.g_j)$: Euclidean distance between two GPS points ($g_i,.g_j$),

$\bar{g}_i$: mapped position of GPS point g_i on a link,

$G_l = \{g_i, g_j, ..., g_N\}$: a set of GPS points mapped to link l such that $m(g_i) = m(g_j) = ... = m(g_N) = l$, and

$m(g_i)$: mapping function of GPS point g_i to a specific link.

For instance, as in FIGURE 2, 11 raw GPS points g_1 to g_4, g_5 to g_6, and g_7 to g_{11} are initially mapped to link l_a, l_c , and $l_d(n_a, n_b)$, respectively mapped to $\bar{g}_1$ to $\bar{g}_{11}$. The covered length of link $l_d(n_a, n_b)$ is estimated as $\bar{L}_{l_d} = len(l_d) - d(g_{n_a}, \bar{g}_7) - d(g_{n_b}, \bar{g}_{11})$. When temporal or spatial resolution of the GPS data is sufficiently high and all GPS points have been sorted in non-decreasing order, the covered link distance may be approximated as $\bar{L}_l \approx \sum_{i=1|m(g_i)=l}^{N-1} d(g_i, g_{i+1})$ until no further activity is detected on link l.

Furthermore, with the calculated link coverage, the network expansion model propagates the network links, beginning with a starting link which should be the only link that a GPS point can be matched to within a certain distance (e.g. 35 ft;10.7 m), as described in (*3*). Marchal et al. (*2*) and Schuessler and Axhausen (*6*) also applied a similar type of network expansion model. The identified starting link is also labeled as the first permanent parent link and all other directly connected links will become its children links. Each child link is reviewed as an alternative link to identify the next permanent parent link. The next permanent link should have the best link coverage among all children links and in the meantime its link coverage should exceed a pre-specified threshold (e.g. 40%). The process continues until no permanent parent link can be identified when searching either has reached the destination or cannot proceed in the middle of trip. For the latter case, the network expansion model will then start over to find the next starting link and repeat the above-described process.

Identification of Anchor Links

Although the network expansion model finds a set of links with the best link coverage (also called permanent parent links), these links do not necessarily form a complete and clean trip path.

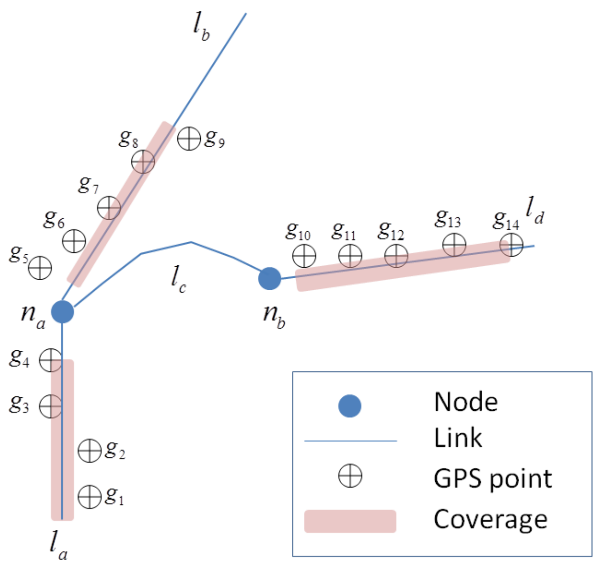


FIGURE 3 Example of Compatibility between GPS Points and Transportation Network

For example, as in FIGURE 3, l_a, l_b, and l_d are identified by the network expansion model as the best covered links for 14 raw GPS points, g_1 through g_{14}. As one may notice, these three links do not represent a connected path. Therefore, a rule-based filtering model is developed to determine the "anchor links" out of the best covered links based on rules as in FIGURE 4.

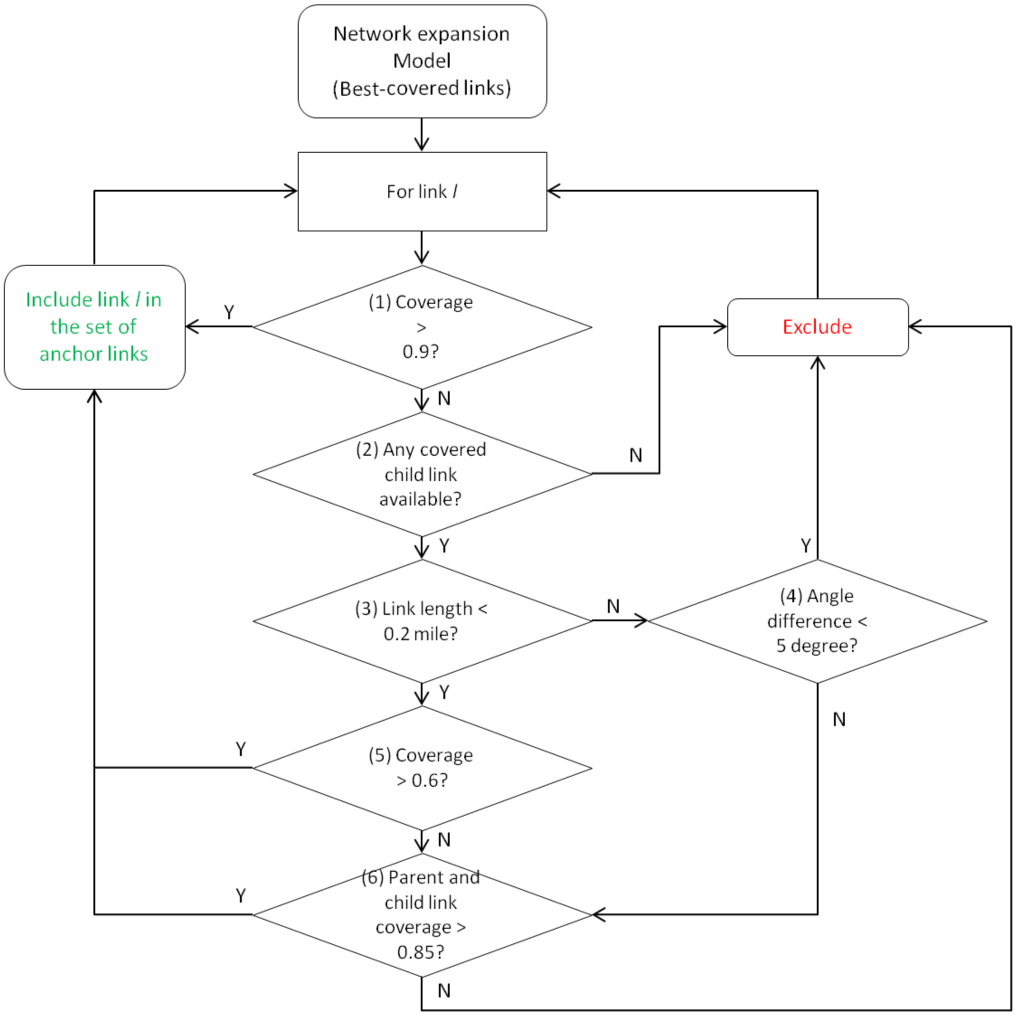


FIGURE 4 Rule-based Filtering Process for Finding a Set of Anchor Links

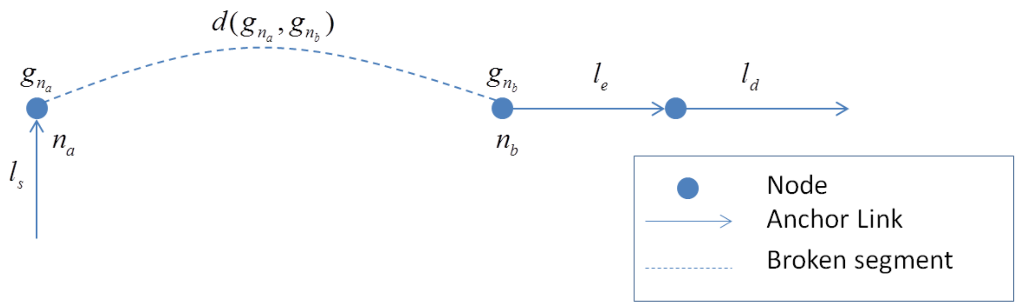
Rule (1) could be customized and calibrated for each network facility type based on the quality of GPS data and network resolution; rule (2) considers the network connectivity; rules (3) and (5) account for the relatively lower link coverage caused by shorter link length; rule (4) takes into consideration the imprecision of geo-matching due to closely located links with a small angle of different between them; and, rule (6) determines anchor links by evaluating the case of a short link with low coverage if confined in its parent and children link with good coverage.

Building a Compatible Path

The "anchor links" determined by the rule-based filtering model represent a portion of an actual trip path that consists of several sub-paths. In order to complete the trip path, a shortest path model or its variant is typically used to search a feasible path between two unconnected GPS points (*2*,*14*,*18*,*24*). In this study, a link-based shortest path model is developed to link the unconnected anchor links for a complete trip path, called "compatible path." The shortest path model is applied with a buffer based on the

Euclidean distance of the two disconnected anchor links as described in the following example.

FIGURE 5(a) shows an example of 3 anchor links, $\{l_s, l_e, l_d\}$ with l_s and l_e disconnected. The buffer is calculated as $\alpha \cdot d(g_{n_a}, g_{n_b}) + len(l_e)$. Where, α is a parameter to adjust buffer area in order to improve shortest path search result. As you noticed, the buffer includes the length of link $len(l_e)$. Since the link-based shortest path model updates the label at the end of each link (*27*), if the length of an ending anchor link l_e is relatively long, the chance of this link to be accessible within a buffer $\alpha \cdot d(g_{n_a}, g_{n_b})$ will be low. In addition, U-turn often exists in collected GPS data but shortest path model often faces the difficulty of finding a reasonable path as addressed in (*6*). The link-based shortest path model has the advantage of finding U-turn traffic behavior without further expanding network (*27*).



(a) Example of Shortest Path Search using Buffer between Anchor Links

(i) Overall model

01: For $l_s, l_e \in A$
02: If $l_e \in F(l_s)$
03: $P = P \cup \{l_e\}$
04: Else
05: $\overline{P} = SP(l_s, l_e)$
06: If $\overline{P} \neq \phi$
07: $P = P \cup \overline{P}$
08: Else
09: $k = 1,\ \tilde{P} = \{l_s\}$
10: While $\overline{P} = \phi$ or $k = K$
11: $\overline{P} = SP(l_{s-k}, l_e)$
12: $\tilde{P} = \tilde{P} \cup \{l_{s-k}\}$
13: $k = k + 1$
14: $P = (P - \tilde{P}) \cup \overline{P}$

(ii) $SP(l_s, l_e)$:Link-based Shortest Path Model

15: $B = \max(1.0\ ,\ \alpha \cdot d(g_{n_a}, g_{n_b}) + len(l_e))$
16: $c_{l_a} = \infty \quad \forall l_a \in L$
17: $\hat{c}_{l_s} = 0\ ,\ \overline{P} = \phi\ ,\ Q = \phi$
18: $\overline{P} = \overline{P} \cup \{l_s\}\ ,\ Q = Q \cup \{l_s\}$
19: While $Q \neq \phi$
20: Retrieve $\arg\min\ l_a \in Q\ ;\ Q = Q - \{l_a\}$
21: If $\hat{c}_{l_a} < B$
22: For $l_b \in F(l_a)$
23: If $\hat{c}_{l_a} + c_{l_b} < \hat{c}_{l_b}$
24: $\hat{c}_{l_b} = \hat{c}_{l_a} + c_{l_b}$
25: $\overline{P} = \overline{P} \cup \{l_b\}$
26: $Q = Q \cup \{l_b\}$
27: return $\overline{P}$

A: a set of given anchor links; P: a compatible path, or a subset of links, from origin to destination; $\overline{P}$: a subset of links searched by shortest path model; $\tilde{P}$: a subset of links overlapped over a shortest path; $F(l)$: a subset of children links of a parent link l; c_{l_a} : link cost of link l_a; L: a set of links on a network; $\hat{c}_{l_a}$: permanent label of link l_a; Q: a set of links for heap structure

(b) Shortest Path Search Algorithm using Buffer

FIGURE 5 Shortest Path Model Linking Anchor Links

With the calculated buffer B (line 15) in FIGURE 5(b), shortest path model (ii) $SP(l_s, l_e)$ (line 15-27) is applied to bridge the gap between two disconnected anchor links l_s and l_e. Once the subsequent child link of link l_s is anchor link l_e (line 02) or a sub-path $\overline{P}$ is found to connect the two anchor links (line 05), every link searched is added into path P

(line 03 and 07). However, one may run into the situation where no shortest path can be found between the disconnected anchor links (from line 08). If it so happens, the shortest path model is performed to search the secondary shortest path between the predecessor links of l_s, l_{s-k} $\forall k = \{1,2,\ldots,K\}$and $l_{s-k} \in P$, and l_e (line 11) instead. If the secondary shortest path would be found, all the links on the secondary shortest path $\overline{P}$ will be added onto path P (line 14). In the meantime, the anchor link l_s and its other predecessor links, $\tilde{P}$, that couldn't produce secondary shortest path to anchor link l_e will be removed from the anchor link list of path P (line 14). The proposed shortest path model in (ii) is a label-setting model as satisfying the optimality condition (line 23) using the proposed buffer (line15 and 21) and heap structure Q in (ii).

Remapping of GPS Points on the Compatible Path

After the compatible path is found for each trip as described above, all GPS points of the trip need to be remapped to the links in its corresponding path. Specifically, each GPS point is categorized into either a node or a link group, based on the distances between the GPS point and the nearest node and link. For this purpose, a threshold distance δ is predetermined. If the distance between a GPS point and the nearest node is shorter than δ, the GPS point will be assigned to node group. Otherwise, it is classified into link group. For example, in FIGURE 6, g_4 to g_6 are within δ distance of node n_a and g_{11} is within δ distance of node n_b, those GPS points are thus assigned to the node group. On the contrary, the other GPS points are assigned to the link group. For the GPS points in link group, they will be remapped onto the nearest links denoted as $\hat{g}_i$ and "Adjusted GPS point" in FIGURE 6.

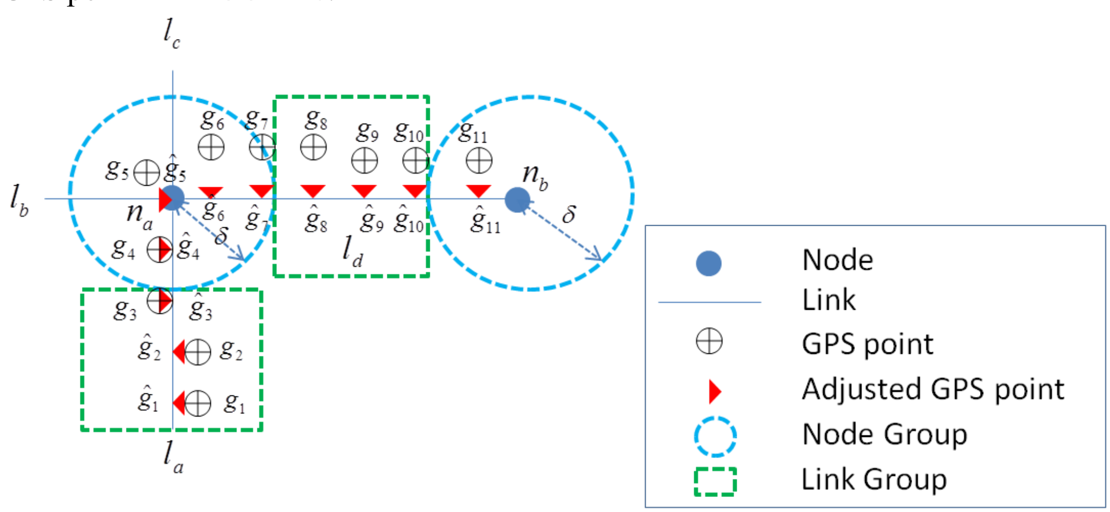


FIGURE 6 Node and Link Group by Node Buffer

In node group, each GPS point g_i may be mapped to either one of the two links that connect at node n_a, notated as $L_{n_a} = \{l_a, l_b\}$ such that $l_b \in F(l_a)$. In determining which link the GPS point should be mapped onto, the heading (direction) information of the

GPS points needs to be compared with the directionality of both possible links and their sub-links. Each link l_a can be split into a set of sub-links, $l_a = \{l_a^1, l_a^2, ..., l_a^N\}$, wherein $len(l_a) = \sum_{j \in \{1,N\}} len(l_a^j)$. Each sub-link is defined by a pair of consecutive geo-coded points of link l_a. The sub-link closest to the GPS point g_i for either l_a or l_b can be easily found as it would always be the first or last sub-link on the list. The heading discrepancy between GPS point g_i and link l_a can be estimated as $\Delta(g_i, l_a) = f\left(dir(g_i) - dir\left(l_a^{\min}(g_i)\right)\right)$[1]. Where, $dir(\cdot)$ stands for heading direction of GPS point or link, exactly the closest sub-link, $l_a^{\min}(g_i) = \arg\min_{l_a^j}\left(d\left(g_i, g_{l_a}^j\right)\right) \quad \forall j = \{1,2,...,N\}$.

Finally, the link with the smaller heading discrepancy with the GPS point is selected as the final mapped link for GPS point g_i, and GPS point g_i is be relocated to $\hat{g}_i$ on the link. In the example, GPS points { g_4, g_5 } and g_6 are mapped onto link l_a and l_d respectively. As a final product of this sub-model, each link is assigned with remapped GPS points.

Activity Detection

In the data collection, off-street activities (e.g. coffee) are common during a trip and when the transportation network has insufficient detail additional step(s) are mandated by the off-street activities. An activity detection model is developed to identify off-street activities to improve the quality of analysis. The off-street activities are presented as a large number of clustering GPS points in a small area, and/or a group of closely located GPS points often combined with slow speed of walking (*1*,*18*,*28*). This study proposes an improved method of activity detection based on a group of GPS points mapped on a specific link by the previous steps. In the proposed activity detection model, an off-network activity is determined through a multiple-step confirmation process, primarily reviewing:

(a) How GPS data derived length of travel on a link compares with the link length,
(b) If travel direction changes abruptly at a middle point of the link that may suggest access to an off-street activity establishment, and
(c) If walking activity is detected and continued for a period of time outside the buffer zone of the link.

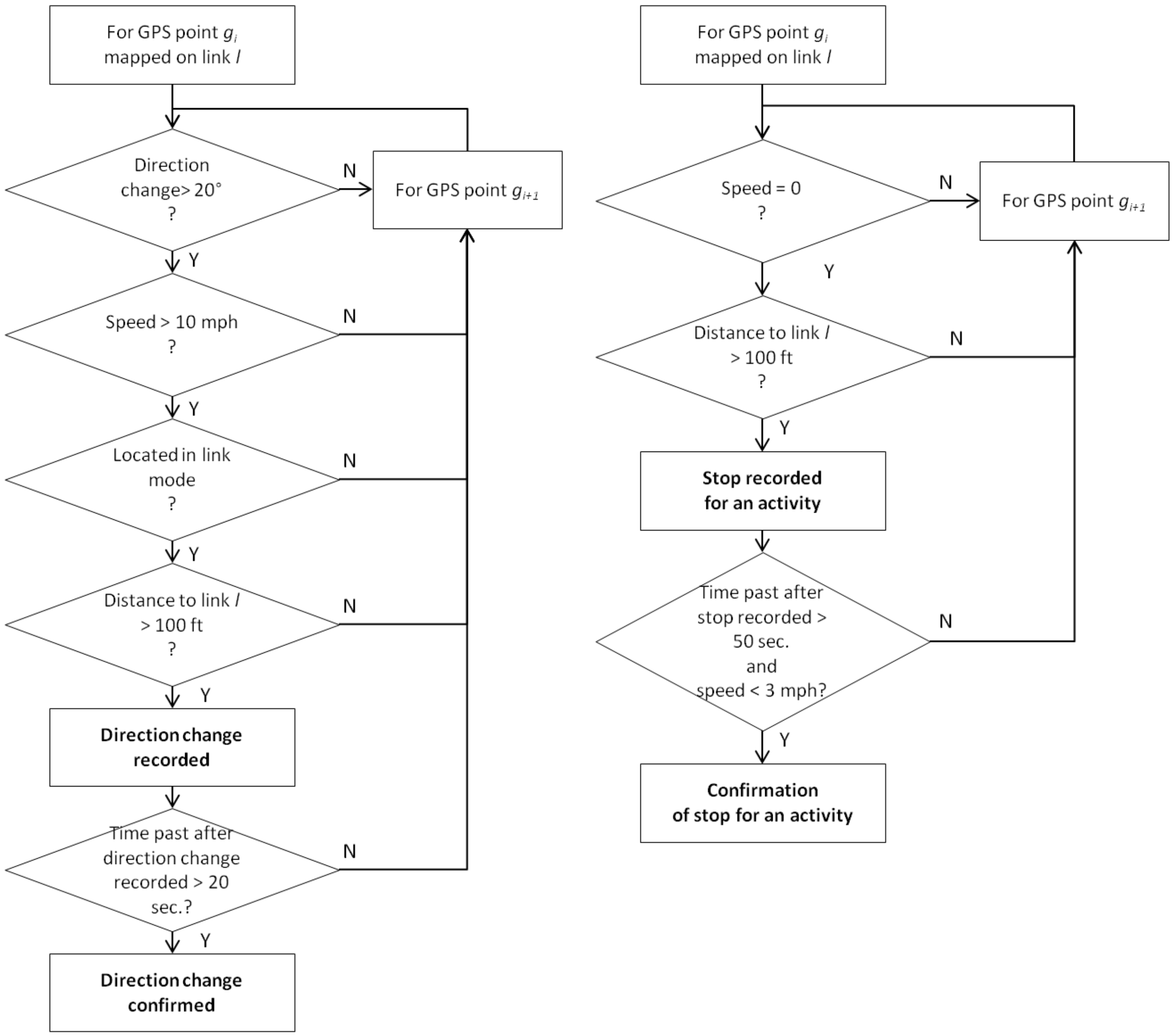
Condition (a) could be evaluated either by comparing the distance of collected GPS points with the link length, $\sum_{i=1}^{N-1} d(g_i, g_{i+1}) > \varphi_1 \cdot len(l_a)$ such that $l_a(n_a, n_b) = \hat{m}(g_i)$ for

---

[1] $\Delta(g_i, l_a) = f\left(dir(g_i) - dir\left(l_a^{\min}(g_i)\right)\right) =$
$\min\left(\max\left[dir(g_i), dir\left(l_a^{\min}\right)\right] - \min\left[dir(g_i), dir\left(l_a^{\min}\right)\right], 360 + \min\left[dir(g_i), dir\left(l_a^{\min}\right)\right] - \max\left[dir(g_i), dir\left(l_a^{\min}\right)\right]\right)$

the activities in the middle of link. If the activities happen around either of the end nodes (or intersections) of a link, this condition might not work, especially for a relatively long link. Instead, apply a rule $\max\left(d\left(g_1, g_{n_a}\right), d\left(g_N, g_{n_b}\right)\right) > \varphi_2 \cdot len\left(l_a\right)$. In addition, if one is interested in where the GPS points are involved in an activity, it could be identified by comparing the distance of original GPS points with the mapped distance: $\left|\sum_{i=1}^{j-1} d\left(g_i, g_{i+1}\right) - \hat{d}\left(\hat{g}_1, \hat{g}_j\right)\right| > \varphi_3$ and $\left|\sum_{i=j}^{N-1} d\left(g_i, g_{i+1}\right) - \hat{d}\left(\hat{g}_j, \hat{g}_N\right)\right| > \varphi_4$. Where, n_a and n_b are upstream and downstream node, φ_1, φ_2, φ_3, and φ_4 are parameters, and $\hat{d}\left(\hat{g}_i, \hat{g}_j\right)$ is the coverage distance of mapped GPS points excluding overlapped portions.

How the activity detection model works, particularly in terms of (b) and (c), is also shown in FIGURE 7. For the mapped GPS points of a link, we apply two processes evaluating the condition (b) and (c) separately, but in a same computation loop, although any activity is usually represented in a combined mechanism. *First*, the left half is designed for condition (b) and inspects if any direction change is relatively sharp while accessing the activity location with low speed. To confirm access to the activity location, the GPS point location should be out of any intersection (or node group) as well as located more than 100 ft (30.5 m) outside of a link. If those conditions are satisfied, the model records this GPS point as the first GPS point in access mode and tracks if the following GPS points are also in access mode and checks if this access mode continues for a significant amount of time. If it is true, this process confirms an access mode has been detected on this link. *Second*, the right half is designed for condition (c) and checks if a vehicle stops completely and if a GPS point is located out of the 100-ft (30.5-m) link buffer; if so, the module examines how long the walking mode continues. We applied a 6-mile/hour (9.7-km/hour) threshold for detecting this walking mode. This process confirms that a traveler has stopped at a non-travel activity location if such a pattern continues for significant amount of time. All parameters could be estimated according to the resolution of the GPS data and transportation network.



Condition (b) Condition (c)

FIGURE 7 Activity Detection Model (Direction Change and Stop for an Activity)

Link-specific Performance Measure Update

Prior to the development of performance measures, the model takes one more step to further clean the mapped GPS points. In addition to excluding the links with low coverage, the final refinement trims the end of GPS points for accuracy. The refinement process first compares the distance covered by GPS points with the length of the link by $\sum_{i=1}^{N-1} d(g_i, g_{i+1}) > len(l_a) + \varepsilon$. Where, ε is a parameter denoting positive measurement error. If this refinement is required, apply the condition $d(g_i, \hat{g}_i) < \varphi_5$ and $d(g_i, .g_n) < \varphi_6$ to find out a boundary GPS point g_i. Where, g_n is the GPS point of the closest node from

g_i and φ_5 and φ_6 are parameters. The GPS point(s) which does not satisfy the condition will be trimmed out.

For the selected link l and the mapped GPS points, $g_1,...g_N$ such that $l = \hat{m}(g_j)\ \ \forall j \in \{1, N\}$, performance measurements are estimated as follows.

Link travel time: $t_l = \tau(g_N) - \tau(g_1)$; Space-mean speed: $s_l = \sum_{i=1}^{N-1} d(g_i, g_{i+1}) \big/ t_l$; Normalized link travel time: $\hat{t}_l = len(l)/s_l$; Free-flow speed: $\phi_l^{p,q} = \max^{15th}\left(\bar{S}_l^{p,q}\right)$; Link delay: $x_l^{p,q} = len(l) \cdot \left(1/s_l - 1/\phi_l^{p,q}\right)$.

Where,

$\tau(g_i)$: Unix time of GPS point g_i .

$\max^{15th}(\cdot)$: 15^{th} max (or 85^{th} percentile) value

$\bar{S}_l^{p,q}$: A set of speeds of link l faster than speed limit q for link class p

RESULTS

We applied the proposed network performance measurement model to high-frequency GPS data collected from March 2015 to May 2016 in the Tucson, Arizona metropolitan region by a smart phone application. The collected raw GPS data contains 56.9 million points grouped around 87,000 trips. The computation time was estimated to the average of 1.77 seconds per trip and the estimated total computation time is flexible from 5 hrs to 9 hrs according to the utilization of different set-up with multiple cores. In terms of data usage, around 59.7% of the collected data was utilized for the final performance measurement. Each location data point includes latitude, longitude, direction, time stamp, accuracy level, trip ID, and so on. The applied regional transportation network, including 2,259 links and 1,404 nodes, and the regional General Transit Feed Specification (GTFS) (*29*) are also inputs to process the collected GPS data. The regional transportation network includes link attributes of length, traffic direction, posted speed limit and functional class.

As a partial output, FIGURE 8 shows some selected links (indicated by red line) after cleanup of “noise” that affected performance measures. Invalid GPS points or “noise” in the data (see the explanation tag, categorized in a, b, and c) are carefully excluded by the proposed noise filtering models: the activity detection model and the filtering process in the link-performance update model.

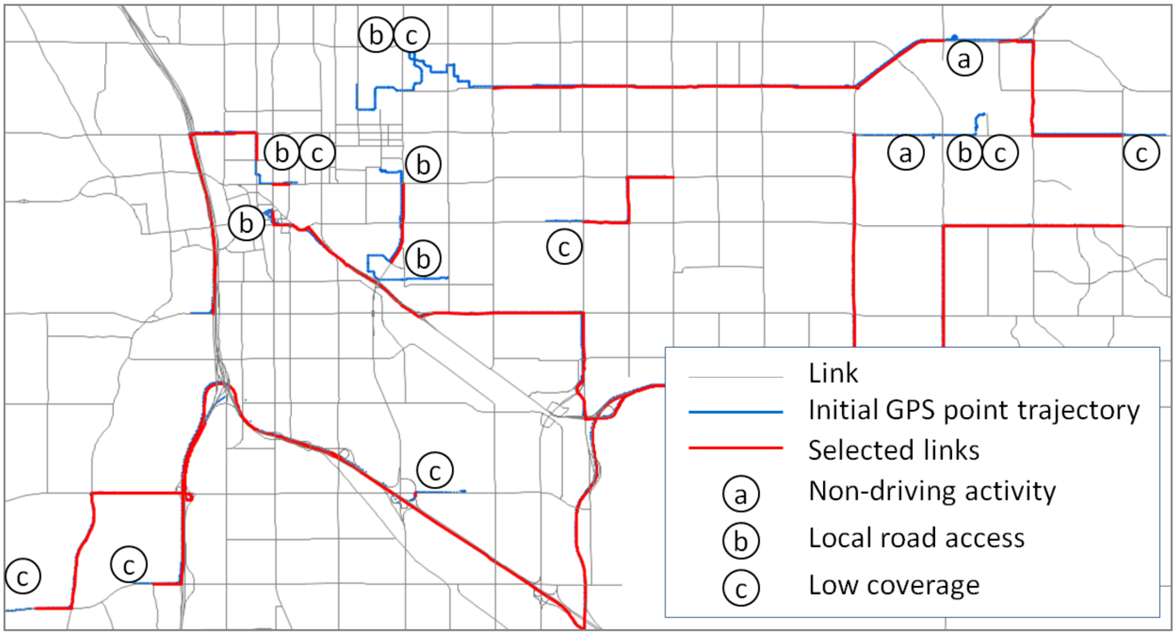


FIGURE 8 Sample Output of Selected Links for Link Performance Measures

As a final product, the proposed model produces link travel time and link speed estimated from the GPS points of selected links. And free-flow speed and delay are further estimated through post-processing using the estimated link speed.

As specified early, free-flow speed is necessary for estimating link delay. In general, free-flow speed is considered as the average travel speed in uncongested free-flow condition no traffic or enough headway allowed for moving forward without any influence (*30*). 85^{th} percentile speed is the customary recommended speed in the condition (*31*). With only the collected GPS data, however, it is hard to identify the condition. For that reason, we relaxed the condition assuming that the free-flow speed starts to form between 80% and 90% of the speed limit. With this relaxation, we prepared two sets of link speed data faster than 80% and 90% of speed limit and found 85^{th} percentile speed for each data set, respectively. Finally, the average of these two 85^{th} percentile speeds was used for the estimated link free-flow speed.

TABLE 1 shows that the estimated free-flow speed and average speed by link class and speed limit. We note that collectors, ramps, frontage roads are not considered in the result, because of insufficient data for collectors and the generic unstable characteristics of ramps and frontage roads as connecting freeway or parkway to arterial roads.

TABLE 1 Free-flow Speed and Average Speed by Link Class and Speed Limit

(a) Free-flow Speed

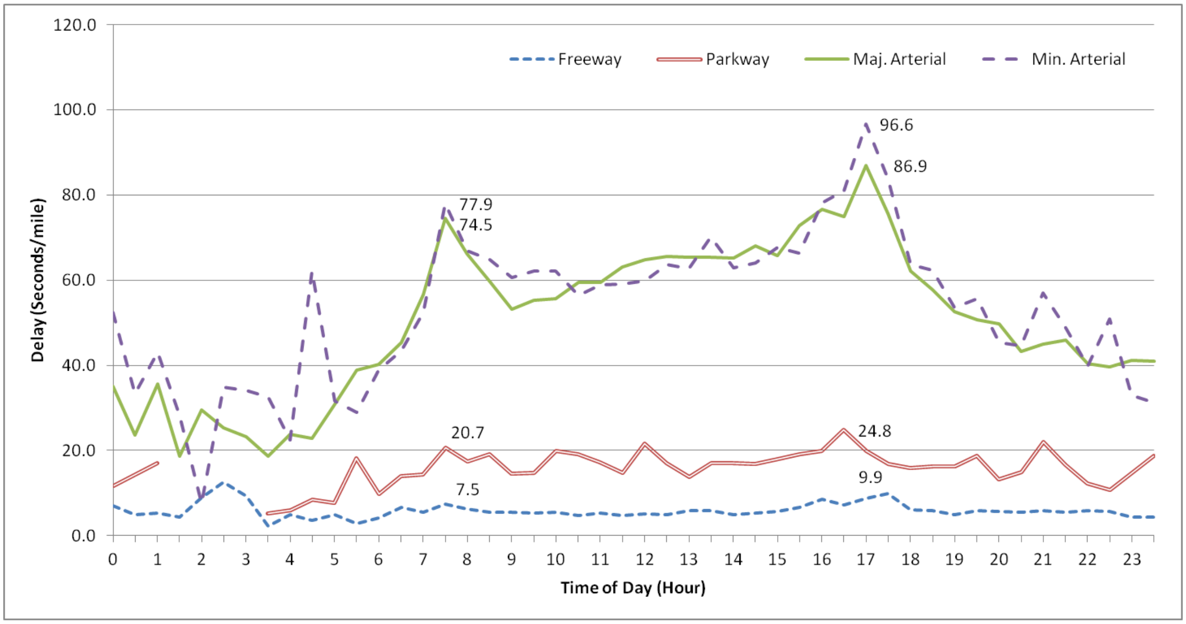
	Speed Limit (mile/hour (km/hour))									
Class	20 (31.2)	25 (40.2)	30 (48.3)	35 (56.3)	40 (64.4)	45 (72.4)	50 (80.5)	55 (88.5)	65 (104.6)	75 (120.7)
Freeway									73.8 (118.8)	80.7 (129.9)
Parkway						52.7 (84.8)	57.8 (93.0)	62.6 (100.7)		
Maj. Art.		32.1 (51.7)	36.0 (57.9)	40.5 (65.2)	44.3 (71.3)	49.6 (79.8)	55.2 (88.8)	55.6 (89.5)		
Min. Art.	24.9 (40.1)	32.2 (51.8)	35.1 (56.5)	40.3 (64.9)	44.7 (71.9)	47.8 (76.9)	52.7 (84.8)	59.3 (95.4)		

(b) Weekday Average Speed

	Speed Limit (mile/hour (km/hour))									
Class	20 (31.2)	25 (40.2)	30 (48.3)	35 (56.3)	40 (64.4)	45 (72.4)	50 (80.5)	55 (88.5)	65 (104.6)	75 (120.7)
Freeway									65.9 (106.1)	75.0 (120.7)
Parkway						41.5 (66.8)	49.2 (79.2)	49.3 (79.3)		
Maj. Art.		14.1 (22.7)	20.2 (32.5)	23.7 (38.1)	26.7 (43.0)	32.9 (52.9)	39.0 (62.8)	39.5 (63.6)		
Min. Art.	11.3 (18.2)	15.9 (25.6)	22.3 (35.9)	27.1 (43.6)	34.1 (54.9)	35.0 (56.3)	40.3 (64.9)	45.8 (73.7)		

Free-flow speeds of freeway/parkway and major/minor arterials are estimated on average to be 7.5 mile/hour (12.1 km/hour) and 4.7 mile/hour (7.6 km/hour) faster than the speed limits, respectively. And weekday average speed (TABLE 1(b)) shows that freeway average speed is very close to the speed limit and the parkway provides a relatively good service around 3.3 mile/hour (5.3 km/hour) slower than the speed limit. On the other hand, the arterial system provides relatively moderate service around 9.1 mile/hour (14.6 km/hour) slower than the speed limit.

Average delay (per mile) was estimated with the estimated free-flow speed, average speed, and a given link length as shown in FIGURE 9. Above all, the delay pattern of freeway is stable even during peak hours (AM:6:30~8:30; PM:16:00~18:00) compared to daily average delays of 6.1 seconds/mile (3.8 seconds/km). The parkway shows relatively worse but stable performance of delay during peak hours, in which the daily average delay was estimated to 17.4 seconds/mile (10.8 seconds/km). On the other hands, average delays on arterials are noticeably increased during peak hours, which also can be compared to the daily average 62.0 and 65.2 seconds/mile (38.5 and 40.5 seconds/km) for major and minor arterial, respectively.



* seconds per mile (seconds per km): 77.9 (48.4); 74.5 (46.3); 96.6 (60.0); 86.9 (54.0); 20.7 (12.9); 7.5 (4.7); 24.8(15.4); 9.9 (6.2)

FIGURE 9 Average Delay (seconds/mile) by Time of Day

CONCLUSION AND FUTURE WORK

In this research, we introduced a method for regional performance measurement using high-frequency GPS data based on rule-based topological map-matching, activity and mode detection models. The proposed method was applied to the collected GPS data over the regional transportation network, after which performance measures for free-flow speed, average speed and delay were estimated. We believe that this effort could support a national program such as CMP and MAP-21, in terms of measuring transportation system performance as integrating any GPS data and transportation network, organically and systematically.

In terms of regional network performance, this research revealed a small portion among many possible transportation performance outcomes. Through collected travelers' behavior and transportation system performances, we plan to continue technical analyses that address: (1) creating detailed performance measurements such as corridor delay analysis, intersection delay, speed profile monitoring, travel time index, and planning time index, (2) analyzing seasonal or monthly travel pattern and its congestion, (3) improving the accuracy of the mode detection model and performance measurement beyond auto mode, and (4) combining the activity detection model with daily activity pattern analyses such as those associated with daily activity diaries.

ACKNOWLEDGEMENT

The survey data was collected by Metropia, Inc. (www.metropia.com) using the Metropia smartphone app with an opt-in process for users to participate in the survey voluntarily.

Metropia applied an anonymizing process to exclude personally identifiable information, removing actual origin and destination locations to protect the privacy of the participants. Special thanks to Eric Kramer (Pima Association of Governments) for thoughtful review.

REFERENCES

1. Doherty, S. T., N. Noël, M. L. Gosselin, C. Sirois, and M. Ueno. Moving Beyond Observed Outcomes: Integrating Global Positioning Systems and Interactive Computer-Based Travel Behavior Surveys, *Transportation Research Circular E-C026: Personal Travel, the Long and Short of It, Transportation Research Board of the National Academies*, Washington, D.C., 2001, pp. 449–466.
2. Marchal, F., J. Hackney, and K. Axhausen. Efficient Map Matching of Large Global Positioning System Data Sets: Tests on Speed-Monitoring Experiment in Zürich. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1935, 2005, pp. 93-100.
3. Pan, C., J. Lu, D. Wang, and B. Ran. Data Collection Based on Global Positioning System for Travel Time and Delay for Arterial Roadway Network. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2024, 2008, pp. 35-43.
4. Quddus, M. A., Noland, R. B. A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. *Journal of Intelligent Transportation Systems*, Vol. 10, No. 3, 2006, pp. 103-115.
5. Quddus, M. A., W. Y. Ochieng, L. Zhao, and R. B. Noland. A General Map Matching Algorithm for Transport Telematics Applications. *GPS Solutions,* Vol. 7, No. 3, 2003, pp. 157-167.
6. Schuessler, N., and K. W. Axhausen. Map-Matching of GPS Traces on High-Resolution Navigation Networks using the Multiple Hypothesis Technique (MHT). *Arbeitsberichte Verkehrsund Raumplanung*, Vol. 568, 2009b.
7. Rehan, A. *Crowdsourcing Real-Time Traveler Information Systems*. Master thesis, University of Connecticut, 2015.
8. Blazquez, C. A. A Decision-Rule Topological Map-Matching Algorithm with Multiple Spatial Data. *Global Navigation Satellite Systems: Signal, Theory and Applications*, 2012, pp. 215-240.
9. Chung, E., and A. Shalaby. A Trip Reconstruction Tool for GPS-Based Personal Travel Surveys. *Transportation Planning and Technology,* Vol. 28, No. 5, 2005, pp. 381-401.
10. Hunter, M., S. Wu, and H. Kim. Practical Procedure to Collect Arterial Travel Time Data using GPS-Instrumented Test Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1978, 2006, pp. 160-168.
11. Ogle, J., R. Guensler, W. Bachman, M. Koutsak, and J. Wolf. Accuracy of Global Positioning System for Determining Driver Performance Parameters. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1818, 2002, pp. 12-24.
12. Quiroga, C., and D. Bullock. Travel Time Information using Global Positioning System and Dynamic Segmentation Techniques. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1660, 1999, pp. 48-57.

13. Chen, J., and M. Bierlaire. Probabilistic Multimodal Map Matching with Rich Smartphone Data. *Journal of Intelligent Transportation Systems*, Vol. 19, No. 2, 2015, pp. 134-148.
14. Quddus, M., and S. Washington. Shortest Path and Vehicle Trajectory Aided Map-Matching for Low Frequency GPS Data. *Transportation Research Part C: Emerging Technologies*, Vol. 55, 2015, pp. 328-339.
15. Shen, L., and P. R. Stopher. Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, Vol. 34, No. 3, 2014, pp. 316-334.
16. Gong, H., C. Chen, E. Bialostozky, and C. T. Lawson. A GPS/GIS Method for Travel Mode Detection in New York City. *Computers, Environment and Urban Systems*, Vol. 36, No. 2, 2012, pp. 131-139.
17. Nitsche, P., P. Widhalm, S. Breuss, N. Brändle, and P. Maurer. Supporting Large-Scale Travel Surveys with smartphones–A Practical Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 212-221.
18. Schuessler, N., and K. Axhausen. Processing Raw Data from Global Positioning Systems without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2105, 2009a, pp. 28-36.
19. Feng, T., and H. Timmermans. Enhanced Imputation of GPS Traces Forcing Full Or Partial Consistency in Activity Travel Sequences: Comparison of Algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2430, 2014, pp. 20-27.
20. Quddus, M. A., W. Y. Ochieng, and R. B. Noland. Current Map-Matching Algorithms for Transport Applications: State-of-the Art and Future Research Directions. *Transportation Research Part C: Emerging Technologies*, Vol. 15, No. 5, 2007, pp. 312-328.
21. White, C. E., D. Bernstein, and A. L. Kornhauser. Some Map Matching Algorithms for Personal Navigation Assistants. *Transportation Research Part C: Emerging Technologies*, Vol. 8, No. 1, 2000, pp. 91-108.
22. Taylor, G., and G. Blewitt. Virtual Differential GPS & Road Reduction Filtering by Map Matching. In *Ion Gps*, Citeseer, 1999, pp. 14-17.
23. Yu, M. *Improved Positioning of Land Vehicle in ITS using Digital Map and Other Accessory Information*. PhD dissertation. Hong Kong Polytechnic University, 2006.
24. Wang, W., J. Jin, B. Ran, and X. Guo. Large-Scale Freeway Network Traffic Monitoring: A Map-Matching Algorithm Based on Low-Logging Frequency GPS Probe Data. *Journal of Intelligent Transportation Systems*, Vol. 15, No. 2, 2011, pp. 63-74.
25. Pyo , J.-S., D.-H. Shin, and T.-K. Sung. Development of a Map Matching Method using the Multiple Hypothesis Technique. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, 2001, pp. 23-27.
26. Quddus, M. A., R. B. Noland, and W. Y. Ochieng. The Effects of Navigation Sensors and Spatial Road Network Data Quality on the Performance of Map Matching Algorithms. *Geoinformatica*, Vol. 13, No. 1, 2009, pp. 85-108.
27. Noh, H. *Capacitated Schedule-Based Transit Assignment Using a Capacity Penalty Cost*. PhD dissertation. University of Arizona, 2013

28. Wolf, J., S. Schönfelder, U. Samaga, M. Oliveira, and K. Axhausen. Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1870, 2004, pp. 46-54.
29. GTFS document (https://developers.google.com/transit/gtfs/) and data(http://suntran.com/developers_feed.php) accessed January, 2016
30. Highway Capacity Manual. Transportation Research Board. *National Research*, 2010.
31. FHWA, Manual on Uniform Traffic Control Devices, 2009, revised in 2012.