Improving Last-Mile Connections to Transit: An Exploration of Data and Analysis Tools

CNU 25 | Seattle, Washington | May 3-6, 2017

Chris McCahill, PhD
State Smart Transportation Initiative, University of Wisconsin-Madison
1180 Observatory Drive, Suite 7122, Madison, WI 53706
608-262-7797
mccahill@ssti.us

Abstract

Last-mile connections are the critical links between an existing transit service and its potential users. New Urbanist principles such as connected street networks, walkable design, and transit oriented development (TOD) are important for improving these last-mile connections and ensuring the success of transit systems. However, to date there are no widely accepted methods for identifying gaps in transit accessibility, evaluating solutions, and prioritizing improvements. This paper explores the potential for two emerging fields – accessibility measurement and trip-making data – to help inform decisions focused on improving last-mile connections.

Based on three case studies of light rail transit corridors in Sacramento, this paper demonstrates the following: 1) the use of accessibility measures in scanning for poor connections and evaluating potential solutions, 2) the use of passive GPS data for understanding vehicle trip generation and travel patterns along transit corridors, and 3) the preliminary use of passive GPS data for understanding pedestrian travel patterns to and from transit stations. With some additional work, these tools will be useful for decision-makers as they work toward improving travel options, increasing transit ridership, coordinating with other key players, and communicating concepts with the public.

Introduction

New Urbanist research and practice focuses on building communities that foster and enable multimodal lifestyles. Transit of all forms is often central to this practice. Equally important, but often neglected, are the so-called first- and last-mile connections to transit services, which ensure that people can easily access transit once it is in place. New Urbanist principles such as connected street networks, walkable design, and transit oriented development (TOD) are important for improving these last-mile connections and ensuring the success of transit systems.

Unfortunately, the various functions that make transit successful – including transit operations, road design, and land development – often take place separately. As a result, we often see good transit service that is difficult to get to or walkable communities built around transit service that doesn't meet the needs of residents, employees, and visitors.

Coordinating these functions on a systemic level can be challenging. Comprehensive, long-range plans can help facilitate the process, but don't always affect decision-making at the right levels or among the right players. Key decisions are often made almost in isolation and without sufficient information to understand how the different parts will fit together.

This paper incorporates insights from two emerging fields – accessibility measurement and trip-making data – and explores their potential to inform decisions for improving last-mile connections. These insights may be valuable at all levels of decision-making and for many various players in the process. Their potential applications include helping to identify critical gaps in transit accessibility, evaluating solutions (both transportation- and land use-related), and prioritizing projects and investments to achieve the greatest impact. Most importantly, the approaches presented in this paper could perform particularly well in ways that conventional methods sometimes fall short.

Accessibility measures

Accessibility measures describe how well the existing transportation and land use system allows people to reach essential destinations. These measures let us characterize how well people can access transit stations, how well the transit system lets them reach other destinations, and how changes in infrastructure and land use affect accessibility.

Traditional measures of transportation system performance, such as travel speed and delay, are sometimes referred to as infrastructure-based accessibility measures (Geurs & van Eck, 2001), but they tell more about mobility (the ability to move around a system quickly) than accessibility (the ability to reach destinations). More accurate accessibility measures include cumulative opportunity and gravity-based measures (El-Geneidy & Levinson, 2006; Handy & Niemeier, 1997). Cumulative opportunity measures describe the number of destinations reachable within a certain time threshold from a particular location – e.g., jobs within 30 minutes. Gravity-based measures represent the sum of all reachable destinations multiplied by their impedance, which assumes that the utility of a destination decreases with travel time.

Cumulative opportunity and gravity-based measures are less commonly used mainly because they require considerable amounts of data and computing power – historically a major challenge. Accessibility estimates require network data, including network conditions, and land use data. They also require the ability to calculate travel times between all origins and destinations in a system, given the network and land use conditions.

Several tools now exist for making these types of calculations, including ArcGIS Network Analyst by Esri and OpenTripPlanner Analyst by Conveyal. Widely available sources of data include OpenStreetMap, General Transit Feed Specification (GTFS), Longitudinal Employer-Household Dynamics (LEHD), InfoUSA, HERE (formerly NAVTEQ), and the U.S. Census, among others.

This study relies on the Sugar Access tool, developed by Citilabs. The tool includes all necessary data, including networks and land uses from HERE, transit networks from GTFS, jobs from LEHD, and demographic information from the Census. The tool runs in Esri ArcMap and accessibility calculations are made on remote servers, to minimize local computing requirements.

Trip-making data

While accessibility measures let us evaluate the transportation-land use system, we can further assess potential gaps in transit accessibility by considering how people use that system. Common methods such as travel surveys, origin-destination studies using license plate matching or Bluetooth sensors, and travel demand models each suffer from limitations. Surveys and origin-destinations are resource-intensive and require substantial preparatory work. Travel demand models simulate travel patterns with varying accuracy and reliability, particularly regarding non-automobile trips.

Emerging data from mobile devices can potentially let us understand people's precise travel patterns in much larger numbers, with less effort, and at lower cost than those methods described above. Location data from cell phones, which is now commercially available, has been useful for producing origin-destination matrices and other output typically associated with travel demand models (Bonnel, Hombourger, Olteanu-Raimond, & Smoreda, 2015; Çolak, Alexander, Alvim, Mehndiretta, & Gonzalez, 2015; Huntsinger & Donnelly, 2014; Jarv, 2013; Jiang, Ferreira, & González, 2015; Widhalm, Yang, Ulm, Athavale, & Gonzalez, 2015). However, the location information provided by cellular signals is less useful for inferring precise location and route information.

Global positioning system (GPS) data, which offers higher spatial resolution, has been used to augment travel surveys since the late 1990s and can be used in conjunction with other data to identify trip segments, travel mode, and trip purpose (Gong, Morikawa, Yamamoto, & Sato, 2014; Shen & Stopher, 2014). Unlike cell phone data, however, passive GPS data is rarely used. Researchers often employ their own GPS units to monitor the travel patterns of study participants (Kang, Moudon, Hurvitz, Reichley, & Saelens, 2013; Wu, Dong, & Lin, 2014) or in some cases gain access to data from a third party, such as fitness tracking applications (Jestico, Nelson, & Winters, 2016; Musakwa & Selala, 2016). These data sources, which require travelers to opt in and sometimes take additional steps to log their data, result in smaller sample sizes and often introduce biases related to trip purpose and demographics.

This paper relies on commercially available GPS data, compiled from sources such as connected vehicles and mobile devices, and provided by StreetLight Data. These data are collected passively, then anonymized, processed, and aggregated before being provided to the research team. These data offer many of the advantages of cell phone data, but with higher spatial resolution. Until recently, this product was limited to vehicle trips, but a growing number of data sources allow the provider to report multiple modes, as this paper outlines. Trip volumes are reported using a relative index, which is then calibrated based on observed traffic counts.

In a later phase, this study will also incorporate a second source of trip-making data, provided by Teralytics. These data, like many previous trip-making studies, use cell phone location records. However, by analyzing the data and incorporating secondary information from GTFS, the data provider separates light rail transit trips from other trips and extrapolates the data to represent the entire study area population, based on knowledge of cell phone market penetration.

Study location and methods

This study examines light rail transit (LRT) corridors in Sacramento, California, and involves stakeholders from the Sacramento Area Council of Governments, the City of Sacramento, Sacramento Regional Transit, the California Department of Transportation, and the Sacramento Downtown Partnership.

For the trip-making analyses, the study area is divided into 250 traffic analysis zones (TAZs), a limit established through data purchase agreements. Census geographies (blocks, block groups, and tracts) form the rough basis of zones, but certain boundaries are redrawn to coincide with physical boundaries (e.g., rail lines) and distinct land use changes. Zones are more granular near transit corridors, including small zones around parking lots, and more aggregated farther from transit corridors.

This study focuses primarily on a subset of 96 zones that comprise four LRT catchment areas, pictured in Figure 1. For this paper, these four areas are taken together to represent one collective transit system catchment area.

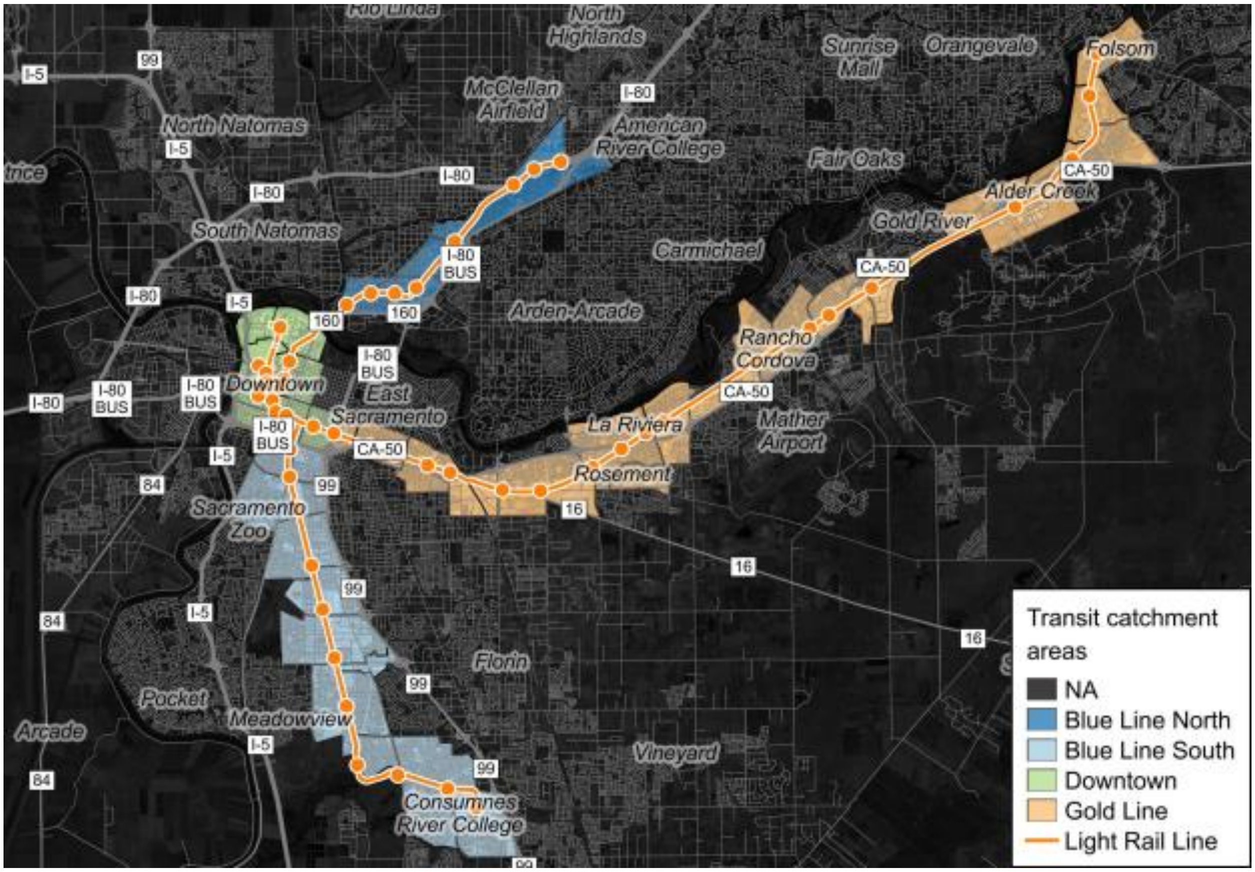

Figure 1. Transit catchment areas

The U.S. Census provides data on jobs in the LEHD Origin-Destination Employment Statistics (2014) and data on households in the American Community Survey (2011-2015). Jobs data are provided at the block level and aggregated to the block group level. Household data are provided at the block group level. Block group level data are reallocated to TAZs based on the portion of their total area that falls within each TAZ. Median values for age and income are estimated from categorical data by taking the midpoint of the bin within which the median value falls.

Accessibility analyses are conducted at the Census block level. Key measures used in this study are access to jobs during the morning period by transit (a gravity-based measure), and access to LRT stations by walking. These measures assume a walking speed of 2.8 miles per hour and use exponential travel time impedance functions derived from the 2009 National Household Travel Survey (NHTS). TAZ-level accessibility measures are estimated by assuming the average value of all the associated blocks. Blocks are assigned to a TAZ if a majority of their area (more than 50 percent) falls within that TAZ.

The ultimate goal of this study is to fully explore the potential applications of these data and the relationships among the different metrics they provide. This paper presents a preliminary investigation of the available tools and data to understand their potential applications in Sacramento. The results are presented as three case studies (Figure 2), developed over the course of a year through coordination between the research team and interested local stakeholders. This paper will inform later work, which will include more exhaustive evaluation and reporting.

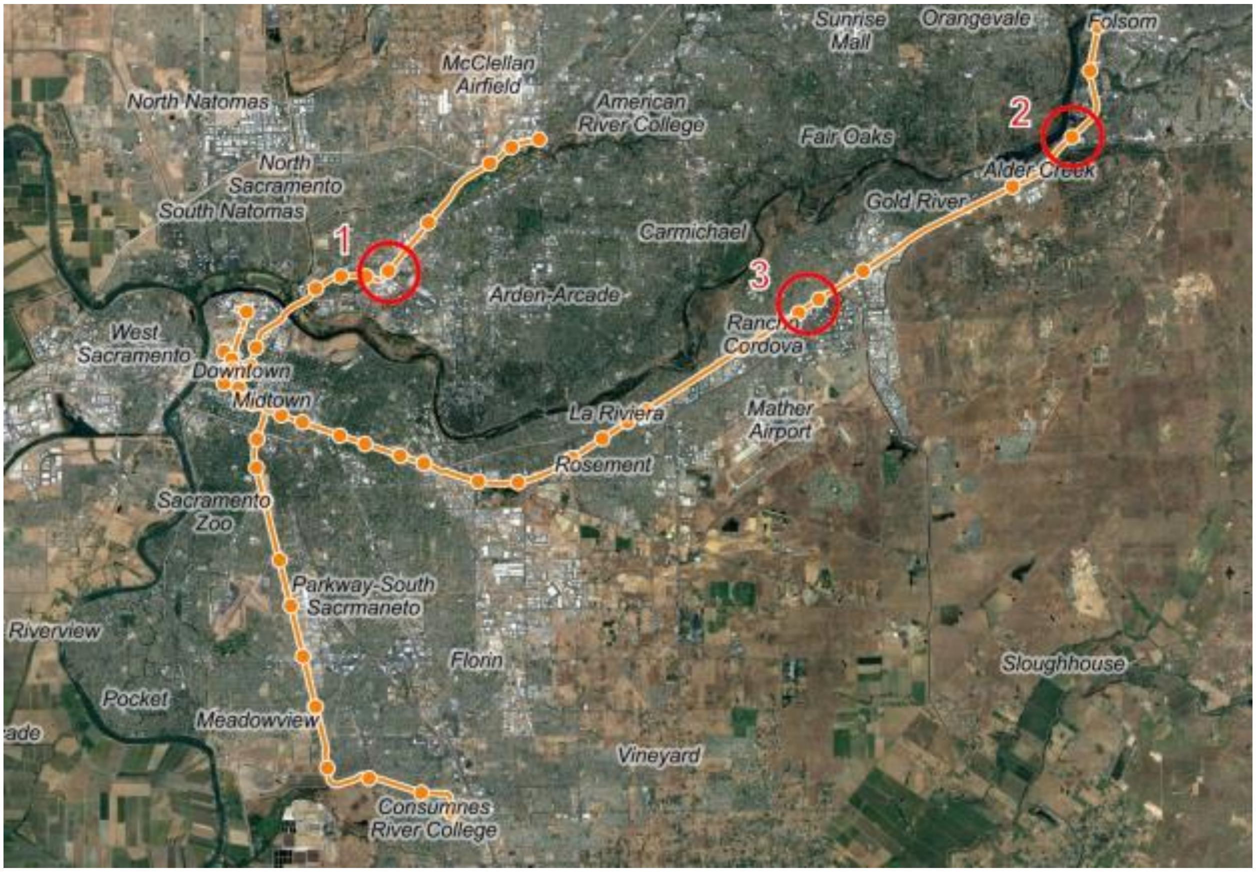

Figure 2. Case study locations

Case Study 1: Accessibility analysis at Swanston station

Several concepts were developed and applied in the accessibility analysis:

- Access to jobs by transit: The number of jobs reachable by transit (bus and LRT) during the morning period, where jobs are discounted based on a travel time impedance function. This measure is shown for the entire region in Figure 3.
- Station utility by walking: The utility of a station, or the likelihood that somebody will walk or bike to the station, based on a travel time impedance function.
- Potential utility improvement: The maximum potential increase in station utility, based on the difference between the straight-line distance and the current network distance.

- Potential impact score: A score from 0 to 100, which accounts for both the potential utility improvement of a zone and the number of residents and employees in the zone.

These measures are helpful in scanning existing conditions throughout the study area and identifying potential opportunities. Actual changes in accessibility and station utility can be calculated for specific projects and network changes, as demonstrated below. The following analyses focus on the area around Swanston station – roughly two miles north of the downtown along the Blue Line – where two of the ten highest potential impact scores are observed.

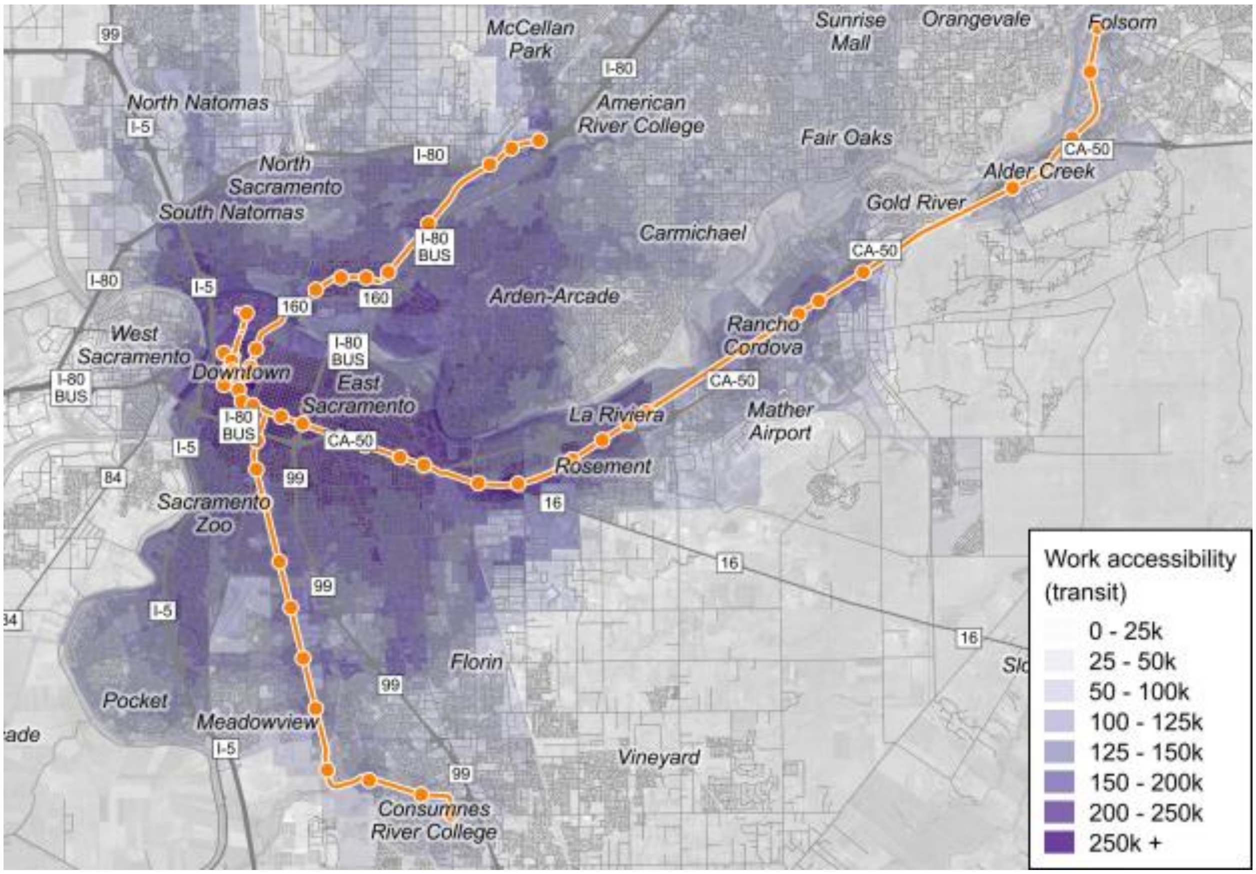

Figure 3. Access to jobs by transit

Figure 4 shows the current utility of Swanston station and other nearby stations by walking. For most of the stations in the area, the utility spreads out evenly into the surrounding neighborhoods. At Swanston station, however, most of the station utility is concentrated west of the station. This is because freight lines and the Capital City Freeway form barriers for neighborhoods to east, as shown in Figure 5. The nearest crossings are Arden Way to the south and El Camino Avenue to the north – neither of which is particularly pedestrian accessible.

As shown in Figure 6, a direct connection from Swanston station to the neighborhood directly to its east could improve the utility of the station for that neighborhood by 74 percentage points – from 9 to 83. Figure 7 depicts the potential impact of the connection, given the considerable number employees in that area. Additional connections to the southeast, across the Capital City Freeway, could also substantially impact the employees and residents there.

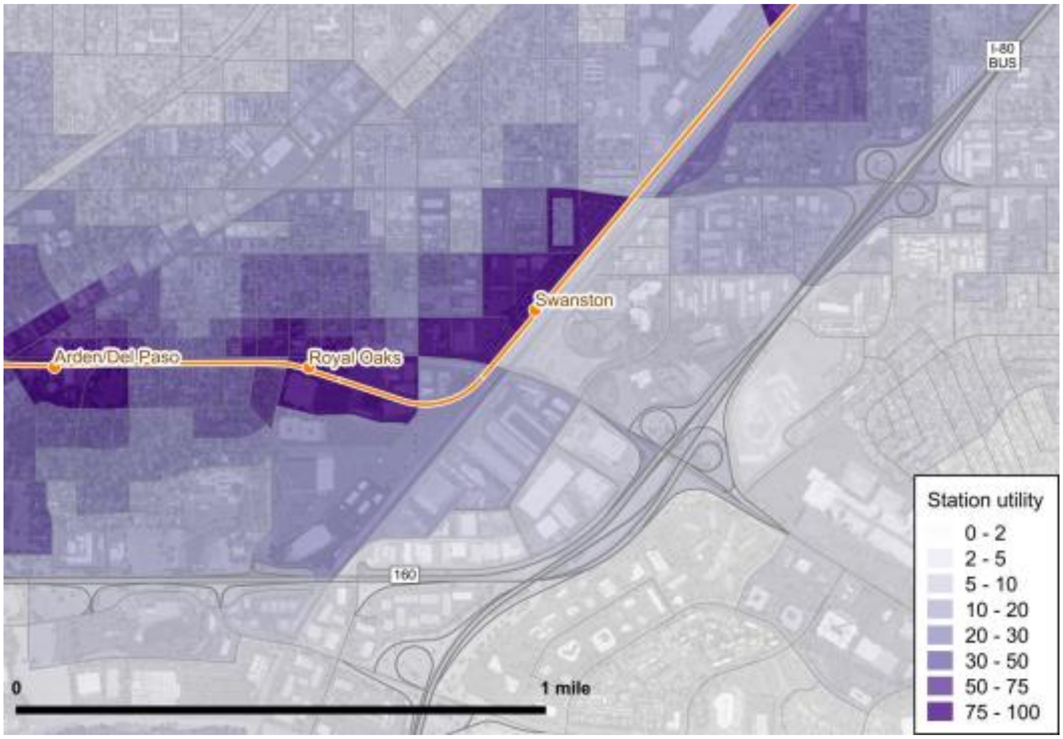

Figure 4. Station utility (0 to 100) around Swanston station

Figure 5. Existing conditions around Swanston station (depicted in orange)

Figure 6. Potential utility improvement (0 to 100) around Swanston station

Figure 7. Potential impact (0 to 100) around Swanston station

Proposed connections at Swanston station

The accessibility analysis shown above is much more than an abstract exercise. Recognizing the accessibility issues around Swanston station, the City of Sacramento proposed new connections to the east in a Transit Village Specific Plan from 2007. The plan, developed through a series of public workshops, is meant to enhance the area around the station as a highly-connected TOD and to maximize development potential. The plan proposes a dense network of bicycle and pedestrian connections including a bridge across the existing freight line to the east, as shown in Figure 8.

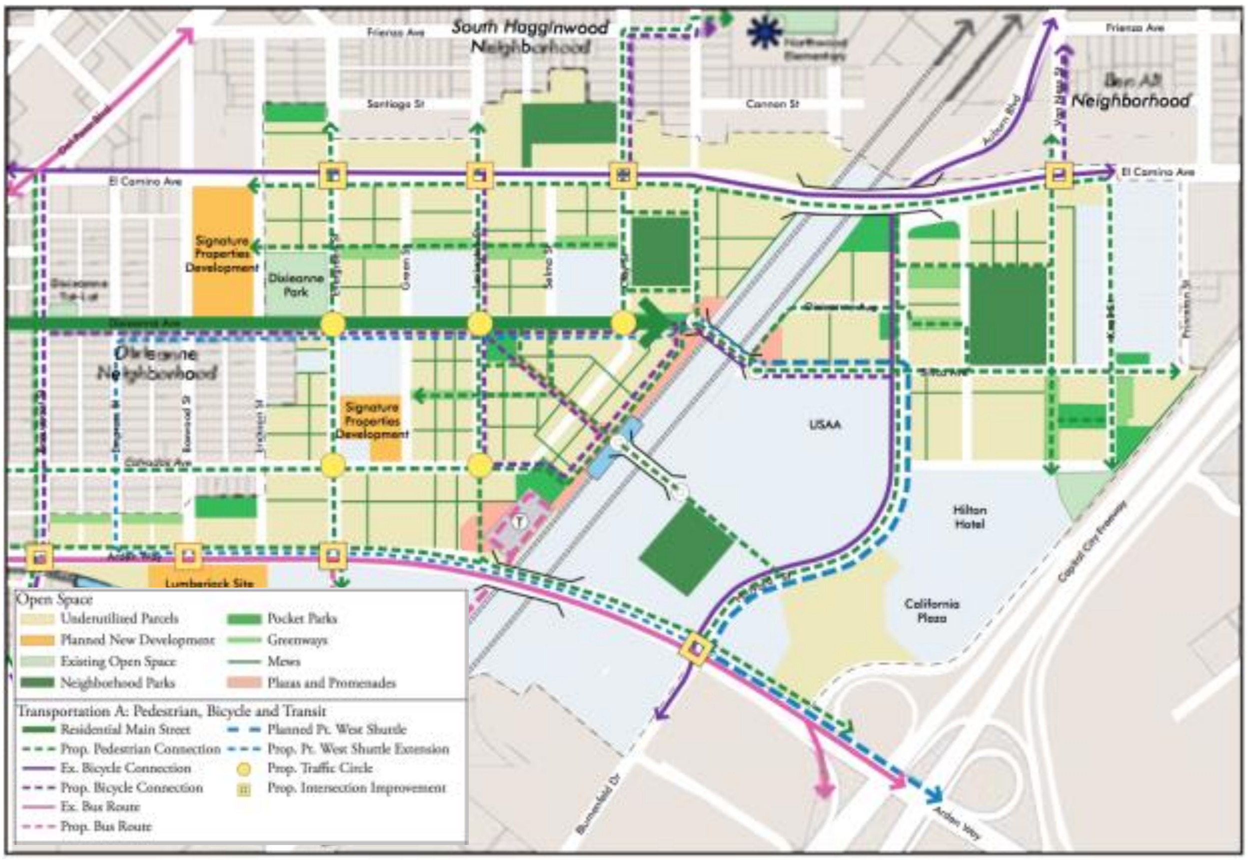

Figure 8. Proposed bicycle and pedestrian connections around Swanston station (source: City of Sacramento)

Using the same analysis tools described above, we can quantify the impacts of the new connections proposed by the City of Sacramento. Figure 9 shows how the proposed connections would improve travel times for those walking to the station, and therefore the station utility. From the neighborhood immediately to the east, travel time to the station decreases by 10 minutes. From neighborhoods further east, travel times decrease by around five minutes. Additional connections west of Swanston station also reduce travel times in the vicinity.

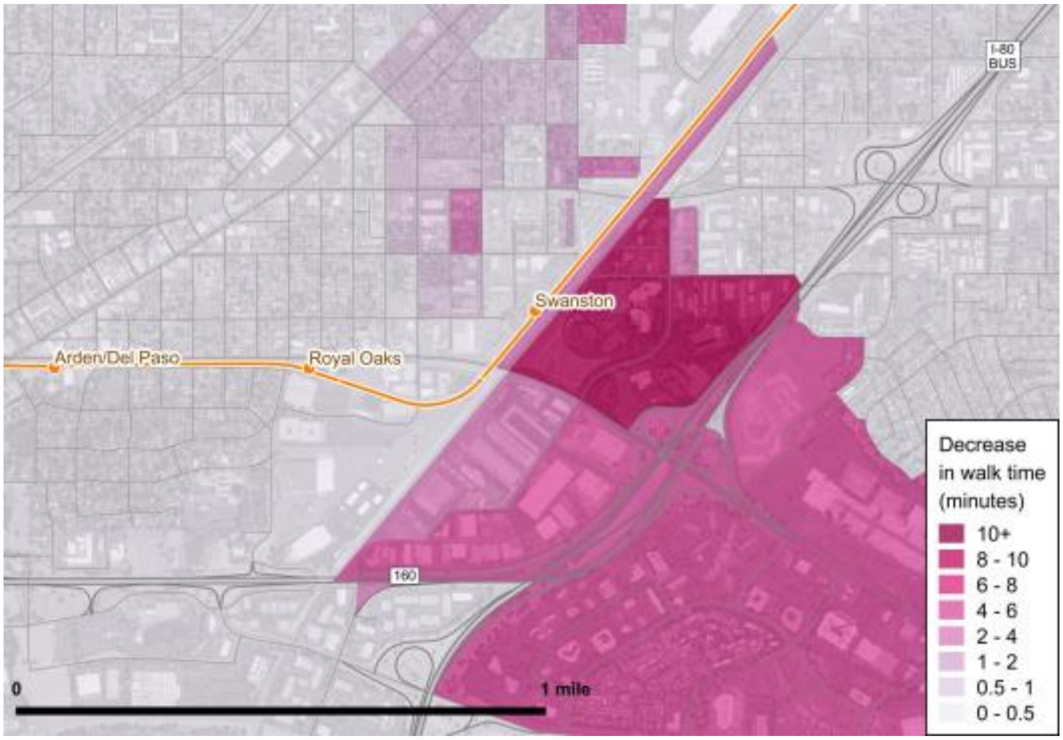

Figure 9. Walk time improvements due to proposed connections to Swanston station

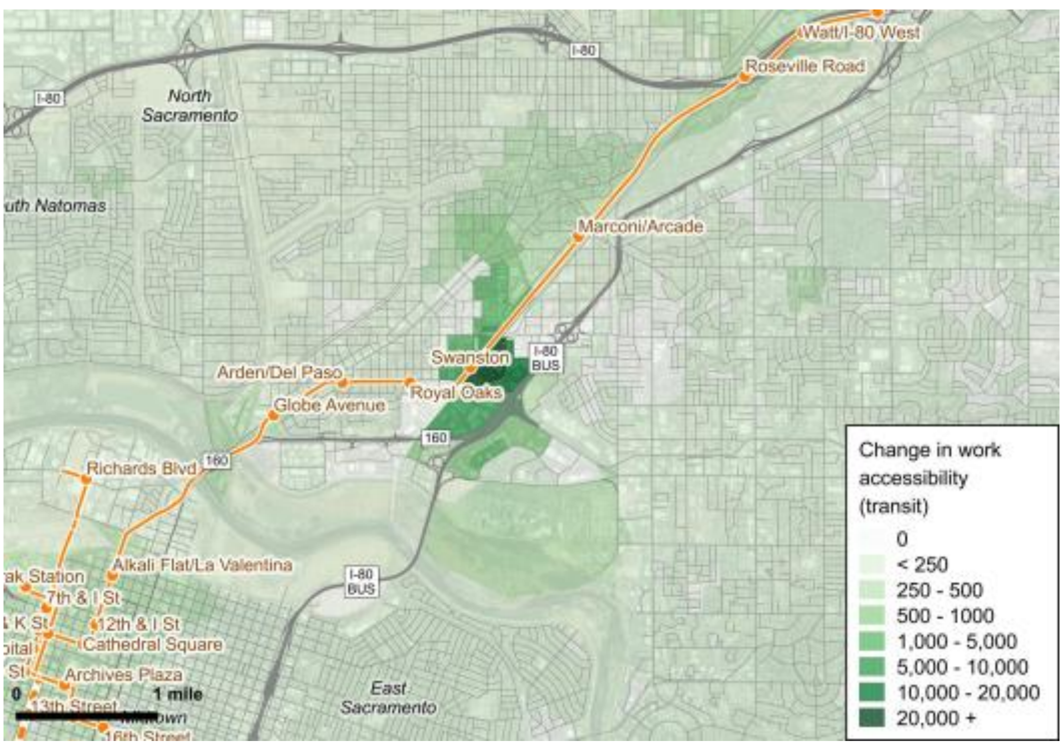

Figure 10. Improvements in access to jobs by transit due to proposed connections to Swanston station

As shown in Figure 10, these improved connections to Swanston station have a pronounced impact throughout the city. Those living immediately to the east gain access to an additional 15,000 to 30,000 jobs by transit from the improvements. Within a half-mile of the station, the average increase is 1,600 jobs. Because the connections also improve access to jobs near the station, the impacts are essentially felt citywide. In total, roughly 33,000 households gain access to an additional 250 jobs or more. The most pronounced changes occur as far as five miles to the north and almost 10 miles south (not pictured).

Case Study 2: Vehicle trip-making at Iron Point station

The vehicle trip-making analysis is conducted using GPS data representing TAZ-to-TAZ flows of personal vehicles. Average weekday flows are reported as a relative “StreetLight Index” value. Based on a regression analysis of average weekday traffic on 30 highway segments, actual trip volumes can be estimated by multiplying the StreetLight Index by 0.85 ($R^2 = 0.9$; p-value < 0.01).

The analyses presented here are meant to depict the number of potential transit trips generated along each transit corridor, as presented in Figure 1. Potential transit trips are defined as vehicle trips that begin and end within a transit catchment area. shows the total daily personal vehicle trips originating in each zone, whereby the entire system is treated as a single catchment area. This represents trips along each individual corridor and trips that would require a transfer Downtown. The highest trip generation rates are observed south of Downtown and east along the Gold Line.

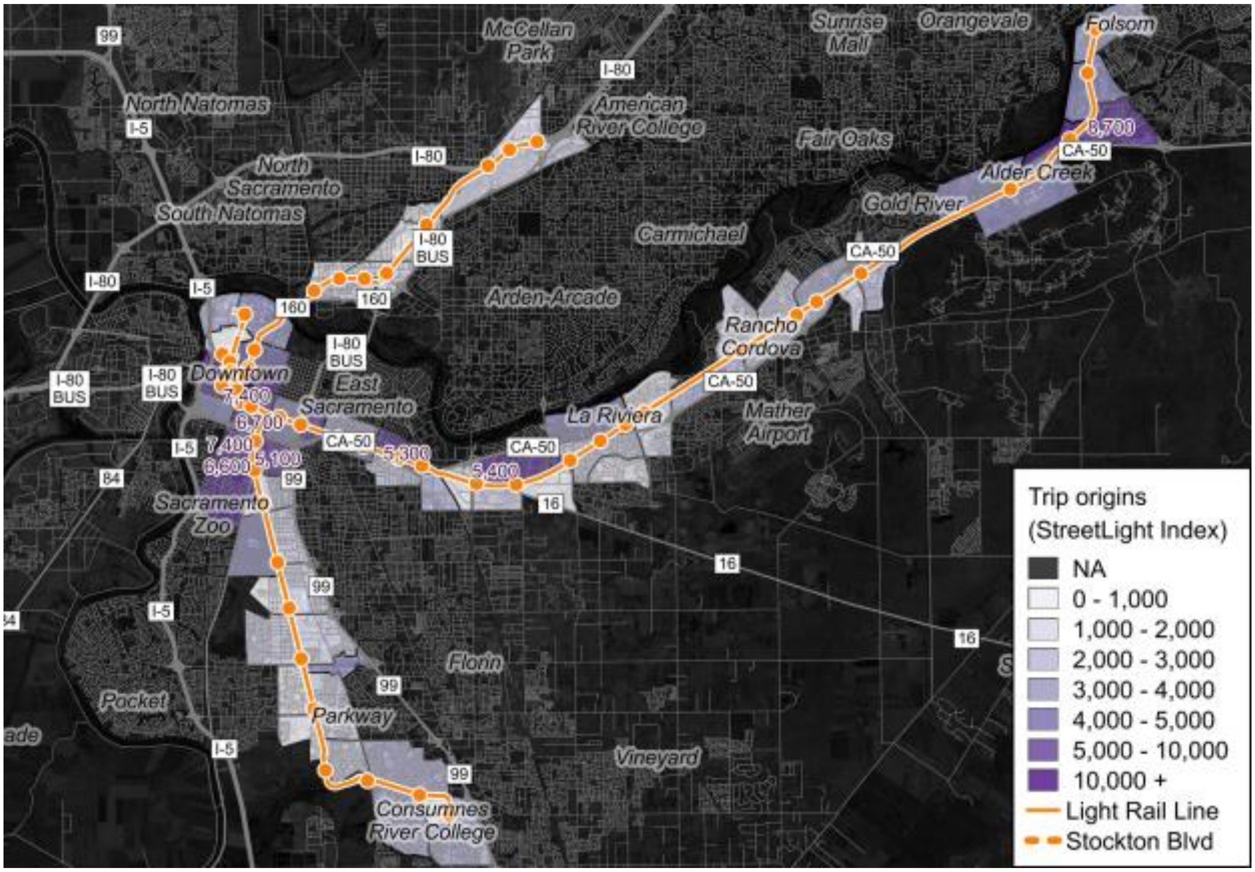

Figure 11. Daily vehicle trips generated along the LRT system plus Stockton Boulevard

Based on the analysis described above, the highest number of StreetLight Index trips (8,700) are generated in the zone surrounding Iron Point station, at the east end of the Gold Line just before it bends northward. To better understand the potential transit ridership of these trips, Figure 12 shows the trip destinations. Most of these trips end near the three closest stations (Hazel, Glenn, and Folsom). These short trips might be difficult to shift to transit unless service became more frequent and better connections to the station are provided. However, roughly 1,800 trips end elsewhere along the Gold Line, including roughly 400 trips ending Downtown.

Demographic analysis offers some additional insight about these trips. The origin zone contains 1,500 jobs and 1,400 households. Households incomes, home ownership rates, and vehicles ownership rates are considerably higher than throughout the rest of the study area. The estimated median household income is $112,500, 74 percent of households are owned, and the average household has 1.9 vehicles.

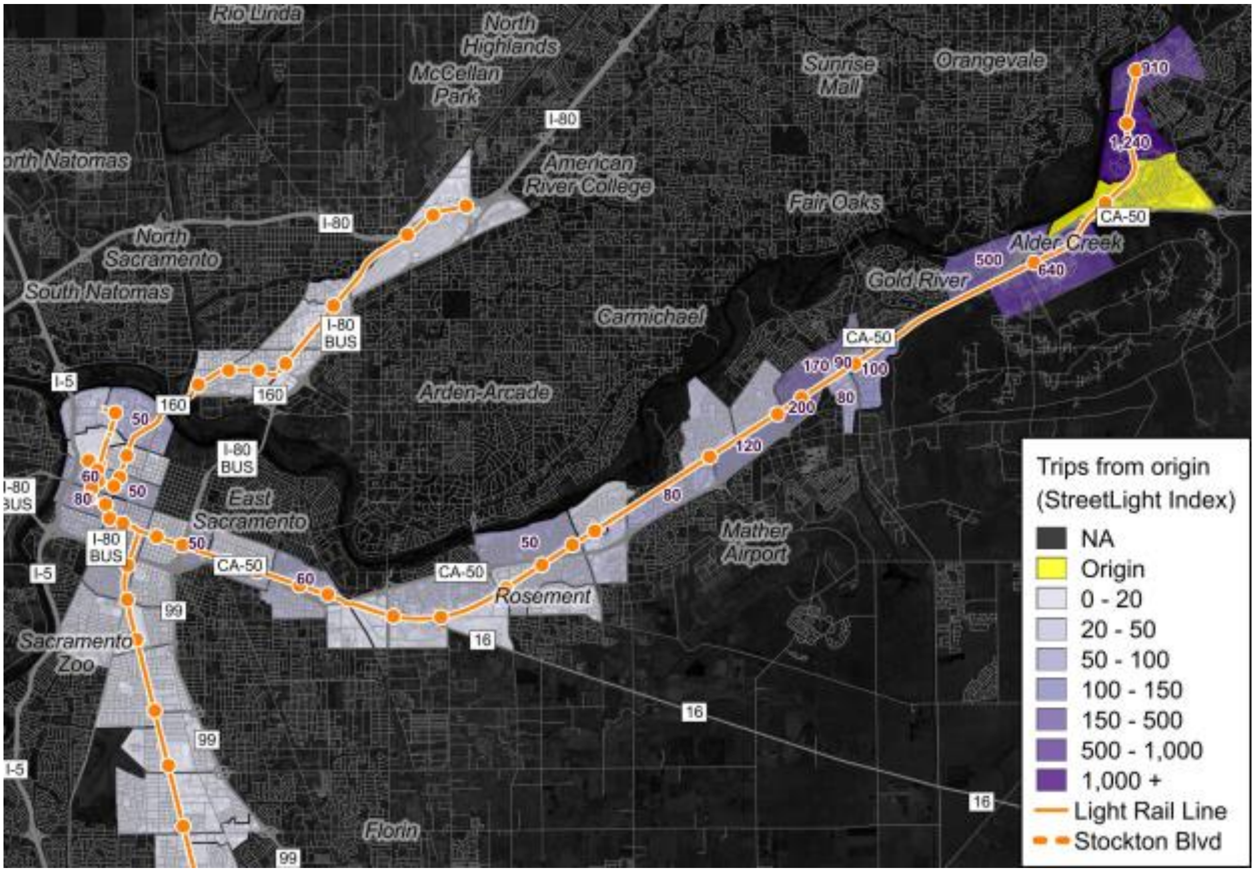

Figure 12. Daily vehicle trip destinations from origin zone (in yellow)

Case Study 3: Pedestrian trip-making at Zinfandel and Cordova Town Center

An essential outcome of this project is the early development of methods for classifying observed GPS traces by mode. Prior to this study, the existing data only let us characterize personal and commercial vehicle trips. This was possible because a large amount of data comes from in-vehicle GPS units. However, there is also abundant data from handheld GPS devices. The ability to detect pedestrian and bicycle trips using passive data is particularly valuable because of the profound scarcity of this type of data and the importance of planning and designing for those types of trips.

A considerable amount of research has recently focused on methods for identifying travel mode from GPS traces based on characteristics such as travel speed. These efforts typically rely on machine learning and other advanced techniques, and many achieve modal recognition on the order of 90 percent accuracy (Gong et al., 2014; Shen & Stopher, 2014; Zong, Bai, Wang, Yuan, & He, 2015). Using similar techniques and a rich source of training data from outside of the study area, StreetLight Data recently developed preliminary modal recognition techniques. As of this study, those methods and metrics are still in a trial phase.

While the penetration rates for this type of data are still quite low, they are nonetheless useful for understanding the relative number of trips between light rail stations and nearby analysis zones. The following analysis is based on two months of data from June and July of 2016. The data includes all trips identified as likely pedestrian trips, excluding those shorter than 500 meters – a temporary artifact of the underlying vehicle trip classification methods.

Figure 13 and Figure 14 show the relative number of pedestrian trips to and from Zinfandel and Cordova Town Center stations, respectively. The number of trips are scaled between 0 and 100. These data show that Zinfandel station attracts considerably more foot traffic than Cordova Town Center, but both stations attract most of their trips from adjacent zones to the northwest and southeast. Zinfandel station attracts more trips from the south and the west, while Cordova Town Center attracts more trips from the two zones to its north.

The large concentration of activities (Figure 15) and high station utility (Figure 16) north of the light rail line explains the large number of trips to and from those areas. However, the large number of trips to and from zones to the south, including those across the Lincoln Highway (US-60), is unexpected. One possible explanation is that these are trips to and from jobs, as shown in Figure 17. In addition, people living in those southern zones are somewhat younger (median adult under 42) and have higher incomes (median above $50,000), compared to other nearby zones.

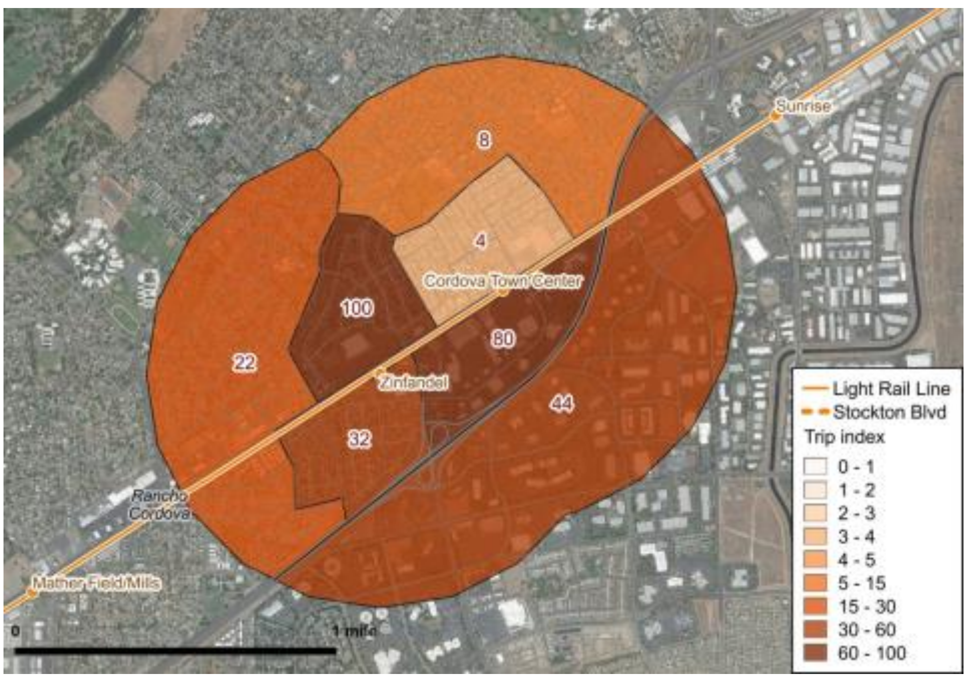

Figure 13. Relative walking trips to and from Zinfandel station platform

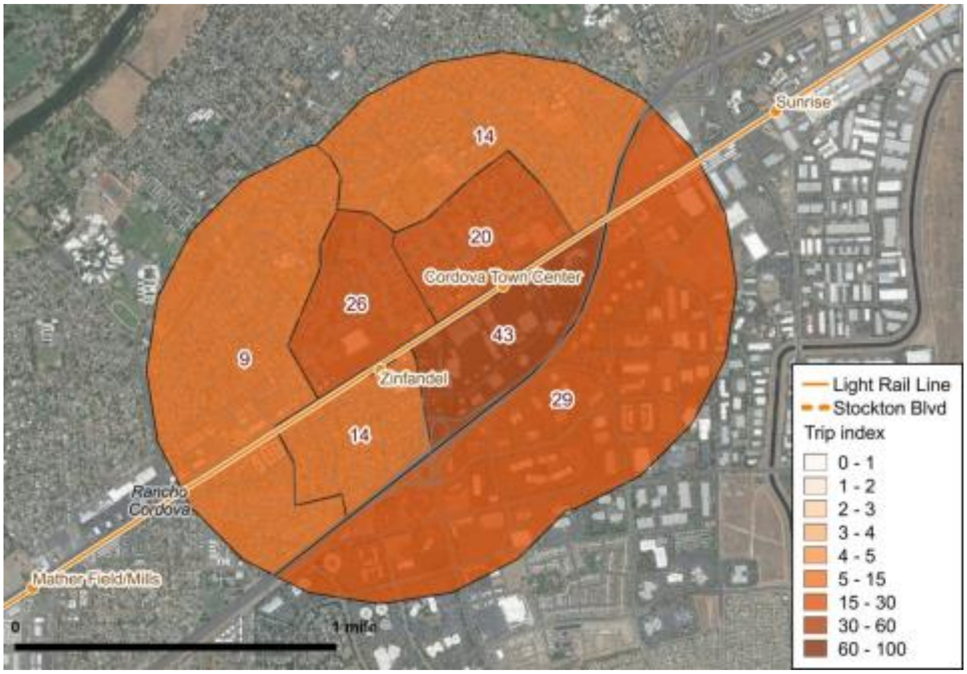

Figure 14. Relative walking trips to and from Cordova Town Center station

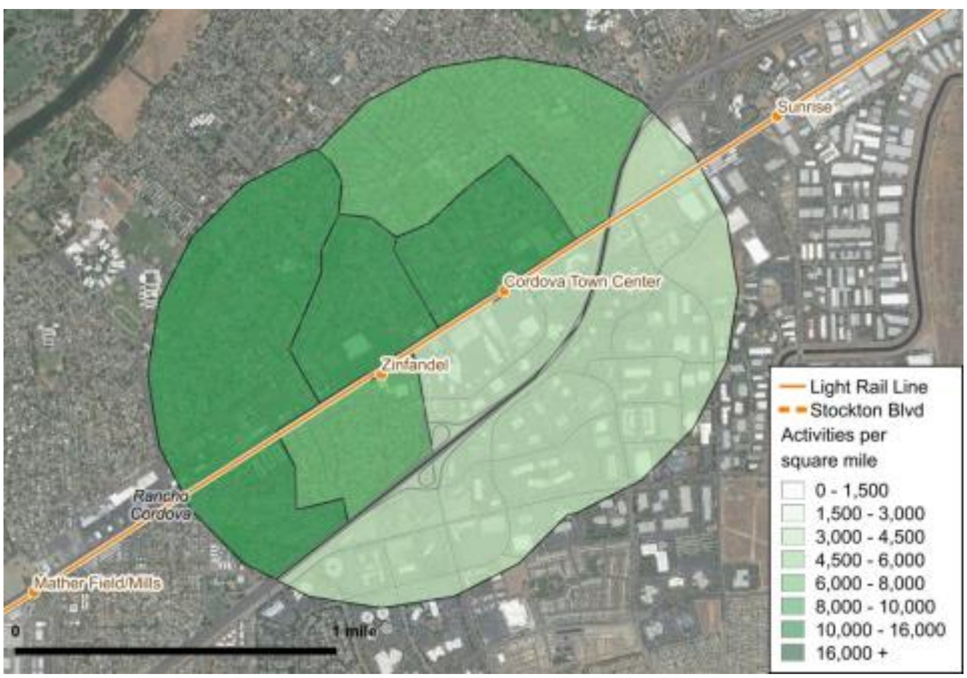

Figure 15. Activities (population + jobs) per square mile around Zinfandel and Cordova Town Center stations

Figure 16. Station utility around Zinfandel and Cordova Town Center stations

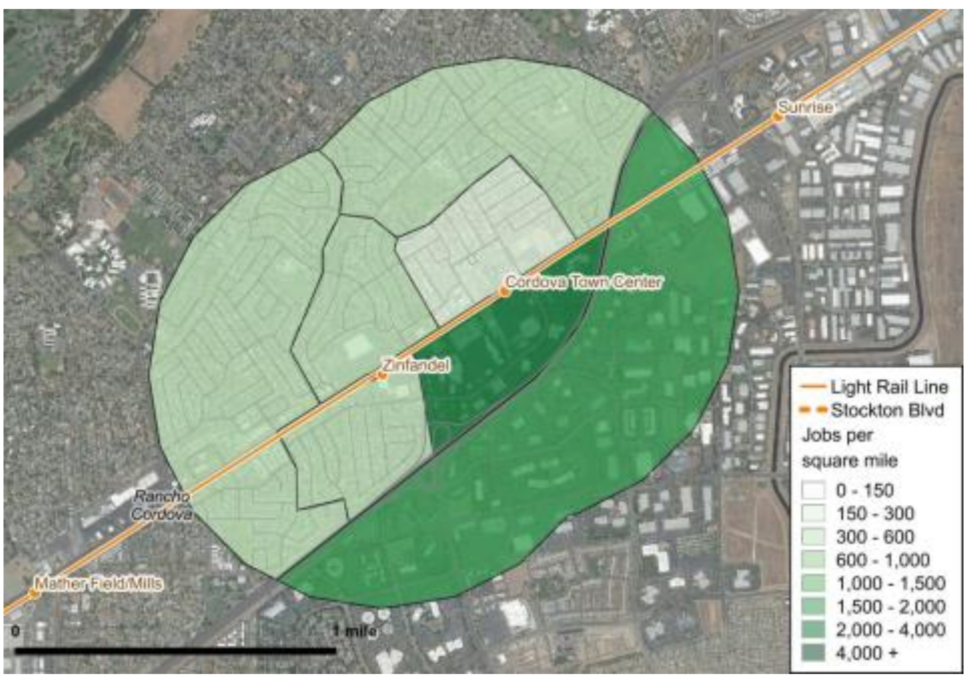

Figure 17. Jobs per square mile around Zinfandel and Cordova Town Center stations

Conclusions

The above case studies highlight three key opportunities arising out of newly available data and analysis tools: 1) using accessibility measures to identify gaps in transit accessibility, prioritize efforts to improve accessibility, and evaluate proposals for doing so; 2) using passive GPS data to identify potential transit riders and infer opportunities to shift certain trips from personal vehicle to LRT; and 3) using passive GPS data to understand pedestrian trip-making around transit stations and identify or validate accessibility gaps.

Admittedly, more work is needed to better understand the shortcomings of these methods and, ultimately, to standardize practices that can be applied broadly and consistently. This paper, however, presents an important first step.

First, this preliminary work demonstrates that accessibility scans can properly identify known gaps in transit accessibility and potentially lesser known ones. Given a concrete proposal to improve last-mile connections, such as the Swanston station area improvements, accessibility measures provide a means for quantifying the impacts, which also makes them useful for comparing and prioritizing different proposals. Moreover, many of the accessibility metrics presented here were used to communicate successfully to a wide variety of stakeholders throughout this study, suggesting that they are particularly useful for conveying existing and proposed conditions to a broad audience.

Second, passive GPS data provides valuable information about potential transit users that would be difficult to ascertain through other means. The information can be obtained quicker and easier than through surveys or conventional travel studies and, in contrast to travel demand models, it represents real on-the-ground behavior. It seems to provide useful information about trip generation and travel patterns anywhere along a transit corridor, although further validation may be needed to ensure the reliability of the observed data. This type of information pertaining specifically to non-transit users can be exceptionally challenging for transit providers to obtain.

Third, more work may be needed before passive GPS data provides reliable information about pedestrian trips, but preliminary data suggests this information will be particularly helpful for understanding people's travel patterns near transit stations. The data presented here seem to confirm some of the findings from earlier accessibility analyses – i.e., that people are more likely to walk to transit stations from better connected neighborhoods. However, it also presents new areas of inquiry. For example, more people than expected (potentially workers) appear to be accessing transit stations from certain poorly connected neighborhoods. This poses questions about the data's validity and raises concerns about the immediate need for accessibility improvements. Like the vehicle trip data described above, this information would be exceptionally difficult to obtain through other means.

Through further development and standardization, the suite of tools presented here can help considerably in efforts to improve travel options and increase transit ridership. In addition to the information they provide, the data and metrics can help bridge existing gaps in the decisions made by transit service providers, infrastructure providers, land use planners, and developers, while providing useful new ways of communicating with the public.

Acknowledgements

The author thanks TransitCenter, the Barr Foundation, and Planet Bike for their support of this research and many partners for their ongoing involvement in this work: the Lincoln Institute of Land Policy, the Sacramento Council of Regional Governments, the City of Sacramento, Sacramento Regional Transit, the California Department of Transportation, and the Sacramento Downtown Partnership.

References

Bonnel, P., Hombourger, E., Olteanu-Raimond, A. M., & Smoreda, Z. (2015). Passive mobile phone dataset to construct origin-destination matrix: Potentials and limitations. *Transportation Research Procedia*, *11*, 381–398. http://doi.org/10.1016/j.trpro.2015.12.032

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiretta, S. R., & Gonzalez, M. C. (2015). Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations and Opportunities. *TRB 2015 Annual Meeting*, 1–17. http://doi.org/10.3141/2526-14

El-Geneidy, A. M., & Levinson, D. M. (2006). *Access to Destinations: Development of Accessibility Measures*. Minneapolis, Minnesota.

Geurs, K. T., & van Eck, J. R. R. (2001). *Accessibility measures: review and applications. Evaluation of accessibility impacts of land-use transportation scenarios, and related social and economic impact*. *RIVM Report*. Netherlands. Retrieved from https://rivm.openrepository.com/rivm/handle/10029/9487

Gong, L., Morikawa, T., Yamamoto, T., & Sato, H. (2014). Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. *Procedia - Social and Behavioral Sciences*, *138*(0), 557–565. http://doi.org/10.1016/j.sbspro.2014.07.239

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A*, *29*(7), 1175–1194. http://doi.org/10.1068/a291175

Huntsinger, L. F., & Donnelly, R. (2014). Reconciliation of Regional Travel Model and Passive Device Tracking Data. *2014 TRB Annual Meeting Compendium of Papers*, *250*(January), 1–12.

Jarv, O. (2013). Mobile phone based data in human travel behaviour studies: new insights from a longitudinal perspective. *Baltic Journal of Economics*, *13*(2), 145–147.

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, *52*, 90–97. http://doi.org/10.1016/j.jtrangeo.2016.03.006

Jiang, S., Ferreira, J., & González, M. C. (2015). Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data : A Case Study of Singapore. *IEEE Transactions on Big Data*, 1–13. http://doi.org/10.1109/TBDATA.2016.2631141

Kang, B., Moudon, A. V., Hurvitz, P. M., Reichley, L., & Saelens, B. E. (2013). Walking objectively measured: Classifying accelerometer data with GPS and travel diaries. *Medicine and Science in Sports and Exercise*, *45*(7), 1419–1428. http://doi.org/10.1249/MSS.0b013e318285f202

Musakwa, W., & Selala, K. M. (2016). Mapping cycling patterns and trends using Strava Metro data in the city of Johannesburg , South Africa. *Data in Brief*, *9*, 898–905. http://doi.org/10.1016/j.dib.2016.11.002

Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, *34*(3), 316–334. http://doi.org/10.1080/01441647.2014.903530

Widhalm, P., Yang, Y., Ulm, M., Athavale, S., & Gonzalez, M. C. (2015). Discovering urban activity patterns in cell phone data. *Transportation*, *42*(4), 597–623. http://doi.org/10.1007/s11116-015-9598-x

Wu, X., Dong, J., & Lin, Z. (2014). Cost analysis of plug-in electric vehicles using GPS-based longitudinal travel data. *Energy Policy*, *68*, 206. http://doi.org/10.1016/j.enpol.2013.12.054

Zong, F., Bai, Y., Wang, X., Yuan, Y., & He, Y. (2015). Identifying Travel Mode with GPS Data Using Support Vector Machines and Genetic Algorithm. *Information*, *6*(2), 212–227. http://doi.org/10.3390/info6020212