TRCLC 16-10
April 30, 2019

A Constraint-Based Bicycle Origin-Destination Estimation Procedure

FINAL REPORT

Zhaocai Liu, Ziqi Song, Ph.D., Anthony Chen, Ph.D.
Department of Civil and Environmental Engineering
Utah State University
Logan, UT 84322

Transportation Research Center
for Livable Communities

Western Michigan University | University of Texas at Arlington | Utah State University | Wayne State University | Tennessee State University

Technical Report
Documentation Page

1. Report No. TRCLC 16-10	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle A Constraint-Based Bicycle Origin-Destination Estimation Procedure		**5. Report Date** April 30, 2019 **6. Performing Organization Code** N/A
7. Author(s) Zhaocai Liu, Ziqi Song, Anthony Chen		**8. Performing Org. Report No.** N/A
9. Performing Organization Name and Address Department of Civil and Environmental Engineering Utah State University Logan, UT 84322		**10. Work Unit No. (TRAIS)** N/A **11. Contract No.** TRCLC 16-10
12. Sponsoring Agency Name and Address Transportation Research Center for Livable Communities (TRCLC) 1903 W. Michigan Ave., Kalamazoo, MI 49008-5316		**13. Type of Report & Period Covered** Final Report 9/1/2016 - 4/30/2019 **14. Sponsoring Agency Code** N/A

15. Supplementary Notes

16. Abstract
Cycling is an active, green transportation mode that improves environmental sustainability and the livability of urban communities. Promoting cycling also has significant public health benefits. Despite the increasing importance of cycling as a transportation mode, it is often ignored in traditional transportation planning procedures. Origin-destination (O-D) matrix estimation methods have focused on estimating O-D matrices from link traffic counts for motorized vehicles, which are regularly collected by transportation agencies for traffic monitoring purposes. However, traffic monitoring of non-motorized traffic is not as comprehensive as motorized traffic monitoring in the United States. Hence, O-D matrix estimation methods developed for motorized traffic cannot be directly used to estimate bicycle O-D trip matrices.

This project proposes a bicycle O-D estimation procedure that is flexible and can be adjusted to different levels of data availability and quality. Bicycle data that are useful for bicycle O-D estimation are first explored. A constraint-based bicycle O-D estimation procedure that utilizes bicycle data from multiple sources is then proposed. A case study is also conducted to demonstrate the proposed methodology. The results demonstrate that in practice, the proposed bicycle O-D estimation procedure is a promising tool.

17. Key Words Bicycle O-D estimation, Path flow estimator, Crowdsourcing bicycle data, Bicycle count data	**18. Distribution Statement** No restrictions.

19. Security Classification - report Unclassified	**20. Security Classification - page** Unclassified	**21. No. of Pages**	**22. Price** N/A

Disclaimer

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the information presented herein. This publication is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. This report does not necessarily reflect the official views or policies of the U.S. government, or the Transportation Research Center for Livable Communities, who assume no liability for the contents or use thereof. This report does not represent standards, specifications, or regulations.

Acknowledgments

This research was funded by the US Department of Transportation through the Transportation Research Center for Livable Communities (TRCLC), a Tier 1 University Transportation Center. The authors would like to thank Salt Lake City and Utah Department of Transportation for sharing bicycle data for our analysis.

Table of Contents

Chapter 1: Introduction

Cycling is an active, green transportation mode that improves environmental sustainability and the livability of urban communities. Promoting cycling also has significant public health benefits. For example, studies (Gordon- Larsen, et al., 2009; Pucher et al., 2010) have indicated that states with high bicycle and walking mode shares generally have low obesity-related health issues. In addition, Dill (2009) demonstrated that a well-connected network of bicycle-specific infrastructure is important in helping adults meet the recommended levels of physical activity by encouraging more cycling among adults.

Although the automobile mode still dominates the transportation system in the United States, increased bicycle use has been observed in recent decades. The National Households Travel Survey (NHTS) (Kuzmyak et al., 2014) reveals that bicycle mode share increased from 0.7% in 1977 to about 1% in 2009. Total bicycle trips increased from 1 million to more than 4 million annually. Additionally, the American Community Survey (ACS) finds that bicycle commuting increased by 61.6% from 2008 to 2012, a larger percentage increase than that of any other commuting mode (McKenzie, 2014).

Despite the increasing importance of cycling as a transportation mode, it is often ignored in traditional transportation planning procedures. A key element in transportation planning is an origin-destination (O-D) trip matrix that describes the travel demand between each pair of O-Ds. An O-D matrix is often obtained from household or roadside surveys, which are generally labor-intensive and hence expensive to carry out. As a result, O-D matrix estimation methods have focused on estimating O-D matrices using link traffic counts for motorized vehicles (Nie and Lee, 2002), which are regularly collected by transportation agencies for traffic monitoring purposes. However, traffic monitoring for non-motorized traffic is not as comprehensive as motorized traffic monitoring in the United States (FHWA, 2013). Hence, O-D matrix estimation methods developed for motorized traffic cannot be directly used to estimate bicycle O-D trip matrices.

The major technical challenge in bicycle monitoring is that cyclists are less confined to fixed lanes than motorized vehicles (FHWA, 2013). Cyclists may ride outside designated bicycle lanes. When there is no designated bicycle lane, cyclists may also ride in motorized lanes or on sidewalks. Therefore, developing reliable bicycle sensors and deploying them to the right locations represent technical barriers to large-scale, continuous monitoring of bicycle traffic. Most communities still rely on manual bicycle counting when they conduct bicycle-related studies, even though the cost for each data point is relatively high when using the manual counting method versus automated counting devices.

Given the rising interest in bicycle travel across the nation, more accurate and cost-effective bicycle sensors have been developed (FHWA, 2014). As a result, more bicycle count data are becoming available. Further, emerging technologies, such as crowdsourcing cell phone applications and bicycle sharing programs, are becoming popular, and they provide additional data sources for bicycle studies. Therefore, it is critically important to develop an innovative bicycle O-D matrix estimation procedure that utilizes bicycle data from multiple sources. With more reliable bicycle O-D matrices, transportation agencies will be able to develop a sound methodology to conduct bicycle demand forecasting and planning, leading to more effective bicycle facilities and improvements in safety for cyclists.

The goal of this research project is to propose a bicycle O-D estimation procedure that is flexible and can be adjusted to different levels of data availability and quality. The study has three major objectives, which are as follows: (1) to explore bicycle data that are useful for bicycle O-D estimation, (2) to develop a constraint-based bicycle O-D estimation procedure that utilizes bicycle data from multiple sources, and (3) to conduct a case study using the proposed methodology. The remainder of this report is organized as follows. Chapter 2 provides a brief introduction about bicycle data from multiple sources. Chapter 3 presents our methodology of using the Path Flow Estimator (PFE) to estimate bicycle O-D matrices. In Chapter 4, a real-world case study is conducted. Finally, Chapter 5 concludes the report.

Chapter 2: Bicycle Data from Multiple Sources

There are multiple sources for bicycle data, including traditional bicycle count data, emerging crowdsourcing bicycle data, and data from bike-sharing programs. This chapter introduces each of the above bicycle data sources and discusses their applicability in bicycle O-D trip estimation.

2.1 Bicycle Count Data

Depending on the method used to collect data, bicycle count data can be categorized into manual count data (as shown in Figure 2.1) and automated count data (as shown in Figure 2.2). Bicycle count data can also be grouped by the locations at which the data were collected. Screen line counts record all bicycles that cross an invisible line on the road. Intersection counts track bicycle turning movement and crossing volumes at intersections.

FIGURE 2.1 Manual collection of bicycle data (Ainsley, 2018)

FIGURE 2.2 An automatic bicycle counter (Andersen, 2015)

A recent survey (FHWA, 2014) revealed that manual counting remains the most prevalent bicycle counting method in the United States. Manual counting is relatively easy to implement and is extremely flexible. Additional information other than traffic count, such as gender and helmet use, can also be gathered. There is no need to install detectors in the public right of way; hence, it entails less technical capacity and does not require the permission of regulatory authorities. By strategically establishing counting locations across the road network, manual counting can cover large areas to investigate the spatial distribution of bicycle travel demand. Nevertheless, the shortcomings of manual counting are notable. The accuracy of manual counts is subject to human error/bias and fatigue of the data collector. Therefore, manual counting is only suitable for short-term counting. The National Bicycle and Pedestrian Documentation (NBPD) Project has made an effort to provide a model for the consistent collection of data that can be employed nationwide. Many transportation agencies and governments, including Salt Lake City, regularly conduct manual bicycle counting.

In recent years, more reliable and affordable bicycle detectors have been developed, making automated bicycle counting possible. Automated counting is suitable for continuous, long-term bicycle monitoring. It can provide crucial information on the temporal variations of bicycle demand. However, installing automated counting devices requires significant financial and technical resources. For example, most bicycle detectors must be calibrated to achieve accurate counting results. Many communities use a combination of both counting methods to fulfill their data collection needs (FHWA, 2013).

2.2 Emerging Crowdsourcing Bicycle Data

As smart-phone technologies have become increasingly popular, people have begun to embrace the idea of recording their bicycle trips using recreational and fitness smart-phone applications, e.g., Strava and MapMyRide, which provide cyclists the ability to retrieve information about their physical activities. Transportation agencies have also recognized the opportunity to leverage this emerging technology and have developed smart-phone applications with the goal of encouraging citizens to record and share their bicycle trip data with these agencies. In 2009, the San Francisco County Transportation Authority (SFCTA) developed an open-source smart-phone application called CycleTracks (Hood et al., 2011). The application has been customized and used by several other North American cities, including Atlanta, Georgia and Reno, Nevada. Bicycle trip data made available through smart phones are usually termed crowdsourcing bicycle data. Brabham (2012) defined crowdsourcing as an "online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals." Because smart cell phones are equipped with Global Positioning Systems (GPSs), bicycle trip data recorded by smart-phone applications can provide detailed information about bicycle trip trajectories, which is particularly valuable in studying cyclists' route choice behavior. Potential problems associated with crowdsourcing bicycle data include sample bias due to participation barriers, demographic bias, and other issues, such as quality control and recruitment.

For the estimation of bicycle O-D trip matrices, crowdsourcing bicycle data can be treated as a sampling of all bicycle trips. However, the sampling rate of crowdsourcing data generally is unknown. Therefore, crowdsourcing data are of limited use in estimating bicycle O-D matrices. Crowdsourcing data can nonetheless be used to complement available bicycle count data. For example, the aggregate bicycle link volume obtained from crowdsourcing data can be used to establish the minimum volume at a location. This information can be incorporated into the procedure used to estimate bicycle O-D matrices.

2.3 Bike-Sharing Data

Bike-sharing programs record the origin and destination of a bicycle trip when a bicycle is checked in or out at a bike-share station. Some programs also implant their bicycles with GPS devices for better tracking. Bike-sharing data can be used to estimate origin-destination patterns, as well as possible routes and travel times (FHWA, 2014). However, bike-sharing data share some of the same issues with crowdsourcing data. First, users of the bicycle share program may engage in riding behaviors that are different from other bicycle users. Second, bike-sharing stations only cover a limited urban area. Therefore, it is impractical to estimate bicycle O-D matrices based solely on bike-sharing data. This data will be used in conjunction with bicycle data from other sources mentioned previously to estimate bicycle O-D trip matrices.

Chapter 3: Methodology

In this project, we propose a constraint-based bicycle O-D estimation procedure that consists of two major stages. The first stage is to generate an efficient bicycle route set that contains a set of Pareto optimal (non-dominated) routes. The second stage is to develop a bicycle Path Flow Estimator (PFE) that is based on the path-size logit (PSL) route choice model.

3.1 Cyclists' Route Choice Criteria

Compared to the route choice model for private motorized vehicles, route choice behavior for bicycles is much more complex, because many influential factors affect cyclists' route choice decisions. Empirical studies on bicycle route choice analysis indicate that cyclists choose routes based on a number of criteria that may include distance, number of intersections, road grade, bicycle facility, safety, and space syntax measures – see, e.g., Stinson and Bhat (2003), Hunt and Abraham (2007), Dill and Carr (2003), and Raford et al. (2007). We developed a multi-objective shortest path procedure to generate the bicycle route set. The generated route set can be compared with bicycle route information obtained from field data, e.g., observed routes from crowdsourcing and bike-sharing bicycle data.

In this study, we adopt three key criteria for cyclists' route choice: route-distance-related attributes, route-safety-related attributes, and route-pollution-related attributes. Note that Ryu et al. (2015) also adopted these three criteria for cyclists' route choices in the development of multi-class, multi-criteria bicycle traffic assignment models.

3.1.1 Route Distance

The distance for a bicycle route consists of two components: the sum of link distances along the route, and the sum of turning movement penalties at intersections. It has been shown that intersection delays are a major deterrent in cyclists' route choices. Route distance is computed as follows:

$$d_k^{rs} = \sum_{a \in A} l_a \delta_{ka}^{rs} + \sum_{a \in IN_i} \sum_{b \in OUT_i} cf_i^t d_i^t \, \delta_{ka}^{rs} \delta_{kb}^{rs} \qquad rs \in RS, k \in K^{rs} \qquad (1)$$

where d_k^{rs} is the distance (in meters) on path k connecting O-D pair rs; l_a is the length (in meters) of link a; δ_{ka}^{rs} (δ_{kb}^{rs}) is the path-link indicator, 1 if link a (b) is on path k connecting O-D pair rs and 0 otherwise; cf_i^t is the conversion factor that converts the penalty of turning movement t at intersection i into an equivalent distance unit (in meters/seconds); d_i^t is the penalty (in seconds) of turning movement t at intersection i; A is the set of links in the study network; IN_i is the set of incoming links at intersection i; OUT_i is the set of outgoing links at intersection i; and K^{rs} is the set of paths connecting O-D pair rs.

In equation (1), the first term is the sum of link distances along the route, and the second term represents the total turning movement penalties at intersections that are on the route. Note that the link distance can be generalized by considering other attributes that affect the physical geometry of the link, such as penalty for elevation gain on the link and restriction on gradient. On the other hand, signalized delays at intersections can also be included in the second term.

3.1.2 Route Bicycle Level of Service (BLOS)

There are numerous measures for assessing the safety aspect of bicycle facilities or the suitability for bicycle travel. In this study, we adopt the bicycle level of service (BLOS) measure developed by the HCM (2010) as a surrogate measure to account for different attributes contributing to the safety of bicycle routes. The BLOS measure is a reasonable bicycle safety measure to use because it is considered a state-of-art method and is widely used across the United States as a guide for bicycle facility design. The route BLOS can be calculated as follows:

$$\begin{aligned} BLOS_k^{rs} = {} & 0.200(ABSeg_k^{rs}) + 0.030\big(exp(ABInt_k^{rs})\big) \\ & + 0.050(Cflt_k^{rs}) + 1.40 \end{aligned} \qquad rs \in RS, k \in K^{rs} \qquad (2)$$

where $BLOS_k^{rs}$ is the bicycle level of service on path $k \in K^{rs}$ connecting O-D pair rs; $ABSeg_k^{rs}$ is the length-weighted average segment bicycle score on path $k \in K^{rs}$ connecting O-D pair rs $\left(ABSeg_k^{rs} = (\sum_{a\in A} l_a \cdot BSeg_a \cdot \delta_{ka}^{rs})/(\sum_{a\in A} l_a \cdot \delta_{ka}^{rs})\right)$; l_a is the link length (in meters); $BSeg_a$ is the segment bicycle score on link a; $ABInt_k^{rs}$ is the average intersection bicycle score on path $k \in K^{rs}$ connecting O-D pair rs $\left(ABInt_k^{rs} = \sum_{i\in I}\sum_{a\in IN_i}\sum_{b\in OUT_i} IntBLOS_i\, \delta_{ka}^{rs}\delta_{kb}^{rs} / N_k^{rs}\right)$; $IntBLOS_i$ is the intersection bicycle scores; N_k^{rs} is the total number of intersections on path $k \in K^{rs}$ connecting O-D pair rs; and $Cflt_k^{rs}$ is the number of unsignalized conflicts per km on path $k \in K^{rs}$ connecting O-D pair rs.

The segment and intersection bicycle scores (*Bseg*$_a$, and $IntBLOS_i$) provided below are calibrated based on the volume and speed of motorized vehicles, the width configuration of bicycle facilities, pavement conditions, etc. The details pertaining to the BLOS development can be found in NCHRP Report 616 (Dowling et al. 2008).

$$BSeg_a = 0.507 \ln\left(\frac{V_a}{4 \cdot PHF_a \cdot La_a}\right) + 0.199 Fs_a (1 + 10.38 \cdot HV_a)^2 + 7.066\left(\frac{1}{PC_a}\right)^2 - 0.005(We_a)^2 + 0.76 \tag{3}$$

$$IntBLOS_i = -0.2144 \cdot Wt_i + 0.0153 \cdot CD_i + 0.0066\left(\frac{Vol15_i}{L_i}\right) + 4.1324 \tag{4}$$

where

PHF_a : peak hour factor of link *a*

HV_a : proportion of heavy vehicles on link *a* (in motorized vehicle volume)

We_a : average effective width on the outside through lane of link *a* (ft)

Fs_a : effective speed factor on link *a*

La_a : total number of directional through lanes on link *a*

V_a : directional motorized vehicle volume on link *a* (vph)

PC_a : *FHWA's* five-point pavement surface condition rating on link *a*

Wt_i : width of outside through lane plus paved shoulder (including bike lane where present) of intersection i
CD_i : crossing distance, the width of the side street (including auxiliary lanes and median) of intersection i
$Vol15_i$: volume of directional traffic during a 15-minute period at intersection i
L_i : total number of direction through lanes of intersection i

3.1.3 Route Pollution

For simplicity, we consider carbon monoxide (CO) as an important indicator of the level of atmospheric pollution. Other pollutants can be modeled in a similar manner. Let $PSeg_a$ denote the amount of CO pollution in grams per hour (g/h) on link (segment) a. To estimate the amount of CO pollution, we adopt the nonlinear macroscopic model introduced by Wallace et al. (1998):

$$PSeg_a(v_a) = 0.2038 \cdot t_a(v_a) \cdot \exp\left(\frac{0.7962 \cdot l_a}{t_a(v_a)}\right) \tag{5}$$

where v_a is the motorized vehicle volume on link a; $t_a(v_a)$ is the link travel time measured in minutes; and l_a is the link length measured in meters. Route pollution can then be calculated as follows:

$$CO_k^{rs} = \sum_{a \in A} PSeg_a \cdot \delta_{ka}^{rs} \qquad rs \in RS, k \in K^{rs} \tag{6}$$

3.2 Multi-Criteria Shortest Path

Based on the three key criteria identified in Section 3.1, we can then generate an efficient bicycle route set that contains a set of Pareto optimal (non-dominated) routes. In our problem, route BLOS has a non-additive route cost structure (i.e., the route cost is not a simple additive sum of the link attributes). To address this non-additive route cost structure, we adopted the two-phase procedure developed by Ulungu and Teghem (1995) to solve the multi-objective shortest path problem. In the first phase, possible routes are generated using one of the objectives, while the second phase determines the more efficient routes (or non-dominated routes) relative to the remaining objectives.

In our problem, we first use route distance to generate a set of realistic routes without exceeding the maximum allowable bound. The corresponding route BLOS and route CO pollution are also calculated. If a route has a BLOS or CO pollution value larger than the threshold value, the route is discarded from the route set. These generated routes are then sorted in ascending order, with the first route in the set designated the route with the minimum route distance (i.e., the first efficient route with minimum distance). The next route is then compared to the routes in the efficient route set to determine whether it satisfies the non-dominated route condition. If the route is non-dominated, it will remain in the efficient route set. Otherwise, it is removed from the efficient route set. The process is repeated until all routes in the set are examined. A pseudo code of the two-phase procedure introduced by Ryu et al. (2015) is provided as follows:

do $rs = 1$ to RS

  $K^{rs} = \emptyset$ // *Initialize route set*

  while $\left(Z_{1k}^{rs} \leq \overline{Z}_{1k}^{rs}\right)$

    $K^{rs} = K^{rs} \cup k$ // *Generate all possible routes for the first objective*

  end while

  if $\left(z_{2k}^{rs} > \overline{z}_{2}^{rs}\right)$ or $\left(z_{3k}^{rs} > \overline{z}_{3}^{rs}\right)$ $K^{rs} = K^{rs} - k$

  end if // *Exclude dominated routes by comparing with other objectives*

  do $n = 1$ to Criteria # -1

    Ascending order with Z_{1k}^{rs}

    $\overline{K}^{rs} = \{1\}$ // *Initialize efficient route set with the first route*

    do $k = 2$ to $|K^{rs}|$ // *Update efficient route set with other routes*

      do $l = 1$ to $|\overline{K}^{rs}|$

        if $\left(z_{n+1,k}^{rs} < z_{n+1,l}^{rs}\right)$ $\overline{\mathrm{K}}_{rs}^{m} = \overline{\mathrm{K}}_{rs}^{m} \cup k$

        else $\overline{\mathrm{K}}_{rs}^{m} = \overline{\mathrm{K}}_{rs}^{m} - \{k\}$

        end if

      end do

    end do

  end do

end do

where z_{1k}^{rs}, z_{2k}^{rs}, and z_{3k}^{rs} are the three criteria for route choice; and $\bar{z}_{1k}^{rs}$, $\bar{z}_{2k}^{rs}$, and $\bar{z}_{3k}^{rs}$ are the threshold values for the three criteria.

3.2 Bicycle Path Flow Estimator

Because of the importance of the O-D matrix in transportation planning, the method for estimating the O-D matrix based on traffic counts has been extensively investigated since the 1970s. O-D estimation methods can be grouped into two categories: (1) proportional-assignment methods that assume link travel time is independent of link flow – see, e.g., Willumsen (1981), Van Zuylen and Willumsen (1980), Cascetta (1984), and Lo et al. (1996); and (2) equilibrium-based methods that incorporate traffic assignment into the O-D estimation procedure to avoid the inconsistency issue related to proportional-assignment methods – see, e.g., Nguyen (1977), Fisk and Boyce (1983), and Yang et al. (1992).

Instead of estimating the O-D matrices directly, Sherali et al. (1994) and Bell and Shield (1995) proposed a Path Flow Estimator (PFE), which was further enhanced by Chen et al. (2005, 2009, 2010) and Chootinan et al. (2005). The PFE is a one-stage network observer capable of estimating path flows and path travel times using only traffic counts from a subset of network links. The basic idea of a PFE is to find a set of unique path flows that can reproduce observed link counts as well as other relevant information, such as the target O-D matrix and link capacities. The advantage of the PFE is that the model is a single-level convex program with linear side constraints, which yield unique path flows that can be used to derive other useful information at different spatial levels. The flexibility of aggregating path flows at different spatial levels makes the PFE a suitable tool for bicycle O-D estimation. A PFE model allows us to make use not only of aggregate bicycle count data, but also bicycle data from other sources, such as crowdsourcing and bike-sharing data. The PSL model is an advanced multinomial logit (MNL) model that can accommodate overlapping among routes (Ben-Akiva and Bierlaire, 1999). The PSL-based PFE can be formulated with route utilities as a convex program with various side constraints. The side constraints will include bicycle count data and other available bicycle data, e.g., minimum volumes, intersection turning movement flows, target O-D matrix, etc.

Mathematically, the PSL-based PFE can be formulated as the following convex program with various side constraints (Ryu et al., 2015):

$$\text{minimize } Z(\boldsymbol{f}) = \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}(\ln f_k^{rs} - 1) - \sum_{rs\in RS}\sum_{k\in K_{rs}} U_k^{rs} f_k^{rs} - \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs} \ln PS_k^{rs}$$

subject to

$$(1-\epsilon_a)\, v_a \le x_a \le (1+\epsilon_a)\, v_a, \quad \forall a \in \bar{A} \tag{7}$$

$$(1-\epsilon_r)\, O_r \le P_r \le (1+\epsilon_r)\, O_r, \quad \forall r \in \bar{R} \tag{8}$$

$$(1-\epsilon_s)\, D_s \le A_r \le (1+\epsilon_s)\, D_s, \quad \forall s \in \bar{S} \tag{9}$$

$$(1-\epsilon_{rs})\, z_{rs} \le q_{rs} \le (1+\epsilon_{rs})\, z_{rs}, \quad \forall rs \in \overline{RS} \tag{10}$$

$$f_k^{rs} \ge 0, \quad \forall rs \in RS, k \in K_{rs} \tag{11}$$

$$x_a = \textstyle\sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}, \quad \forall a \in A \tag{12}$$

$$P_r = \textstyle\sum_{s\in S}\sum_{k\in K_{rs}} f_k^{rs}, \quad \forall r \in R \tag{13}$$

$$A_s = \textstyle\sum_{r\in}\ \sum_{k\in K_{rs}} f_k^{rs}, \quad \forall s \in S \tag{14}$$

$$q_{rs} = \textstyle\sum_{k\in K_{rs}} f_k^{rs}, \quad \forall rs \in RS \tag{15}$$

where f_k^{rs} is the estimated traffic flow on path $k \in K_{rs}$ connecting O-D pair $rs \in RS$; U_k^{rs} is the utility of route $k \in K_{rs}$ connecting O-D pair $rs \in RS$; PS_k^{rs} is the path size factor of path $k \in K_{rs}$ connecting O-D pair $rs \in RS$; v_a is the observed count on link a; x_a is the estimated flow on link a; ϵ_a is the percentage measurement error allowed for the traffic count on link a; $\bar{A}$ is the set of links with measurements; O_r and D_s are the generated or observed trip production of origin r and the generated or observed trip attraction of destination s; P_r and A_s are the estimated trip production of origin r and the estimated trip attraction of destination s; ϵ_r and ϵ_s are the percentage measurement errors allowed for the trip production of origin r and the trip attraction of destination s; $\bar{R}$ and $\bar{S}$ are the sets of zones with planning data or observations; $\overline{RS}$ is the set of target (or prior) O-D pairs; z_{rs} is prior O-D flows of O-D pair $rs \in \overline{RS}$; ϵ_{rs} is the percentage measurement error allowed for the O-D pair $rs \in \overline{RS}$; and δ_{ka}^{rs} is the path-link indicator: 1 if link

$a \in A$ is on path $k \in K_{rs}$ connecting O-D pair $rs \in RS$, and 0 otherwise. Detailed discussions of the above program can be found in Ryu et al. (2015).

Ryu et al. (2015) proposed to define the utility U_k^{rs} as follows:

$$U_k^{rs} = -\left((d_k^{rs})^{\alpha} \cdot (BLOS_k^{rs})^{\beta} \cdot (CO_k^{rs})^{\gamma}\right) \qquad \forall rs \in RS, k \in K^{rs} \tag{16}$$

where α, β, and γ are parameters of the utility function.

The path size factor PS_k^{rs} is given as follows:

$$PS_k^{rs} = \sum_{a \in k} \left(\frac{l_a}{L_k^{rs}}\right) \cdot \left(\frac{1}{\sum_{\hat{k} \in K^{rs}} \delta_{\hat{k}a}^{rs}}\right) \qquad \forall rs \in RS, k \in K^{rs} \tag{17}$$

where l_a is the length of link a; and L_k^{rs} is the length on path k connecting O-D pair $rs \in RS$. The Lagrangian function of the PSL-based PFE formulation is as follows:

$$L(\boldsymbol{f}, \boldsymbol{\rho}^{+}, \boldsymbol{\rho}^{-}, \boldsymbol{\mu}^{+}, \boldsymbol{\mu}^{-}, \boldsymbol{\varphi}^{+}, \boldsymbol{\varphi}^{-}, \boldsymbol{\sigma}^{+}, \boldsymbol{\sigma}^{-})$$

$$= Z(\boldsymbol{f}) + \sum_{a \in \bar{A}} \rho_a^{-} \left((1 - \epsilon_a) v_a - \sum_{rs \in RS} \sum_{k \in K_{rs}} f_k^{rs} \delta_{ka}^{rs} \right)$$

$$+ \sum_{a \in \bar{A}} \rho_a^{+} \left((1 + \epsilon_a) v_a - \sum_{rs \in RS} \sum_{k \in K_{rs}} f_k^{rs} \delta_{ka}^{rs} \right) + \sum_{r \in \bar{}} \mu_r^{-} \left((1 - \epsilon_r)\, O_r - \sum_{s \in S} \sum_{k \in K_{rs}} f_k^{rs} \right)$$

$$+ \sum_{r \in \bar{}} \mu_r^{+} \left((1 + \epsilon_r)\, O_r - \sum_{s \in S} \sum_{k \in K_{rs}} f_k^{rs} \right) + \sum_{s \in \bar{S}} \varphi_s^{-} \left((1 - \epsilon_s)\, D_s - \sum_{r \in} \sum_{k \in K_{rs}} f_k^{rs} \right)$$

$$+ \sum_{s \in \bar{S}} \varphi_s^{+} \left((1 + \epsilon_s)\, D_s - \sum_{r \in} \sum_{k \in K_{rs}} f_k^{rs} \right) + \sum_{rs \in \bar{S}} \sigma_{rs}^{-} \left((1 - \epsilon_{rs})\, z_{rs} - \sum_{k \in K_{rs}} f_k^{rs} \right)$$

$$+\sum_{rs\in\bar{S}}\sigma_{rs}^{+}\left((1+\epsilon_{rs})\,z_{rs}-\sum_{k\in K_{rs}}f_k^{rs}\right)$$

where ρ_a^-, ρ_a^+, μ_r^-, μ_r^+, φ_s^-, φ_s^+, σ_{rs}^-, and σ_{rs}^+ are the dual variables of constraints (7)-(10), respectively.

The first partial derivatives with respect to the path-flow variables can be given as follows:

$$\frac{\partial L}{\partial f_k^{rs}}=0\Longrightarrow \ln f_k^{rs}-U_k^{rs}-\ln PS_k^{rs}-\sum_{a\in\bar{A}}\rho_a^-\delta_{ka}^{rs}-\sum_{a\in\bar{A}}\rho_a^+\delta_{ka}^{rs}-\mu_r^--\mu_r^+-\varphi_s^--\varphi_s^+-\sigma_{rs}^--\sigma_{rs}^+=0,\quad\forall rs\in RS,k\in K^{rs}\qquad(17)$$

From equation (17), the analytical path flow solution can be expressed as follows:

$$f_k^{rs}=\exp(U_k^{rs}+\ln PS_k^{rs}+\sum_{a\in\bar{A}}\rho_a^-\delta_{ka}^{rs}+\sum_{a\in\bar{A}}\rho_a^+\delta_{ka}^{rs}+\mu_r^-+\mu_r^++\varphi_s^-+\varphi_s^++\sigma_{rs}^-+\sigma_{rs}^+),\quad\forall rs\in RS,k\in K^{rs}\qquad(18)$$

An iterative balancing scheme presented in Ryu et al. (2014) is adopted to solve the PFE model. The iterative procedure is outlined as follows:

Step 1: *Initialization*

1.1 Set inner iteration (n)=0
1.2 Set primal variables: $x_a^n, P_r^n, A_s^n, and\ q_{rs}^n\ =0$
1.3 Set dual variables: $(\rho_a^-)^n, (\rho_a^+)^n, (\mu_r^-)^n, (\mu_r^+)^n, (\varphi_s^-)^n, (\varphi_s^+)^n, (\sigma_{rs}^-)^n$, and $(\sigma_{rs}^+)^n=0$.

Step 2: *Compute Dual and Primal Variables*

The values of ρ_a^+, μ_r^+, φ_s^+, and σ_{rs}^+ are restricted to be non-positive, while the values of ρ_a^-, μ_r^-, φ_s^-, and σ_{rs}^- must be non-negative.

2.1 Update dual variables

a. For each measured link $a\in\bar{A}$, update the dual variables.

$(\rho_a^+)^n=Min\left\{0,(\rho_a^+)^{n-1}+\ \ln\left(\frac{(1+\epsilon_a)v_a}{x_a^n}\right)\right\}$, and

$(\rho_a^-)^n = Max\left\{0, (\rho_a^-)^{n-1} + \ln\left(\frac{(1-\epsilon_a)v_a}{x_a^n}\right)\right\}.$

b. For each zonal production flow $r \in \bar{R}$, update the dual variables.

$(\mu_r^+)^n = Min\left\{0, (\mu_r^+)^{n-1} + \ln\left(\frac{(1+\epsilon_r)\,O_r}{P_r^n}\right)\right\}$, and

$(\mu_r^-)^n = Max\left\{0, (\mu_r^-)^{n-1} + \ln\left(\frac{(1-\epsilon_r)\,O_r}{P_r^n}\right)\right\}.$

c. For each zonal attraction flow $s \in \bar{S}$, update the dual variables.

$(\varphi_s^+)^n = \mathrm{Min}\left\{0, (\varphi_s^+)^{n-1} + \ln\left(\frac{(1+\epsilon_s)\,D_s}{A_s^n}\right)\right\}$, and

$(\varphi_s^-)^n = \mathrm{Max}\left\{0, (\varphi_s^-)^{n-1} + \ln\left(\frac{(1-\epsilon_s)\,D_s}{A_s^n}\right)\right\}.$

d. For each target O-D flow $rs \in \overline{RS}$, update the dual variables.

$(\sigma_{rs}^+)^n = \mathrm{Min}\left\{0, (\sigma_{rs}^+)^{n-1} + \ln\left(\frac{(1+\epsilon_{rs})\,z_{rs}}{q_{rs}^n}\right)\right\}$, and

$(\sigma_{rs}^-)^n = \mathrm{Max}\left\{0, (\sigma_{rs}^-)^{n-1} + \ln\left(\frac{(1-\epsilon_{rs})\,z_{rs}}{q_{rs}^n}\right)\right\}.$

2.2 Compute primal variables

a. Compute path flows

$$(f_k^{rs})^n = \exp\Bigg(U_k^{rs} + \ln PS_k^{rs} + \sum_{a\in\bar{A}}(\rho_a^-)^n\delta_{ka}^{rs} + \sum_{a\in\bar{A}}(\rho_a^+)^n\delta_{ka}^{rs} + (\mu_r^-)^n + (\mu_r^+)^n + (\varphi_s^-)^n + (\varphi_s^+)^n + (\sigma_{rs}^-)^n + (\sigma_{rs}^+)^n\Bigg), \quad \forall rs \in RS, k \in K^{rs}$$

b. Compute link flows

$x_a^n = \sum_{rs\in RS}\sum_{k\in K_{rs}}(f_k^{rs})^n\delta_{ka}^{rs}, \quad \forall a \in A$

c. Compute zonal production flows

$P_r^n = \sum_{s\in S}\sum_{k\in K_{rs}}(f_k^{rs})^n, \quad \forall r \in R$

d. Compute zonal attraction flows

$A_s^n = \sum_{r\in}\ \sum_{k\in K_{rs}}(f_k^{rs})^n, \quad \forall s \in S$

e. Compute O-D flows

$q_{rs}^n = \sum_{k\in K_{rs}}(f_k^{rs})^n, \quad \forall rs \in RS$

Step 3: *Convergence and Divergence Test*

$$\xi = Max\{|(\rho_a^+)^n - (\rho_a^+)^{n-1}|, |(\rho_a^-)^n - (\rho_a^-)^{n-1}|, |(\mu_r^+)^n - (\mu_r^+)^{n-1}|, |(\mu_r^-)^n - (\mu_r^-)^{n-1}|, |(\varphi_s^+)^n - (\varphi_s^+)^{n-1}|, |(\varphi_s^-)^n - (\varphi_s^-)^{n-1}|, |(\sigma_{rs}^+)^n - (\sigma_{rs}^+)^{n-1}|, |(\sigma_{rs}^-)^n - (\sigma_{rs}^-)^{n-1}|\}$$

If $\xi \le \eta_0$, where η_0 is a convergence tolerance (e.g., 10^{-6}), then terminate and output the results. If $\xi > \eta_0$, then set all parameters in the next iteration as equal to those of the current iteration, set $n = n + 1$, and go to step 2.

Chapter 4: Case Study

In this section, a real-world case study in Salt Lake City, Utah is provided to demonstrate the proposed methodology.

4.1 Bicycle Counts

The bicycle count data used in this project were obtained from Salt Lake City's Transportation Division. In 2010, Salt Lake City joined the National Bicycle/Pedestrian Documentation Project. Since then, it has recruited volunteers every year to record the number of bicyclists at key intersections located throughout the city. In this project, we used the latest bicycle count data, which was collected in September 2015. The data collection process was conducted on Tuesday, September 15; Wednesday, September 16; Thursday, September 17; Saturday, September 19; and Sunday, September 20. In total, 19 intersections were involved. Counting duration at each intersection took place for two hours each day. The time of day for the counts was 5–7 p.m. on weekdays, and 12–2 p.m. on weekends. Table 4.1 shows the statistical summary for the bicycle count data. Because we could not identify the specific location of one intersection on the University of Utah campus, we deleted it from our analysis. Figure 4.1 presents a map of the bicycle count locations.

TABLE 4.1 Summary statistics for two-hour bicycle counts

Statistic	All Counts	Weekday	Weekend
Number of counts	95	57	38
Minimum	2	7	2
Maximum	161	129	161
Median	47	47	42
Mean	54.8	54.1	55.9
Standard deviation	35.7	31.3	41.8

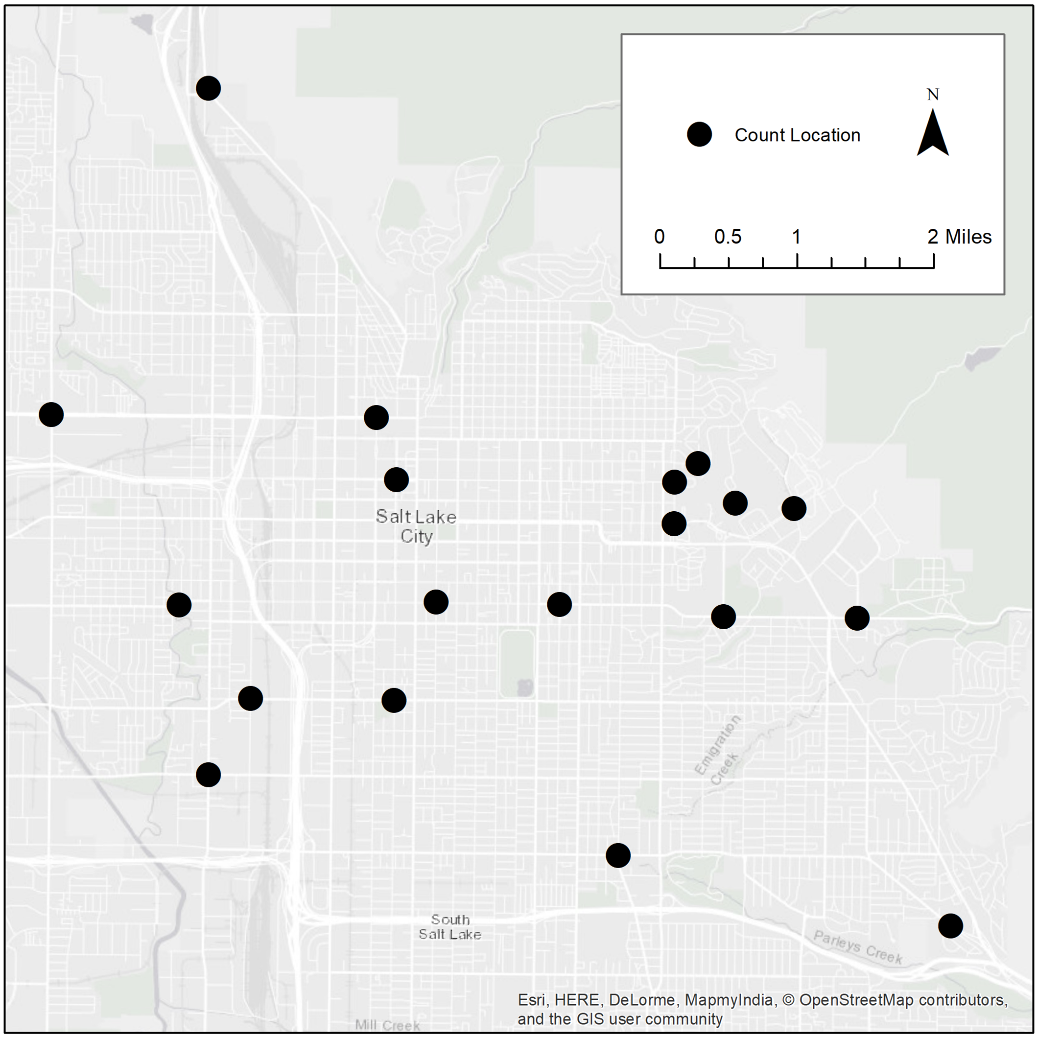

FIGURE 4.1 Map of locations of bicycle counts

4.2 Strava Data

Strava Metro is a suite of data services that provide a window into cycling and pedestrian activity at a high level of granularity (Strava, 2015). Millions of people upload their rides and runs to Strava every week via their smartphone or GPS device (Strava, 2019). Strava claims that in denser metropolitan areas, commutes comprise more than one-half of the data; thus, Strava Metro data provides valuable insight into the needs of those riding exclusively for transportation purposes

(Strava, 2015). This data is aggregated and de-identified by Strava Metro and is then made available for purchase. Departments of transportation and city planning groups can utilize Strava Metro data to analyze and understand real-world cycling and pedestrian route preference and to improve infrastructures for bicyclists and pedestrians.

In Utah, nearly 58,000 cyclists and 43,000 runners, walkers, hikers and joggers use the Strava app to track and share their activities. In April 2017, the Utah Department of Transportation (UDOT) signed a two-year contract with Strava to obtain all of that data (Smith, 2017). Figure 4.2 shows the Strava heatmap for cycling activities in Salt Lake City, Utah.

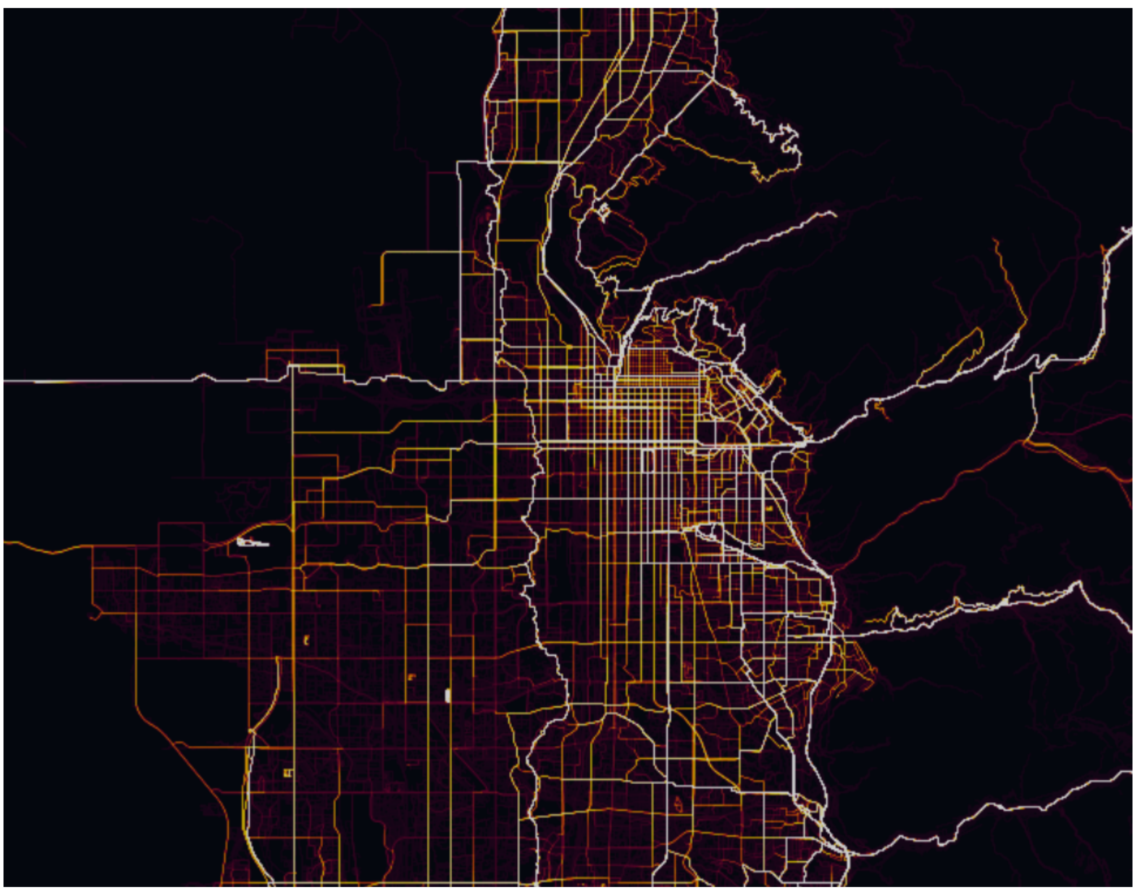

FIGURE 4.2 Strava heatmap for cycling activities in Salt Lake City

The Strava data we obtained from the UDOT is the link-level ride counts. Figure 4.3 shows the annual rides on each road segment in our study area, in which wider lines represent more rides.

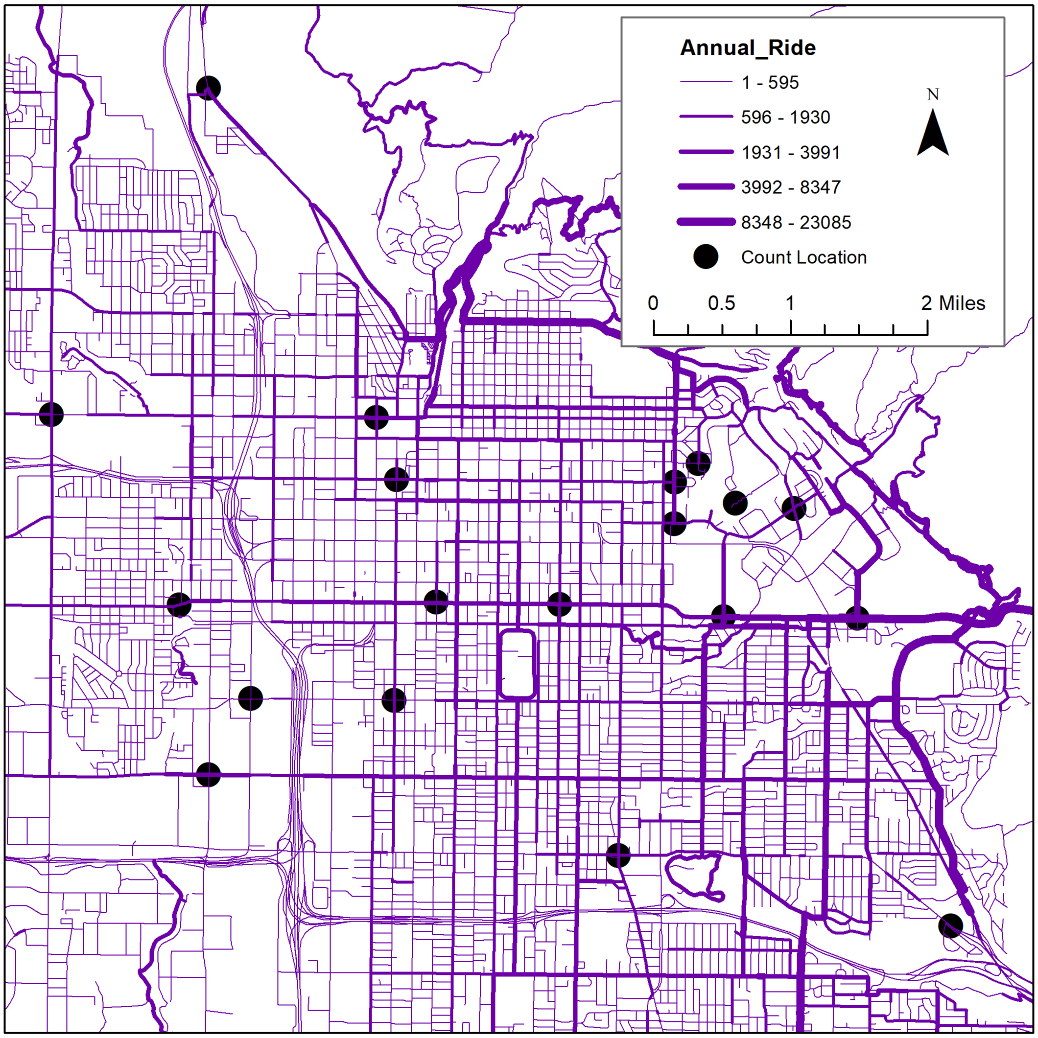

FIGURE 4.3 Annual rides on each road segment in the study area

4.3 Correlation Analysis between Bicycle Counts and Strava Data

The bike count data we obtained is comprised of bicycle volume counts at intersections. The Strava data, however, is link-based ride data. To analyze the correlation between bicycle counts and Strava data, we considered the sum of rides for all entering legs at an intersection. Figure 4.4 shows the comparison between bicycle counts and Strava data for the 19 intersections. One can see that, with the exception of two obvious outliers, the count data and Strava data are linearly correlated. If the two outliers are removed, the R-square value is increased to 0.622, as shown in Figure 4.5. The relationship between the count value and the Strava value is given by $(v_a)_{Strava} = 14.212 \cdot$

$(v_a)_{Count}$. With this equation, we can infer the bicycle volume using the Strava data. Note that the bicycle counts are in two-hour intervals while the Strava data are annual total counts. If we assume bicycle travel is uniformly distributed within 12 hours in a day and 365 days in a year, the penetration ratio of the Strava app can be approximately estimated as 14.212/365/6=0.649%.

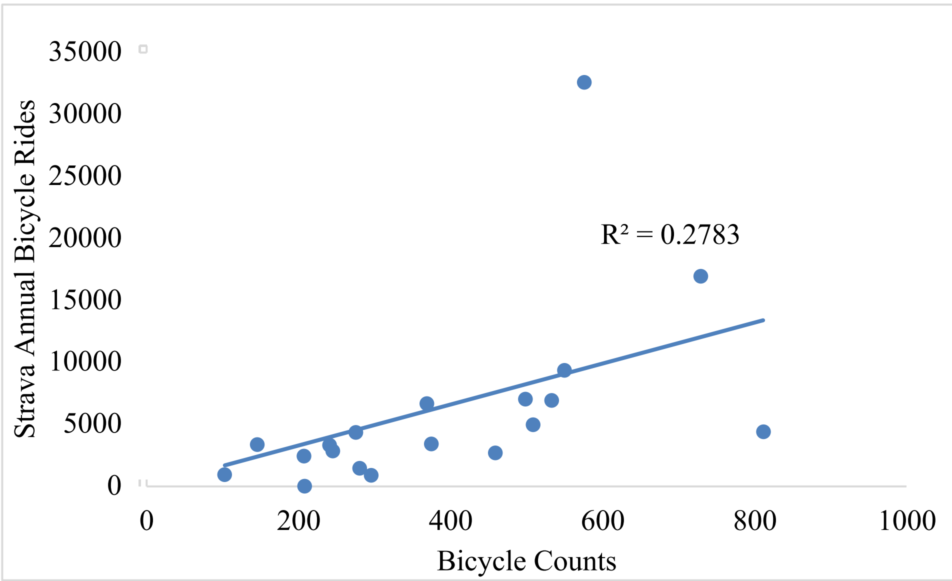

FIGURE 4.4 Comparison between bicycle counts and Strava data for all 19 intersections

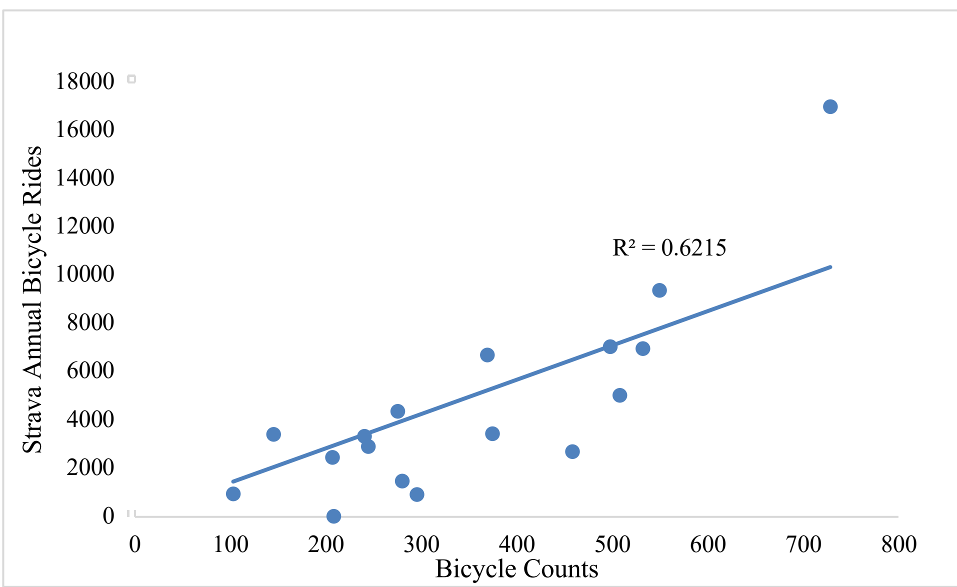

FIGURE 4.5 Comparison between bicycle counts and Strava data for 17 intersections

4.4 Bicycle O-D Estimation

A sub-network in Salt Lake City, as shown in Figure 4.6, is chosen to demonstrate the proposed method. In total, the sub-network consists of 42 traffic assignment zones, 119 nodes, and 380 links.

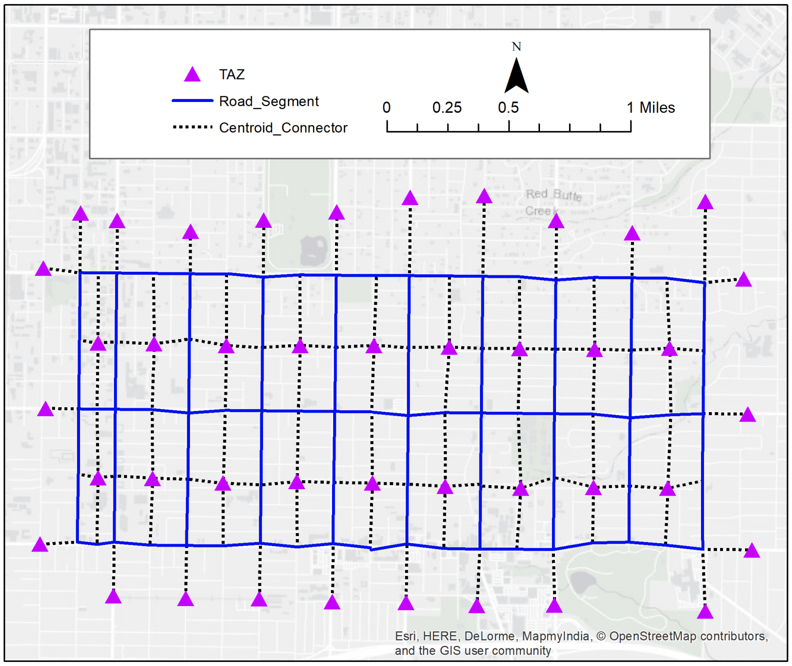

FIGURE 4.6 A sub-network in Salt Lake City

The motor vehicle volume on a link was estimated based on annual average daily traffic (AADT) data, which was obtained from the UDOT. The speed limit information on a link was also provided by a UDOT dataset. The data regarding the number of lanes on a link was manually collected in Google Maps (https://www.google.com/maps/). The bicycle lane information was provided by the Salt Lake Transportation Division (http://bikeslc.com/WheretoRide/SLCBikeMap.html). For unavailable data, including the peak hour factor of a link, the proportion of heavy vehicles, the average effective width on the outside through lane of a link, and FHWA's five-point pavement surface condition rating, we used the default values recommended in the HCM (2010).

In total, 67 Strava link flow data points, whose link flows are greater than or equal to 1,000 (per year), were adopted for the O-D estimation. Considering the inconsistency among the Strava data, the error bounds for link counts were specified as ±15% to ensure solution existence. Note that there could be many possible O-D demand patterns that match the observed link counts.

Figure 4.7 shows the scatter plots of observed and estimated link flows. The R^2 value is provided to demonstrate the accuracy of the estimated flows. One can see from this data that the estimated link flow may closely match the observed values. Figure 4.8 shows the estimated production flows and attraction flows in the study area (in unit of bicycle trips per two hours). Note that Figure 4.8 only shows O-D flows that are greater than or equal to 10.

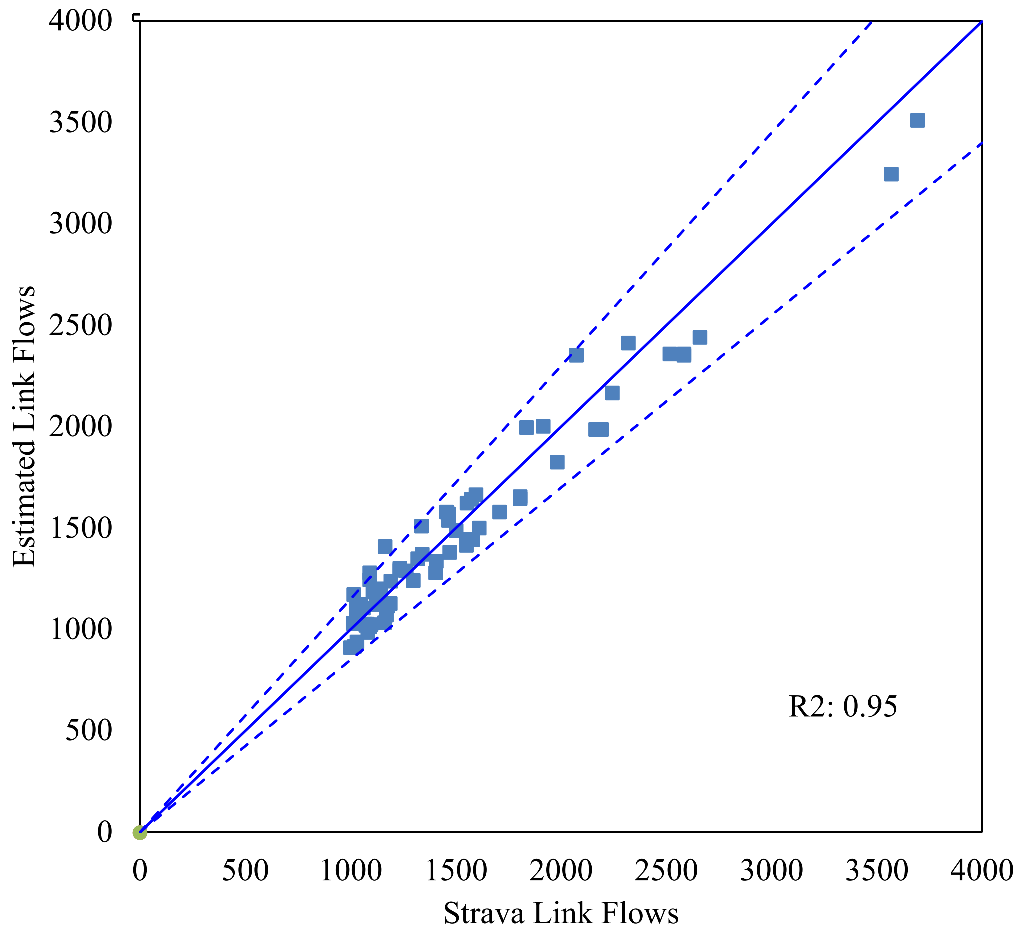

FIGURE 4.7 Comparison between estimated flows and Strava flows

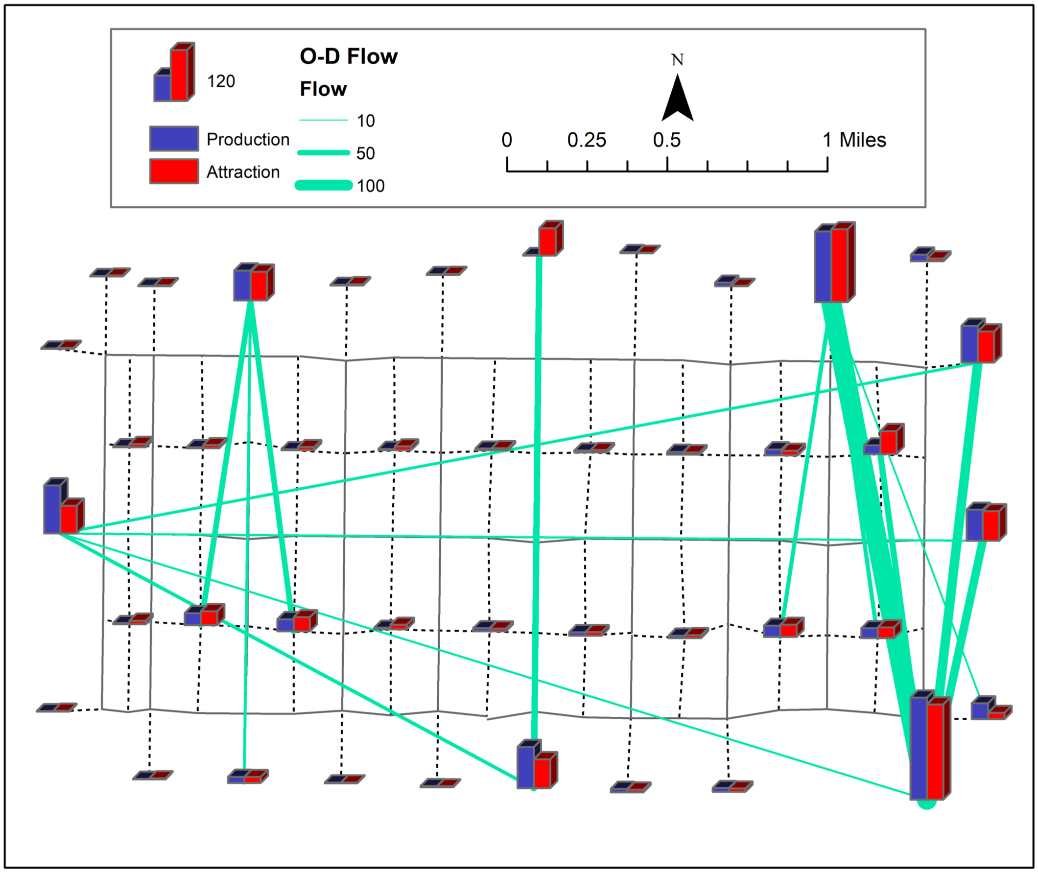

FIGURE 4.8 Estimated zonal bicycle production and attraction flows

4.5 Discussion

In this project, the available bicycle count data only had 19 data points. Although the count data and Strava data are linearly correlated with a high R-square value, other large-scale datasets are clearly needed to further validate the correlation between the count data and the Strava data. The Strava data has the potential issue of biases. It is difficult to quantify these biases from user-submitted content without comparing against reference data sources.

Although crowdsourced data such as Strava data have some potential issues, they offer a new source of data for transportation planning. Crowdsourced data may serve as an economical complement to official count programs, which are typically highly demanding in terms of resources. Crowdsourced data can effectively increase the spatial and temporal resolution of official count programs.

The PFE model can effectively employ the bicycle count data and the Strava data to estimate path flow and O-D flow. However, it is difficult to accurately match the data due to inconsistencies and inaccuracies in the data. In future studies, a more advance PFE model, such as the L_∞-norm PFE (Chen et al., 2009), should be adopted to address inconsistencies in the data.

Chapter 5: Concluding Remarks

This report proposes a methodology to utilize multiple data sources including crowdsourcing data and traditional bicycle count data in bicycle traffic origin-destination (O-D) matrix estimation. Bicycle data that are useful for bicycle O-D estimation are first explored. There are multiple sources for bicycle data, including traditional bicycle count data, emerging crowdsourcing bicycle data, and data from bike-sharing programs. The characteristics of each data sources are discussed. A two-stage constraint-based bicycle O-D estimation procedure is then proposed. The first stage generates an efficient bicycle route set that contains a set of Pareto optimal (non-dominated) routes. The second stage is a bicycle PFE model that is based on the path-size logit (PSL) route choice model. The procedure is flexible and can be adjusted to different levels of data availability and quality. A real-world case study was conducted in Salt Lake City, Utah to demonstrate the proposed methodology. The results show that the proposed O-D estimation procedure can effectively utilize the bicycle count data and crowdsourcing data, i.e., the Strava data, to estimate bicycle O-D matrix. The Strava data have the potential to serve as an economical complement for official count programs.

The findings in this project, if could be further verified with more extensive data sets, have importation implications in bicycle facility assessment. Crowdsourcing data is demonstrated to be a useful data source in bicycle modeling and planning. The combination of traditional bicycle count data and emerging crowdsourcing data provides many opportunities and challenges for future studies.

References

1. Ainsley, M., 2018. PLANNER RESOURCES: LET'S TALK ABOUT VALIDATION COUNTS. Eco-Counter Blog. https://www.eco-compteur.com/blog/2018/10/17/planner-resources-lets-talk-about-validation-counts/ (Accessed April 15, 2019)
2. Andersen, M., 2015. This $50 device could change bike planning forever. Bike Portland. https://bikeportland.org/2015/01/13/50-device-change-bike-planning-forever-130891 (Accessed April 15, 2019)
3. Ben-Akiva, M., Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. Handbook of Transportation Science. R.W. Hall ed., Kluwer Publishers.
4. Bell, M.G.H., Shield, C.M. (1995). A log-linear model for path flow estimation. In: Proceedings of the 4th International Conference on the Applications of Advanced Technologies in Transportation Engineering, Carpi, 695–699.
5. Brabham, D. C. (2012). The myth of amateur crowds: a critical discourse analysis of crowdsourcing coverage. *Information, Communication & Society*, 15(3), 394–410.
6. Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B*, 18, 289–299.
7. Chen, A., Chootinan, P., Recker, W. (2005). Examining the quality of synthetic origin-destination trip table estimated by path flow estimator. *Journal of Transportation Engineering*, 131 (7), 506–513.
8. Chen, A., Chootinan, P., Recker, W. (2009). Norm approximation method for handling traffic count inconsistencies in path flow estimator. *Transportation Research Part B*, 43 (8), 852–872.
9. Chen, A., Pravinvongvuth, S., Chootinan, P. (2010). Scenario-based multiobjective AVI reader location models under different travel demand patterns. *Transportmetrica,* 6 (1), 53–78.
10. Chootinan, P., Chen, A., Recker, W., (2005). Improved path flow estimator for estimating origin–destination trip tables. *Transportation Research Record*, 1923, 9–17.
11. Dill, J., Carr, T. (2003). Bicycle commuting and facilities in major US cities: If you build them, commuters will use them. *Transportation Research Record,* 1828, 116–123.
12. Dill, J. (2009). Bicycling for transportation and health: the role of infrastructure. *Journal of Public Health Policy*, 30 (1), 95–110.
13. FHWA. (2013). Traffic Monitoring Guide. U.S. Department of Transportation, Publication FHWA PL-13-015.
14. FHWA. (2014). Guidebook on Pedestrian and Bicycle Volume Data Collection. U.S. Department of Transportation, NCHRP Report 797.

15. Fisk, C.S., Boyce, D.E. (1983). A note on trip matrix estimation from link traffic count data. *Transportation Research Part B*, 17, 245–250.
16. Gordon-Larsen, P., Boone-Heinonen, J., Sidney, S., Sternfeld, B., Jacobs, D., Lewis, C. (2009). Active commuting and cardiovascular disease risk: the CARDIA study. *Archive of Internal Medecine*, 169 (13), 1216–1223.
17. Hood, J., Sall, E., Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California." Transportation Letters: *The International Journal of Transportation Research*, 3(1), 63–75.
18. Hunt, J.D., Abraham, J.E. (2007). Influences on bicycle use. *Transportation*, 34(4), 453–470.
19. Kuzmyak, J., Walters, J., Bradley, M., Kockelman, K. (2014). Estimating Bicycling and Walking for Planning and Project Development: A Guidebook. National Cooperative Highway Research Program, Report 770.
20. Lo, H.P., Zhang, N., Lam, W.H.K. (1996). Estimation of an origin–destination matrix with random link choice proportions: a statistical approach. *Transportation Research Part B*, 30, 309–324.
21. McKenzie, B. (2014). Modes Less Traveled—Bicycling and Walking to Work in the United States: 2008–2012. American Community Survey Reports, ACS-25.
22. Nie, Y., Lee, D.-H. (2002). An uncoupled method for the equilibrium-based linear path flow estimator for origin–destination trip matrices. *Transportation Research Record*, 1783, 72–79.
23. Nguyen, S. (1977). Estimating an OD matrix from network data: a network equilibrium approach. University of Montreal Publication, No. 60.
24. Pucher, J., Dill, J., Handy, S. (2010). Infrastructure, programs and policies to increase bicycling: An international review. *Preventative Medicine*, 50(1), 106–125.
25. Ruy, S., Chen, A., Su, J., Christensen, K., & Choi, K. (2015). Development of multi-class, multi-criteria bicycle traffic assignment models and solution algorithms (No. TRCLC 15-10). Western Michigan University. Transportation Research Center for Livable Communities.
26. Ryu, S., Chen, A., Zhang, H. M., & Recker, W. (2014). Path flow estimator for planning applications in small communities. Transportation Research Part A: Policy and Practice, 69, 212-242.
27. Raford, N., Chiaradia, A., Gil, J. (2007). Space Syntax: the role of urban form in cyclist route choice in central London. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C.
28. Strava, 2015. Strava Metro Comprehensive User Guide. https://www.ubdc.ac.uk/media/1323/stravametro_200_user_guide_withoutpics.pdf (Accessed April 20, 2019)

29. Strava, 2019. Better Data for Better Cities. https://metro.strava.com/ (Accessed April 20, 2019)
30. Smith B., 2017. Utah is using your social-media data to make roads safer for bikers and pedestrians. https://www.sltrib.com/news/2017/11/21/utah-is-using-your-social-media-data-to-make-roads-safer-for-bikers-and-pedestrians/ (Accessed April 20, 2019)
31. Stinson, M.A., Bhat, C.R. (2003). An analysis of commuter bicyclist route choice using a stated preference survey. *Transportation Research Record,* 1828, 107–115.
32. Van Zuylen, J.H., Willumsen, L.G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B*, 14, 281–293.
33. Willumsen, L.G. (1981). Simplified transport models based traffic counts. *Transportation*, 10, 257–278.
34. Yang, H., Sasaki, T., Iida, Y., Asakura, Y. (1992). Estimation of origin–destination matrices from link traffic counts on congested networks. *Transportation Research Part B*, 26, 417–434.