DERIVING PUBLIC TRANSIT (PT) RELIABILITY MEASURES USING
SMART CARD DATA: 2 CASE STUDY PT LINES FROM KONYA CITY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

FATİH SARIYÜZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCEMASTER OF SCIENCE
IN
GEODETIC AND GEOGRAPHIC INFORMATION
TECHNOLOGIESGEODETIC AND GEOGRAPHIC INFORMATION
TECHNOLOGIES

SEPTEMBER 2020

Approval of the thesis:

DERIVING PUBLIC TRANSIT (PT) RELIABILITY MEASURES USING SMART CARD DATA: 2 CASE STUDY PT LINES FROM KONYA CITY
submitted by **FATİH SARIYÜZ** in partial fulfillment of the requirements for the degree of **Master of Science** in **Geodetic and Geographic Information Technologies, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar

Dean, Graduate School of **Natural and Applied Sciences**

                        _____

Prof. Dr. Zuhal Akyürek

Head of the Department, **Civil Engineering**

                        _____

Assoc. Prof. Dr. Hediye Tüydeş Yaman

Supervisor, **Civil Engineering, METU**

                        _____

Examining Committee Members:

Prof. Dr. Ela Babalık

City and Regional Planning, METU

                        _____

Assoc. Prof. Dr. Hediye Tüydeş Yaman

Civil Engineering, METU

                        _____

Prof. Dr. Ebru Vesile Öcalır Akünal

City and Regional Planning, Gazi University

                        _____

Assist. Prof. Dr. Hande Işık Öztürk

Civil Engineering, METU

                        _____

Assist. Prof. Dr. Meriç H. Gökdalay

Aviation Management, University of TAA

                        _____

Date: 18.09.2020

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Fatih, Sarıyüz:

Signature :

ABSTRACT

DERIVING PUBLIC TRANSIT (PT) RELIABILITY MEASURES USING SMART CARD DATA: 2 CASE STUDY PT LINES FROM KONYA CITY

Sarıyüz, Fatih
Master of Science, Geodetic and Geographic Information Technologies
Supervisor: Assoc. Prof. Dr. Hediye Tüydeş Yaman

September 2020, 153 pages

In Intelligent Transportation Systems (ITS) applications for Public Transit (PT), it is almost customary to monitor and collect bus trajectory data in addition to smart card systems, which includes geocoded information on ticketing. These datasets enable system managers to derive PT reliability and Level-of-Service (LOS) measures, such as average travel time, dwell time, etc, at a PT line level. However, it is also necessary to derive such measures spatio-temporally to get PT characteristics in metropolitan regions with heterogeneous land use and congestion patterns. This study focuses on processing and evaluation of smart card data (SCD) to determine PT reliability, using the case study results from Konya PT smart card system. A major goal of the study was to convert geocoded SCD to create time-space diagrams using linear referencing to a bus route. A set of data preprocessing studies have been conducted by using a database management software and a GIS software (PostgreSQL and ArcGIS) for converting data. Afterward, GIS-based technics known as Linear Referencing (LR) and Dynamic Segmentation (DynSeg) have been implemented to BTD coordinates gathered from public BTD in ArcGIS software to obtain the kilometers (aka Station Km) of the smart card records on the relevant route segments. Arrival times to the bus stops have been estimated, which also allowed

estimation of inter-stop travel times (ISTT) along a route. Arrival time estimations (ATEs) have been used to visualize bus trajectory with time-space diagrams and to calculate total travel times and Travel Time (TT) reliability indicators. As a result, a route based evaluation has been made over two study PT lines in the Konya PT network by using these calculated TT reliability indicators and time-space diagrams.

Keywords: Public Transit, Smart Card Data, Linear Referencing, Dynamic Segmentation, Travel Time Reliability, Arrival Time Estimation

ÖZ

AKILLI KART VERİLERİNİN KULLANILARAK TOPLU ULAŞIM GÜVENİLİRLİK ÖLÇÜTLERİNİN ELDE EDİLMESİ: KONYA ŞEHRİNDEN 2 ÖRNEK HAT ÇALIŞMASI

Sarıyüz, Fatih
Yüksek Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri
Tez Yöneticisi: Doç. Dr. Hediye Tüydeş Yaman

Eylül 2020, 153 sayfa

Toplu Ulaşımdaki Akıllı Ulaşım Sistemleri uygulamalarında, biletleme sırasında coğrafi-kodlanmış akıllı kart sistemlerine ek olarak otobüs yörünge verilerinin izlenmesi ve toplanması alışılagelmiştir. Bu veri setleri, sistem yöneticilerinin, bir toplu ulaşım hattında, ortalama seyahat süresi ve bekleme süresi gibi toplu ulaşım güvenilirlik ve hizmet seviyesi ölçütlerini elde etmesini mümkün kılmaktadır. Bununla birlikte, heterojen arazi kullanımı ve sıkışıklığa sahip metropol bölgelerdeki toplu ulaşım karakteristiğini elde edebilmek için bu ölçütler konum-zamansal olarak da toplanmalıdır. Bu çalışma, Konya toplu ulaşım akıllı kart sisteminden örnek çalışma sonuçları kullanılarak, toplu ulaşım güvenilirliğinin tespit edilmesi için akıllı kart verilerinin işlenmesi ve değerlendirilmesi üzerine odaklanmıştır. Çalışmanın ana hedefi, coğrafi-kodlanmış akıllı kart verilerini bir otobüs rotasına doğrusal referanslama yaparak zaman-konum diyagramları oluşturmaktır. Bir dizi veri ön işleme çalışması, bir veritabanı yönetim yazılımı ve Coğrafi Bilgi Sistemi (CBS) yazılımı (PostgreSQL and ArcGIS) kullanılarak gerçekleştirilmiştir. Daha sonra, ait oldukları rota bölümlerindeki akıllı kart kayıtlarının kilometrelerini elde edebilmek için ArcGIS yazılımında toplu ulaşım otobüs yörünge verilerinden toplanan koordinatlara Doğrusal Referanslama ve Dinamik Segmentasyon olarak bilinen CBS

tabanlı teknikler uygulanmıştır. Otobüs duraklarına varış zamanları tahmin edilmiştir ki bu aynı zamanda bir rota boyunca duraklar arası seyahat sürelerinin tahmin edilmesine de imkan sağlamıştır. Tahmini varış zamanları, zaman-konum diyagramları ile otobüs yörüngelerini görselleştirmek, toplam seyahat sürelerini ve seyahat süresi güvenilirlik ölçütlerini hesaplamak için kullanılmıştır. Sonuç olarak, hesaplanan bu seyahat süresi güvenilirlik ölçütleri ve zaman-konum diyagramları kullanılarak Konya toplu ulaşım ağındaki iki örnek çalışma hattı üzerinde rota bazlı değerlendirme yapılmıştır.

Anahtar Kelimeler: Toplu Ulaşım, Otobüs Yörüngesi, Doğrusal Referanslama, Dinamik Segmentasyon, Seyahat Süresi Güvenilirliği

To my dear father i have recently lost, Kazım Sarıyüz...

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

xx

LIST OF ABBREVIATIONS

AFC   Automated Fare Collection Systems

APTS   Advanced Public Transportation System

ATE   Arrival Time Estimation

AVL   Automated Vehicle Location

BTD   Bus Trajectory Data

DBMS   Database Management System

DynSeg   Dynamic Segmentation

GIS   Geographic Information Systems

GPS   Global Positioning System

GTFS   General Transit Feed Specification

ISTT   Inter-Stop Travel Time

ITS   Intelligent Transportation Systems

LE   Linear Extrapolation

LI   Linear Interpolation

LR   Linear Referencing

LRS   Linear Referencing System

PT   Public Transit

SC-AFC   Smart Card Automated Fare Collection Systems

SCD   Smart Card Data

TT   Travel Time

CHAPTER 1

INTRODUCTION

The importance of Public Transit (PT) is increasing significantly with the effects of rapidly increasing urban population and mobility needs to be met in a sustainable way. According to the United Nations (UN) vision, the rate of urbanization is expected to increase, reaching 68% in 2050. This included a projection for relative change of 34% in the rural population in Turkey for the 2018-2050 period (United Nations, 2018), which is expected to increase the mobility demand in the urban regions. Besides the commute trips between work and home, PT facilitates people to access basic services in daily life (i.e. hospitals, schools, universities, shopping malls, governmental buildings, etc.); thus, accessibility with PT is a crucial issue, especially for low-income people who cannot afford private means of transportation. On the other hand, it is important to increase the desirability of the PT services so that more people use them and thus, share the cost increasing the economic sustainability of the system.

A key issue in the preferability of the PT services is the reliability of the system, which includes the punctuality of the bus services. Since the complexity of the PT network and operations in the cities increases exponentially with the size of the city (coverage area, urban and suburban services, etc.) and PT supply (routes, stops, service frequency and hours, etc.), it is very challenging to monitor and manage them manually and is generally supported by the Intelligent Transportation Systems (ITS) services.

ITS can be defined as an integrated system that implements a wide range of technologies, including vehicle sensors, communication and control tools, to solve and manage transportation and traffic problems. Despite the existing ITS experience for more than two decades in the developed countries, there is rather limited experience with ITS services in developing ones, especially with the Advanced

Public Transportation System (APTS) services. The usage area of APTS is to increase the operational efficiency of PT and ridership by improving the reliability of the transportation system. Most APTS applications require Geographic Information Systems (GIS) or web-based platforms. While web-based platforms have an advantageous side in giving real-time information to users, GIS platform makes available advanced spatial analysis techniques when needed. On the other hand, the Global Positioning System (GPS) is a crucial technology that takes place in ITS applications (Singh and Gupta, 2015). Today's GPS equipped vehicles are facilitating the establishment of the APTS systems. The vehicle trajectory consists of spatio-temporal data (including time and coordinates) for the vehicle/driver on a PT network. Use of smart cards further enables the collection of spatio-temporal data of the passengers (portraying the PT ridership), which is mostly the boarding information due to its convenience and connection to automated fare collection (AFC) systems.

1.1 Motivation

ITS applications generally monitor vehicle (bus, tramway, etc.) in real-time and keep trajectory data for security and advance notice. As an added value, it is also possible to use this trajectory data for monitoring and increasing the performance and quality of the PT network by deriving and evaluating the PT reliability measures. Although vehicle GPS-based trajectory data is a sufficient data source for deriving PT reliability measures, the storage, process and analysis of such huge datasets bring together management difficulties considering large and complex PT networks, especially the bus networks and routes.

Most metropolitan municipalities have GIS-based network data and/or GIS-based route, stop and schedule information to some extent. But, local governments and their PT administrations should manage the PT trajectory data on a well-planned geographic information system and create network datasets with routes, stops and schedules to accompany these data in specific standard formats like General Transit

Feed Specification (GTFS). However, most PT administrations do not /cannot store bus trajectory GPS datasets over long periods, since these datasets cover a massive amount of disk place and are not effectively utilized for further evaluation, yet. However, the use of PT smart card datasets can be more effective and accessible, as they cover a considerable amount of less disk space and include more detailed data about passengers, which was the main motivation behind this study. But, the key issue is the assessment of how effective it would be to use smart card data (SCD) to estimate the PT reliability and what the prerequisite conditions should be in terms of GIS-based data storage. There are some major challenges in this context:

- First, the lack of a GIS-based network dataset and a GTFS-based schedule data makes it difficult to carry out further analysis on PT reliability. Although it is possible to obtain a GIS-based network dataset from some commercial and non-commercial road network data sources, adapting these network datasets to the PT network of a large city requires a separate and challenging process.

- Secondly, relating and mapping raw smart card point data to the relevant route segments is another major problem. Smart card records contain GPS based location data (latitude and longitude) recorded during a smart card reading. GPS signals can be inaccurate due to the lost or corrupt signals and these points can deviate from their actual route segments. Without using a GIS-based network dataset with a GTFS based stop and schedule dataset, a route-based trajectory production can be possible by conducting some geoprocessing task in the GIS environment.

- Thirdly, although some PT reliability measures like schedule adherence and bus bunching studies need to schedule data in a specific format, some of the PT reliability studies like travel time (TT) reliability can be conducted over the resulting smart card dataset after the above-mentioned processes.

1.2   Scope of the Study

This study aims to answer the following research questions:

i.   How effective is it to use SCD to estimate the PT reliability?

ii.   What are the prerequisites in using SCD in PT reliability estimation?

iii.   How should the SCD be used to estimate PT reliability?

The scope of this study includes

a) the development of a framework to create vehicle time-space diagram (showing the location of the bus along its predetermined route) using SCD with GIS-based techniques and

b) generation of PT performance and reliability measures.

Konya PT-SCD and bus trajectory data (BTD) have been used for the case study. In the study framework;

➢ A set of data preprocessing actions have been conducted by using a database management system (DBMS) for rapid converting of trajectory data to a proper format in order to facilitate the further processes and preprocesses in the GIS environment.

➢ BTD was converted to create time-space diagrams using GIS-based techniques known as Linear Referencing (LR) and Dynamic Segmentation (DynSeg).

➢ After LR and DynSeg steps, arrival time estimations (ATE) of each bus trip to the designated bus stops and determination of the inter-stop travel time (ISTT) were conducted by linear interpolation and extrapolation calculations using SCD and BTD.

➢ A comparison between BTD and SCD has been conducted by using times-space diagrams to see how the smart card dataset is affective in creating time-space diagrams of bus trajectory. As a result, these obtained data tables are visualized by time-space diagrams to evaluate the travel-time reliability

of the public transit network. Finally, the selected reliability indicators have been calculated and evaluated by using statistics of total travel times of the bus trips.

1.3   The Layout of the Thesis

In Chapter 2, a literature review was presented to give a background on the use of SCD and GIS in PT operations, followed by GIS-based techniques employed in this study. A short summary is also provided in the last part of the chapter to introduce the parameters used to evaluate PT reliability. Chapter 3 mainly introduces the methodology followed in the study, including the overall depiction of the research framework and description of the major parameters and processes used, while details of the GIS-based model buildings and pseudo codes were presented in the Appendix part in more details.

In Chapter 4, PT network in Konya is introduced briefly, followed by example calculations of the selected PT reliability parameters from the SCD, some of which are compared to the values obtained from the original BTD of the same trips. Time-space diagrams, ISTT diagrams, and the TT reliability analysis for selected PT lines will be presented. The final chapter includes the conclusion and future recommendations of this study.

CHAPTER 2

BACKGROUND

2.1   Smart Card Use in Public Transit

Smart card use is a common technology in PT nowadays. Although several improved forms of smart cards like NFC mobile wallets or credit cards has emerged with the improving technology, the conventional smart cards are very portable devices as the size of a credit card and they are basically produced to store data for the purpose of identification, authorization, and payment (Lu, 2007). Smart card technology has improved in years and many other features have been added to the technology after the 1970s, which is the years of the invention (Shelfer and Procaccino, 2002). The use of smart card technology is varying from health care to telephone services, banking, human resources and transportation. Today, smart card-based AFC systems have a wide variety of usage in the area of public transit networks all over the world and, of course, in Turkey.

Essentially, AFC data included in a database comprised of records for card transactions. Each transaction is recorded when a smart card is tapped-on and/or of in a PT vehicle like a bus. Each AFC system can store transaction data at different levels of details in the AFC database. Basically, the data stored in these systems include the location and time of boarding and/or alighting stops (Faroqi et al., 2018). Some AFC systems do not include the name or id of the boarding and/or alighting stops, only include the location data in terms of longitude and latitude. Contrary to the boarding transactions, the alighting transactions may not be recorded in the dataset in some AFC systems. AFC data can also include the route id and the direction of the vehicle. Other information such as fare types, transfers and sociodemographic information may be included in the AFC data. The type of

validation in terms of just tap-on or both tap-on and tap-off is related to the fare structure. This can be flat, distance-based, or zone-based (Devillaine et al., 2012).

Although AFC systems are mainly established to gather revenue more efficiently, archived or real-time output data from these systems are used to conduct research on many subjects. Several studies have been conducted in recent years on the use of SCD in public transit. Pelletier et al. (2011) have been categorized these studies under three titles. First is the strategic-level studies that related to long-term network planning, demand forecasting and customer behavior analysis. Second is the tactical-level studies that related to the schedule adjustment, longitudinal and individual trip patterns. Finally, operational-level studies are related to smart card system operations and supply-demand indicators (Pelletier et al., 2011). Especially the category of operational-level studies is subject to this study.

The output dataset from the smart card AFC (SC-AFC) systems can be used to calculate the public transit performance measures from passenger tap-in locations, vehicle-kilometers and schedule adherence for each trip (run), route and day (Trépanier et al., 2009). Timestamps included in the smart card records can also be used to estimate the average arrival and boarding time at bus stops for a specific route and vehicle. In fact, smart cards can collect data indirectly, similar to the way in which automated vehicle location (AVL) systems do so (Hickman, 2002).

On the other hand, there could be some faulty in the records of smart card transactions. Steve Robinson et al. (2014) was identified as the causes of faculties in the records of smart card systems in their study. They propose a set of rules to aggregate individual rides into a single trip then provide an approach to identify erroneous tap-out data caused by several problems.

Fourie (2014), in his study, reconstructed the bus vehicle trajectories by using the full day of a tap-in and tap-out smart card transactions in Singapore city. Some records with GPS errors are filtered by removing transactions having the stop identifiers that irrelevant to the assigned route. After calculating the speed between consecutive transactions, transactions with speed higher than 80 km/h are removed

from the dataset. The trajectories and the provided GTFS based transit schedule for each route were recorded in a format compliant with the outputs of an agent-based simulation system called MATSim, which uses the specifications developed by Rieser (2010). The outputs in this format were visualized and analyzed possible in compliant software called Senozon Via.

2.2   GIS Use for Smart Card Data (SCD)

PT administrations are using SC-AFC for revenue management and administrative reasons in terms of monitoring of traveler's access to the service. On the other hand, these data collected from SC-AFC systems bring some crucial spatio-temporal information about vehicle and passenger movement. When integrated with an AVL system, if combined with GIS and other related network data, SC-AFC is a very beneficial system for the development of public transit travel survey data. (Chapleau et al., 2008)

Many SC-AFC and AVL systems are not designed or established to determine public transit performance measures. Thus, supplementary data preprocessing and visualization procedures are needed. GIS software can provide powerful tools to process, produce, analyze and visualize the spatial information. Some studies were carried out to integrate public transit data from SC-AFC and AVL to GIS for presenting transit performance measures or developing transit traveler information systems.

 In Chapleau et al. (2011), the data from SC-AFC, GPS and GIS were integrated to build a framework for evaluating public transit performance in a GIS platform. Liao and Liu (2010) developed a stand-alone visualization software interface to conduct the time point–level travel time and schedule adherence analysis by using the data produced by AVL and SC-AFC in Minnesota. The processed information can be integrated into a mapping system to improve public transit service and optimize the transit route/schedule. Public transit performance measures in this study include

travel time analysis, reliability analysis for schedule adherence and travel time variation. Curries and Mebah (2011) developed a GIS visualization platform. This platform was used to examine the variations in spatio-temporal patterns to determine the transit performance of the tram network in Melbourne, Australia. They have used the ArcGIS software to determine public transit performance in a GIS environment.

Brown and Racca (2012) generated TT reliability measures for two corridors in New Castle County by using the vehicle GPS data of State Fleet. This study was conducted for various times of day, days of the week, and seasons. The processing was done using the tools that were included in ArcGIS Desktop software and Excel spreadsheets. A linear referencing system (LRS) that was created in advance for a previous study was used to group the continuous road segments into a single route. After that, processed GPS data was referenced with the road centerlines in the LRS. GPS readings can, therefore, be mapped directly to a route and milepost, independent of how links (route segments) were constructed (segmentation scheme) or the cartographic representation. The referencing system also provided the ability to determine distances within links more accurately. The direction of each vehicle is identified in the standard referencing scheme, and direction is essential when estimating trip time. A GIS-based network dataset with a routable LRS was used in this study. After processing GPS records in the LRS, two types of TT reliability indexes buffer index and planning index was calculated to evaluate the reliability. This study was conducted without using a GTFS or non-GTFS based schedule information.

2.3   ArcGIS Software Use with Model Builder

Many commercial or non-commercial GIS software packages provide LR tools. ArcGIS Desktop software also provides a set of LR tools along with many other tools used in many studies about public transit subjects. It allows for creating maps, performs spatial analysis, and manage data. Multiple data formats can be imported

and powerful analytical tools and workflows can be used to identify spatial patterns, trends, and non-obvious relationships. (Esri, 2019)

In addition to the studies mentioned in the previous section of this chapter, ArcGIS software is used in many different studies. Curries and Mesbah (2011) developed a GIS platform in ArcGIS software to visualize the spatial and temporal patterns of changes in Melbourne, Australia, for analyzing the transit performance in their study. Zhang et al. (2018), in their article, used ArcGIS software to calculate passenger flow data within a specified period.

Model Builder is a visual programming language for building geoprocessing workflows. This application takes place in ArcGIS software, and it can provide the output of one tool to another tool as input. The Model Builder is a handy application since it facilitates tedious works conducted under a GIS and helps to solve some problems encountered in a study without getting into deep coding. The tools created under the model builder can be converted to Python scripting codes and implemented to other codes or other models created under model builder.

Bigham and Kang (2013) developed a geoprocessing model using the ArcGIS Model Builder. A database file is imported and a new field is created for route identifier to geo-reference the collision events to the mileposts in the LRS. A batch process was developed using ArcGIS Model Builder to automatically produce the continuous routes by iterating all the geoprocessing tasks needed in the study of Dayan et al. (2014).

2.4   Linear Referencing (LR)

LR is a term that emerged in the engineering area. Basically, it consists of locating an object along a linear feature like rivers pipelines and roads by referencing the location of that object to another fixed location. (Curtin et al., 2007). The well-known example of a LR is the mile markers throughout the Highways in USA and Kilometer markers in Turkey (Figure 2.1). In other words, LR is a methodology for locating

data of objects (polygon, line or point) based on a position along a route (Route and Measure) rather than X/Y coordinates (Esri, 2019).

Figure 2.1. Mile and KM markers

The LR term itself covers a group of concepts and techniques used for relating the features to their actual locations along a network, rather than referencing those locations to a traditional spherical or planar coordinate system. We use LR when the location of events and objects on a network are more important than the location in 2D or 3D space. (Curtin et al., 2019)

LR can be used for several tasks. Management, monitoring and analyzing of highways and streets, pipelines, railways, oil and gas exploration and water resources can be conducted by LR. Transit applications are another area of LR used. LR is a critical component in transit applications, and it facilitates such activities as:

• Route planning and analysis

• Automatic vehicle location and tracking

• Bus stop and facility inventory

• Rail system facility management

• Track, power, communications, and signal maintenance

• Accident reporting and analysis

• Demographic analysis and route restructuring

- Ridership analysis and reporting
- Transportation planning and modeling

(Esri, 2019)

Users often want to record many additional attributes about the roads. Without the use of LR, it could require that roads must be split into many smaller segments at each location where an attribute value change. As an alternative, these situations can be handled as LR events along the roads, as in Figure 2.2.
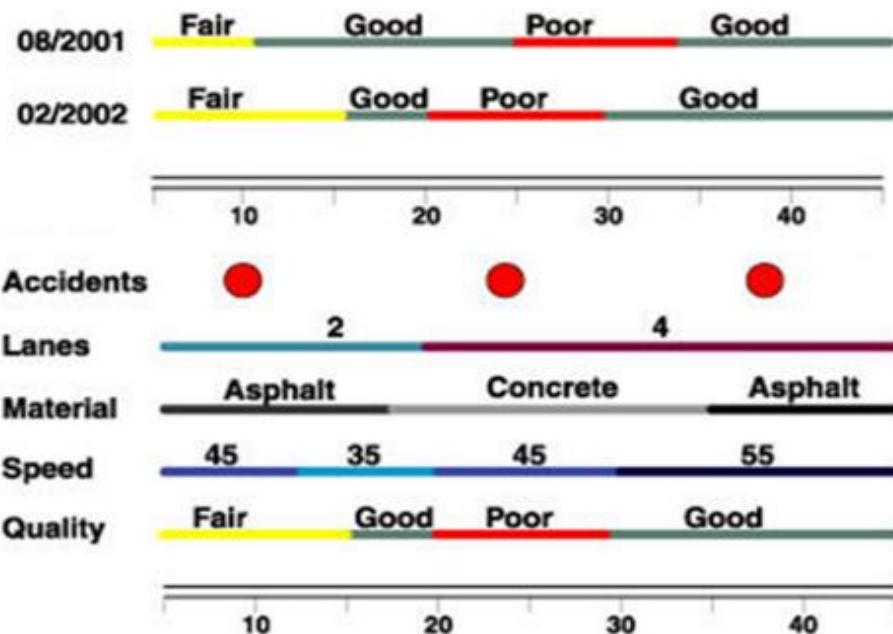
Figure 2.2. Multiple sets of attributes for road features. (Curtin et al., 2019)

There are some GIS software packages which can be used to implement a LRS. But, especially the documents of these software packages focus on creating the events and locating these events to the linear features. There are also some analyzing tools to overlay and intersect different linearly referenced dataset. However, these documents do not mention about data preparation or data capture phase of the study. Hence it is not easy to implement LR unless following a pre-defined model or

framework. Moreover, creating and managing an up to date LRS in advance for a specific transit network could be beneficial for further studies.

2.4.1　　　Route Feature Classes

To locate an event along a linear feature, this linear feature needed to be converted to a route feature class. A route feature is a line feature that has a defined measurement system. This measurement system can be used to locate events, assets, and conditions along linear features. These types of feature classes are called "M aware" feature class in terms of "M aware Point" or "M aware Polyline" (Figure 2.3). Differently from a line or polyline, route term refers to any linear feature class such as highway, river or pipeline that has a common measurement system and a unique identifier field in its attribute table. "Polyline M" values can be seen in the *SHAPE* field of the sample attribute table of a route feature class in Figure 2.4.
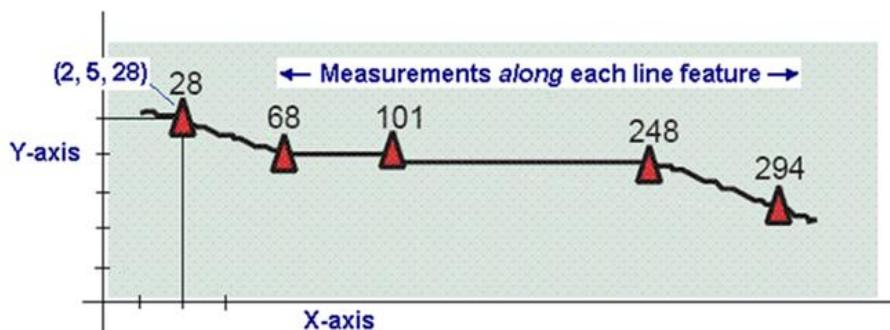
Figure 2.3. Route feature class (Esri, 2019)

Figure 2.4. Route attribute table (Esri, 2019)

2.4.2   Linear Referencing (LR) Tasks

The main tasks to conduct in a LR process are;

•   Creating Routes

•   Creating Event Tables

•   Creating Event Layers

As the first main task of a LR process, the concept and requirement of the route was mentioned in the previous title. Creating event tables is may be the core part of the LR process and can be conducted by using the "Locate features along routes" tool. Event tables include the events in terms of entities that are linearly referenced along routes. Events can be the mile markers or traffic accidents along highways or construction zones along railways, or pipe diameters along pipelines as mentioned in previous titles. Creating event layers is the third main task and can be conducted by using the "Make Route Event Layer" tool. This task is also called Dynamic Segmentation. All Linear Referencing tasks provided in ArcGIS software can be seen in Figure 2.5.
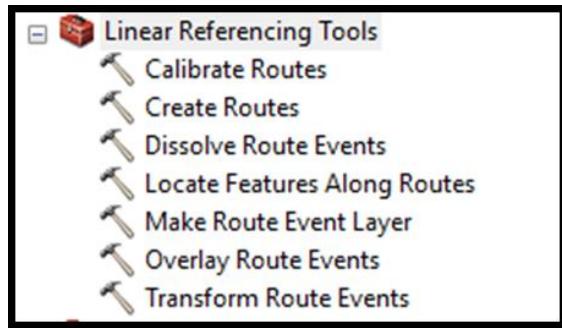
Figure 2.5. Linear Referencing (LR) Tasks in ArcGIS Toolbox.

2.4.3   Dynamic Segmentation (DynSeg)

The DynSeg process consists of computing the locations of events on a map. These events must be stored and managed in an event table by using a LRS in advance. DynSeg facilitates locating the line or point features along a route, without splitting or segmenting features each time an attribute value changes. Segments, defined by the measures in the event tables created by the LR process, can be "dynamically" located. DynSeg provides associating multiple sets of attributes with any portion of an existing linear feature, no matter where it begins or ends. These sets of attributes can be displayed, queried, analyzed and edited without affecting the original linear feature's geometry. (Esri, 2019)

The result of the DynSeg is an event layer. Event layers are layers in ArcGIS that are created from an event table. These event layers can be further exported to a feature class in a geodatabase or shapefile. We need primarily two data types to implement LR in a GIS environment: a) Route feature classes and b) Event tables. Essentially, locating the events in the event tables to the line features on a map called DynSeg. (Figure 2.6)

Transmission Pipe 10 = A1

Event ID	Pipe ID	From_M	To_M	Rating
1	A1	0	3	fair
2	A1	3	75	good
3	A1	75	10	poor
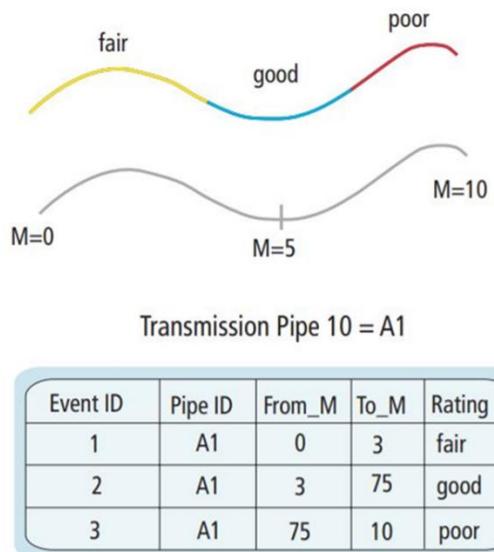
The dynamic segmentation process for a pipeline

Figure 2.6. Dynamic Segmentation (Esri, 2019)

2.4.4 Steps of Linear Referencing (LR) Process

Before starting to use a LRS, we need to define the steps necessary for implementing the objects of the study. Curtin et al. (2007) were defined a general set of processes to follow while implementing such a system in their study.

This Linear Referencing process consists of a seven-step procedure (Figure 2.7).

1. Identifying an application to which LR is pertinent and using that information to decide what network representation should be employed and the topological rules that must be followed.

2. Determining the route structure or the underlying datum to which events can be linearly referenced.

3. Identifying the way in which measurements will be made along those routes.

4. Defining the way in which linear events will be defined, captured, and maintained.

5. The cartographic output of the linearly referenced events.

6. Analyzing the events until all of them referenced.

7. Keep the linearly referenced data for later use in case of sharing with others or using with different applications to analyze. (Curtin et al., 2007)

Although this seven-step procedure can clarify the application of LR in terms of basic steps to take, a more specific study needs to be done considering the provided dataset and problems encountered in the study.
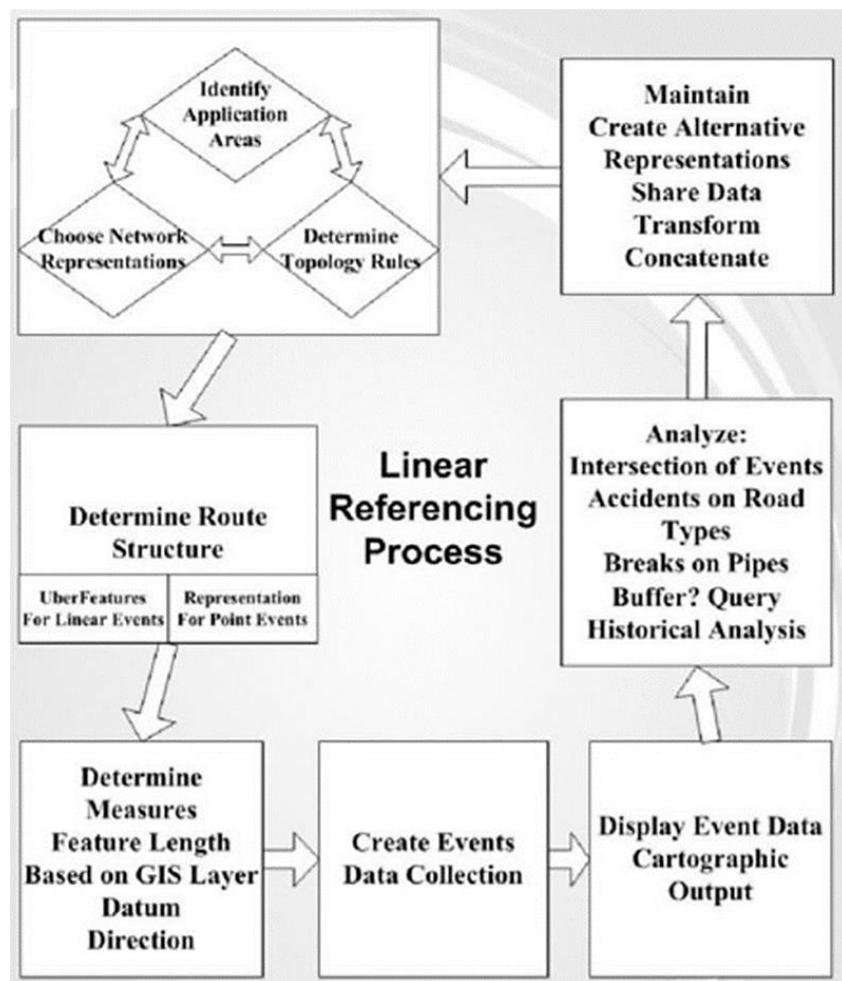
Figure 2.7. Linear Referencing (LR) Steps (Curtin et al., 2007)

18

2.4.5 Challenges and Disadvantages of LR in Bus PT

There are GIS software packages that can be used to implement a LR system. Especially the documents of these software packages focus on creating the events and locating these events to the linear features. However, these documents do not mention about data preparation or data capture phase of the study.

Mostly straight lines as train lines and pipelines or highways are more suitable and easier to implement for LR since these types of lines are not as complicated as public transit bus lines in terms of intersections U-turns and round trips. Train lines, pipelines and highways mostly consist of one-way straight lines without a neighboring close line. Another disadvantage of implementing LR for a bus public transit line is a possible detour of the vehicle. Train lines, pipelines and highways follow a fixed-line. A detour is not possible in these types of lines.

Hence it is not easy to implement LR, unless following a pre-developed process order specialized for the study dataset.

2.5 Public Transit Performance Measures

Measuring the performance of a transit system was stated as the first step toward efficient and proactive management (Bertini and Geneidy, 2003). As PT departments are required to provide quality service under increasing demand in ever-growing urban regions with limited resources, the use of performance measures for PT planning and operations is gaining more attention. In this regard, the use of APTS in PT provides many crucial data for PT management to inspect transit performance measures.

Transportation Research Board (TRB), an independent adviser of the Federal Government in the USA and one of the major divisions of the US National Research Council, published the first edition of the Transit Capacity and Quality of Service

Manual (TCQSM) in 1999 (Danaher et al., 1999). This manual recommended the evaluation of transit systems using six performance measures:

- Service frequency,
- Hours of service,
- Service coverage,
- Passenger loading,
- Reliability,
- Transit vs. automobile travel time

Typically, these measures are determined for a service on a particular day from surveys or on-board statistics. (Trépanier et al., 2015)

2.5.1 PT Reliability Studies

Reliability is one of the main factors in terms of performance measure for PT, some which are;

- Ticket price,
- Accessibility in time and space,
- Travel time spent from origin to destination,
- Comfort,
- Image of the system,
- Service reliability.

The last of these main quality aspects, Service Reliability, expresses whether the actual passenger journey meets the expected quality aspects such as waiting, travel time and comfort (OORT, 2011). One of the early studies over PT reliability is "Transit Service Reliability", as stated in a technical report prepared by the US Urban Mass Transportation Administration (Abkowitz et al., 1978). A comprehensive overview about transit service reliability is explained and a framework for research studies presented. In this report, it is indicated that transit service reliability is a significant determinant of traveler mode and departure time choices and it is crucial

in influencing the costs of providing transit service. The report recommends some reliability measures for use in future evaluation studies. In this report, reliability measures determined both for the traveler and the operator oriented. Traveler oriented measures are:

- Total travel time,
- Wait time (including transfer time),
- In-vehicle time,
- Seat availability.

Operator oriented measures are:

- Adherence to schedule at origin-destination and intermediate points,
- Headways,
- Seat availability.

In this thesis, only travel time-based reliability measures which is one of the traveler-oriented measures, could be studied since other ones need the schedule information and alighting (tap-out) information that our case study is lacks of.

2.5.2   Travel Time (TT) Reliability

According to the definition by the US Department of Transportation (USDOT), Federal Highway Administration (FHWA), primary resource for identifying the measures of TT reliability, travel time reliability was the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day (USDOT, 2006). Depending on congestion or incidents in traffic or any other reason like weather conditions, travel times can vary from expected travel times. In this context, TT reliability is more important than average travel times because travelers tend to remember unexpected delays (Lyman and Bertini, 2008).

Travelers can manage their times when TT reliability information is provided. On the other hand, reducing the traffic congestion can be an indirect result. Transit TT reliability brings together efficiency and service attractiveness to the public transit

network (Kittelson and Associates, 2003). The variability in travel time for a specific route has a direct relation and a negative correlation with TT reliability. The variability in travel times is an important factor for passengers in terms of guaranteeing their on-time travels. (Rakha et al., 2010)

Song (2018) studied to model spatial-temporal patterns of bus delays between bus stops using data from the vehicle GPS. A LR based group of processes applied to the dataset. (Figure 2.8). As a result, Song (2018) presented weekly and daily time-space diagrams of the bus trips to determine the variances of travel time between trips in terms of scattering in the diagrams and evaluated the source of the delay visually (Figure 2.9).

2.5.3 Travel Time (TT) Reliability Indicators

The conventional statistical measures, such as coefficient of variation and standard deviation, have been used to evaluate the TT reliability. However, it is not that easy for passengers with no technical knowledge to understand these statistical measures. Besides, these statistics treat early and late arrivals with equal weight. On the other hand, passengers tend to remember late arrivals more than early ones. The Federal Highway Administration has recommended several route-level transit TT reliability indicators that can be used to measure the variance of in-vehicle travel time and to quantify the TT reliability. Two of them are;

 • **90th or 95th percentile travel time,** which represents the travel time in the worst case traffic conditions,

 • **Buffer index**, which measures the extra time a passenger needs to spend in addition to the average travel time to ensure on-time arrival for 95% of the trips. The extra time can be defined as the time difference between 95th percentile travel time and average travel time, and the buffer index is then calculated as the ratio of the extra time and average travel time. The smaller

the buffer time index is, the more reliable a transit route is. (Ma and Wang, 2014)

In this study, the 95% travel time and buffer index are adopted to measure the transit travel time for a particular route. In addition, the segment-level TT reliability indicator of each transit route is also calculated for further evaluation.
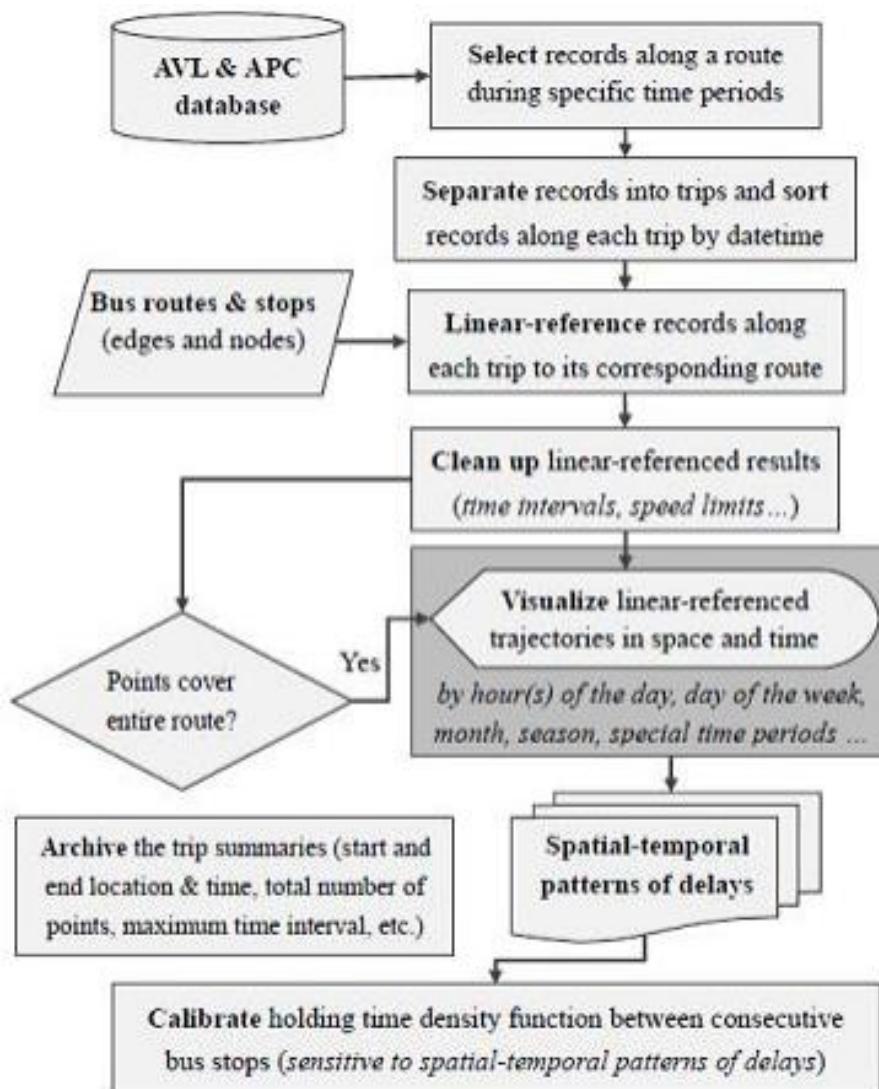
Figure 2.8. The process to model bus delays between stops. (Song, 2018)
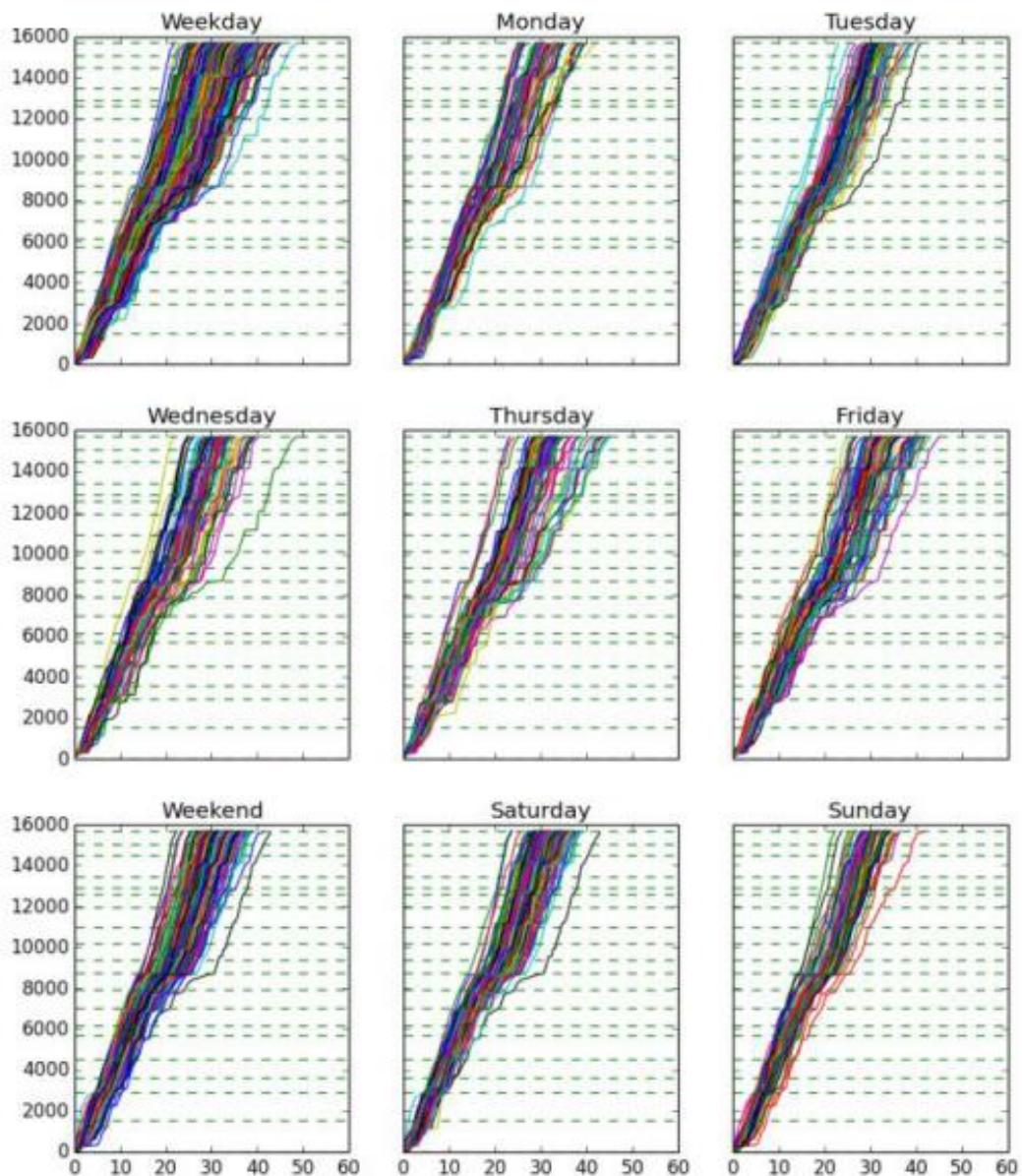
Figure 2.9. Linear referenced trips by days of the week. (Song, 2018)

The other TT reliability indicators are travel time index and planning time index. The travel-time index is calculated by the ratio of the average travel time to the free-flow travel time. The planning-time index (PTI) is calculated by the ratio of the 95th percentile travel time to the free-flow travel time. Because the only difference between the TTI and the PTI is that 95th percentile travel times are used instead of

24

the average travel times, the PTI is very useful to compare with the TTI. (Rakha et al., 2010). Pu (2011) listed the TT reliability indicators recommended by different sources and compare them by examining the advantages and disadvantages. (Figure 2.10)

Travel Time Reliability Measure	Lomax et al. (2003)	FHWA Guide (2006)	NCHRP Report 618 (2008)	SHRP 2 (2008)
95th or other percentile travel time	N/A	√	N/A	N/A
Standard deviation	N/A	×	×	N/A
Coefficient of variation	N/A	×	×	N/A
Percent variation	√	N/A	√	N/A
Skew statistic	N/A	N/A	N/A	√
Buffer index	√	√	√	√
Planning time index	N/A	√	√	√
Frequency of congestion	N/A	√	N/A	N/A
Failure rate (percent on-time arrival)	N/A	N/A	√	√
Misery index	√	N/A	√	√

NOTE: √ = use encouraged; × = use discouraged; N/A = not applicable.

Figure 2.10. Travel time (TT) reliability indicators recommended by different sources. (Pu, 2011)

The most recommended indicator is the Buffer Index by all four sources. On the other hand, Pu (2011) examined the Buffer Index in two categories. The conventional Buffer Index is the mean (average) based Buffer Index, which is calculated by dividing to the mean travel times. The second Buffer Index examined in the study is the median-based Buffer Index, which is calculated by dividing the median of the travel times. After analysis of both, the median-based Buffer Index is recommended, especially when travel time distributions become heavily right-skewed.

Anderson (2019) evaluated urban arterial reliability performance indicators in his study by using three reliability indicators. These were Buffer Index, Normalized Standard Deviation and Planning Index. The travel times dataset gathered from Bluetooth sensors placed at arterial highways are used to calculate these indicators.

He also compared these three indicators by calculating the correlation between them. Eventually, the indicator chosen to be used in the following parts of the study was the Planning Index. It is indicated in the study that Planning Index is an outlier free indicator since it is calculated based on 95th Percentile Travel Time.

2.5.4 Estimation of Travel Times

Llnaes (2017) conducted a network-based study for the analysis and monitoring of the bus network using GPS datasets. First, a GIS-based network dataset was created with the help of GTFS based schedule, route and stop dataset. Then travel times between route segments were estimated. These segments are the sub-routes or the minimal units of routes that are created for monitoring the behavior of buses. Linear interpolation is used to discover the timestamps of segment endpoints between GPS observations to estimate the arrival time of buses.

Chu and Chapleau (2008) proposed methods for the estimation of the arrival times to the bus stop using temporal constraints. Given the constraints, the arrival times are estimated by using linear interpolation where there is not any smart card reading (boarding). Also, linear interpolation is used in the estimation of the vehicle position between the last boarding and the destination terminus, assuming that the vehicle will arrive at the terminus at the scheduled time. If the calculated speed since the last smart card reading too high or low, linear extrapolation with an average speed is used for estimation.

A GTFS based detailed schedule and the records of the alighting passengers (tap-off data) are provided in these studies just like in many other studies.

2.6 Findings of the Literature Review

By examining all these studies in the literature, it has become clear whether the reliability measure of a PT network can be obtained (and what kind of reliability

criteria can be derived) from the available SCD . The studies and the methods investigated in this chapter showed that SCD can be processed and analyzed using LR tools in the GIS environment, and as a result, travel time reliability measures can be evaluated by generating time-space diagrams. In addition, by calculating travel time reliability indicators, PT reliability measures can be derived numerically. On the other hand, almost all studies investigated in this chapter have been used rich datasets, including both tap-on and tap-off SCD, GIS-based network datasets and a detailed schedule dataset in GTFS or other software-specific formats.

Additionally, most of these studies have been conducted over non-complex PT networks like tram or train relative to the bus networks. The provided dataset in this study belongs to a complex bus PT network with a tap-on only passive smart card system. Detailed schedule information is not provided in GTFS or another format. Even so, the studies in the literature show that establishing a customized LRS for bus PT network by developing and using GIS tools with Model Builder in ArcGIS software can solve the problems with this kind of poor dataset. In the following methodology chapter, the generated framework and the methods to conduct these processes will be explained in more detail.

CHAPTER 3

METHODOLOGY

To obtain PT reliability measures from SCD, it is necessary to establish the time and space relationship for PT line and the stops along. Since many SC-AFS are established for the purpose of fare allocation, smart card datasets lack of boarding and alighting stop information, which is a critical data for these types of studies. While it is relatively easy to derive reliability measures from SC-AFS of the train or tram lines with a fixed route, strict schedule and rich of information in terms of boarding and alighting stops, it is hard to derive reliability measures for a bus PT which is lack of these specifications. On the other hand, another challenge is, these systems are not based on GIS and a LRS, which facilitates analyzing reliability. The first phase of the methodology is obtaining the kilometers (aka Station Km or St Km in short) of the trajectory data. In order to obtain kilometer values traveled for a vehicle that following a route from the beginning, we need to convert location data from x/y (latitude/longitude) domain to M (measure) domain. GIS techniques called LR and DynSeg can be used to obtain kilometer measures of the vehicle from the beginning of the route.

The second phase of the methodology is estimating the bus-stop arrival times by using linear interpolation and extrapolation calculations. While current GIS software packages providing spatial interpolation tools in X/Y domain, they do not provide a ready to use tool for linear interpolation in the M domain. Therefore, these calculations are conducted in Microsoft Excel, which is a spreadsheet software providing ready to use linear interpolation functions.

3.1   Framework

A methodology has been developed to derive PT reliability measures from SCD by using GIS-based techniques (see Figure 3.1) with the following steps:

I) Static Route Data Analysis (performed for each bus line):

> **I.1** Producing "**bus route feature classes**" from **"bus lines**" by converting the inter-stop segmented polylines to "polyline m objects (routes)" by LR tasks.

> **I.2** Generating bus stop station kilometers (St Km) using LR tasks.

II) Dynamic Bus Line Service and Performance Analyses (repeated for each bus service):

> **II.1**   Preprocessing of smart card and/or bus trajectory in DBMS and GIS environment.

> **II.2**   Generating route and trip based St Km of smart card and bus trajectory data by implementing LR and DynSeg to the daily dataset.

> **II.3**   Postprocessing for data cleaning and rearranging

III) PT Performance measure estimations:

> **III.1**   Merging SCD with bus-stops

> **III.2**   Arrival time estimations (ATE) of each bus stop by linear interpolation using kilometers and timestamps.

IV) Preparation of the outputs:

> **IV.1**   Generating times-space diagrams
> **IV.2**   Inter-stop travel time (ISTT) diagrams.
> **IV.3**   TT reliability indicators

V) Evaluation of Reliability

> **V.1**   Evaluating reliability in terms of ATE.
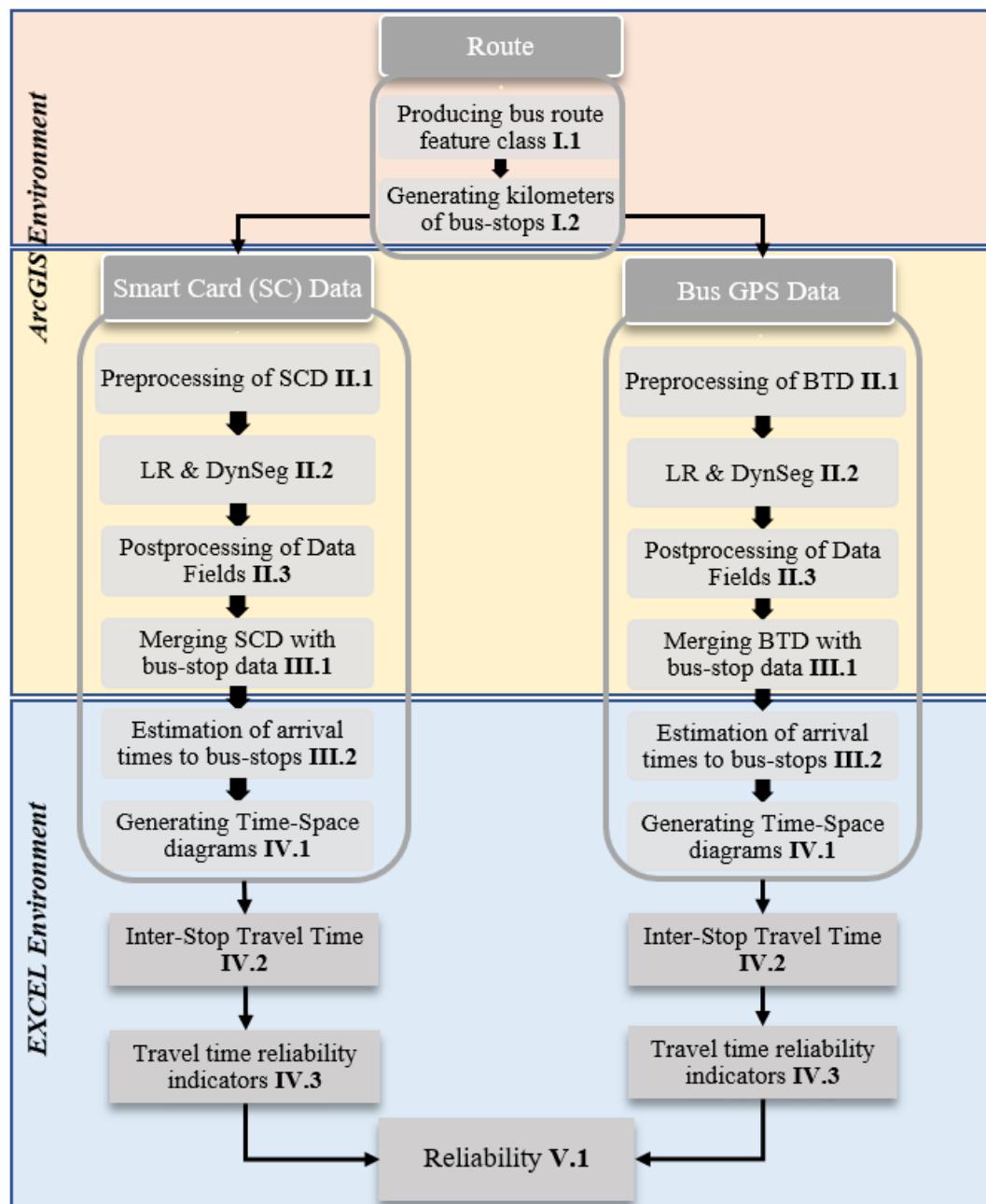
Figure 3.1. The framework of the methodology

While the major concepts of these steps are discussed in the following subsections, the details of the step/procedures are presented in the Appendix for further discussion and implementations.

Key Points to Consider

There is a strict hierarchy between datasets and processes in the developed framework. The diagram in Figure 3.2 presents the relation between datasets and processes along this hierarchy.

LR: Linear Referencing, DynSeg: Dynamic Segmentation, LI: Linear Interpolation, LE: Linear Extrapolation
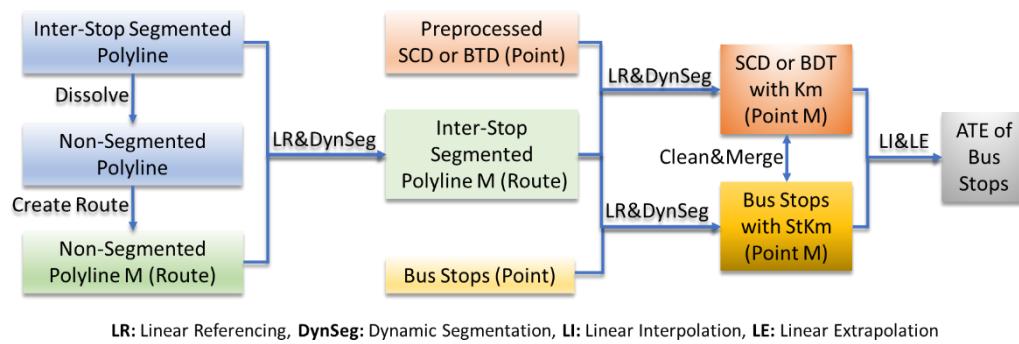
Figure 3.2. Hierarchy and the relation of datasets with the processes

Coordinate System: Raw bus trajectory and smart card data contain geographic coordinates (latitude and longitude) in WGS84 Datum. A projected coordinate system has to be used in order to conduct geoprocessing tasks and to represent the bus and smart card locations coordinates precisely. All location data has to be converted to TUREF_TM33, which is an ITRF based geodesic datum corresponding to the central meridian of Konya.

Common Route Identifiers: It is necessary to carefully define input and output route identifiers in order to conduct a precise LR and DynSeg. Managing and using route identifiers carefully can help us to avoid matching the BTD/SCD points to the wrong route segments, where these segments are too close to each other.

Sorting Attribute Table: Since all processes will be conducted in linear order at the further interpolation and extrapolation operations while estimating arrival times, all objects in a dataset (feature class/.csv/xlsx) must be sorted according to the bus stop order, timestamp and kilometers, respectively.

32

Other definitions and abbreviations of GIS Tools and Concepts to consider are listed in Table A.1 and Table A.1.2 in the Appendix section.

3.2   Static Route Data Analyses

This process aims to obtain a route feature class necessary for the LR of bus-stop locations, BTD and SCD. In the static "Route" analysis step, to produce bus route feature class from the polyline feature class (Step I.1),

- Inputs: Multi-part (inter-stop segmented) polyline feature classes
- Outputs: Single-part and multi-part route feature classes

The details of the tasks conducted to produce these outputs will be explained in this section.

The provided bus line object at the beginning of the process consists of some polyline types of objects between bus stops, as can be seen in Figure 3.3a. As can be seen in the shape field of the initial attribute table from Figure 3.3b, all segments between bus stops are of "polyline" object type. Each polyline object has attribute values that identifying the bus stop names and orders at the beginning and the end of the segment. Besides, the direction attributes of each object in terms of inbound and outbound is occurring in the attribute table as "0" and "1". The attribute list and the description of the initial bus line and the final bus route are listed in Table 3.1.

The route feature classes need to be comprised of inter-stop segmented "polyline m" type of objects having cumulatively continuing station kilometers at each bus-stop. However, it is not possible to create a route feature class directly from this feature class without losing its attributes. On the other hand, we cannot obtain the "from measure" (*FMEAS*) field and "to measure" (*TMEAS*) field that indicates cumulative measure values in the attribute field of the resulting route feature class. These fields are necessary for precisely identifying and matching of SCD and BTD to the related polyline object between bus stops in the further LR processes. Hence, a single-part route feature class needs to be created first as an intermediate product. In the next

phase of this methodology, we will use this intermediate single part route feature class to produce the actual route needed for us. The geoprocessing steps for creating such an intermediate route feature class in the GIS environment are as follows:

- Create a single part polyline feature class using the "Dissolve" tool from the "Data Management " toolbox.

  Input: Polyline feature class of bus line segmented (split) between bus-stops.

  Output: Single part polyline feature class of bus line. The resulting attribute table of the feature class in this step is as shown in Figure 3.3c.

- Create a single-part route feature class by using "Create Route" tool from the "Linear Referencing" toolbox.

  Input: Single-part polyline feature class of bus line.

  Output: Single-part route feature class as an intermediate product (see Figure 3.4) .

(b) LINE 381

OBJECTID *	Shape *	ULID	DIR_CHG_CH	DEST	S_ORD_1	S_NO_1	Y1	X1	S_ORD_2	S_NO_2	Y2	X2	R_LENGTH	Shape_Length
1	Polyline	381	0	SAMANPAZARI	1	2191a	37.86635	32.47722	2	1652a	37.86652	32.47838	0.118658	118.657544
2	Polyline	381	0	SAMANPAZARI	2	1652a	37.86652	32.47838	3	11a	37.86705	32.48261	0.377065	377.064994
3	Polyline	381	0	SAMANPAZARI	3	11a	37.86705	32.48261	4	26a	37.86818	32.48601	0.391688	391.68845

(c) ULID381SPLT_Dissolved

OBJECTID *	Shape *	ULID	Shape_Length
1	Polyline	381	7827.952762

Figure 3.3. Initial (a) inter-stop segmented bus line and (b) attribute table, (c) resulting attribute table of a single part (non-segmented) bus line (polyline)

Table 3.1. Initial and resulting field attributes of bus lines

34

Polyline Field Name	Attribute Description	Notes
Initial Data		
ULID	Unique Line ID	LID*10+SLID
DIR_CHG_CH	Direction Value (Inbound/Outbound)	0/1
DEST	Destination	-
S_ORD_1	Order Value of the Current Bus Stop	1, 2 , 3, …
S_NO_1	Number of the Current Bus Stop	-
Y1	Latitude of the segment origin	-
X1	Longitude of the segment origin	-
S_ORD_2	Order Value of the Next	2, 3, 4, …
S_NO_2	Number of the Next Bus Stop	-
Y2	Latitude of the segment destination	-
X2	Longitude of the segment destination	-
Resulting Data		
FMEAS	Cumulative Km from Bus Stop	-
TMEAS	Cumulative Km to Bus Stop	-
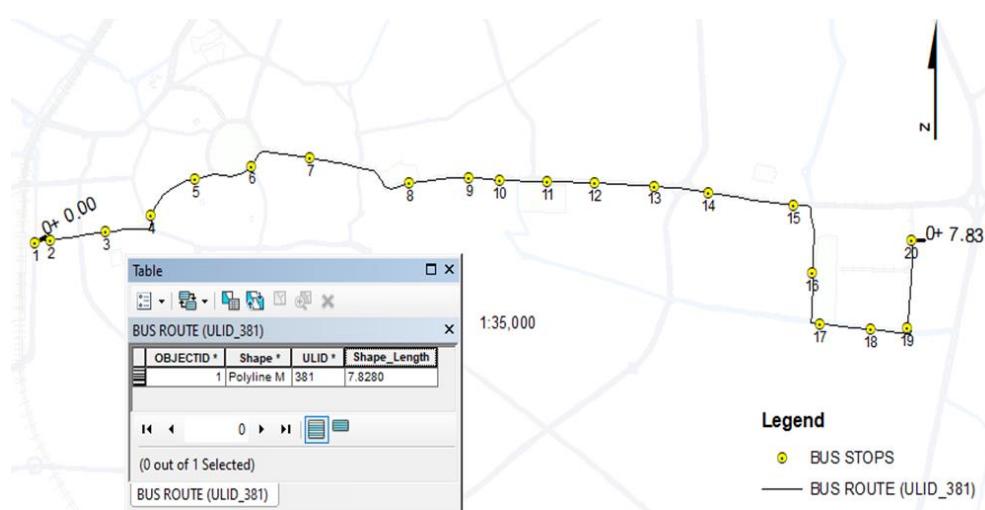
Figure 3.4. Single part (non-segmented) route feature class and attribute table (polyline m).

After the creation of this intermediate single-part route feature class, following LR and DynSeg tasks will be conducted in the GIS environment to produce the final route feature class;

- Create the event table of route segments between bus-stop locations by using the "Linear Referencing" tool from the "Linear Referencing" Toolbox with,

  Input: Polyline feature class of bus line split between bus-stops, Single-part route feature class of bus line,

  Output: Event Table of route segments between bus-stop locations (see Figure 3.5a)

- The DynSeg of the Event Table to the single-part route feature class with,

  Input: Event Table of route segments and the single-part route feature class.

  Output: Event Layer of the inter-stop segmented route feature class (see Figure 3.5b)

- Exporting the Event Layer to the feature class with,

  Input: Event Layer of the segmented route feature class

  Output: Inter-stop segmented route feature class of bus line (see Figure 3.5c).

ULID381SPLT_EventTable

OBJECTID*	RID	FMEAS	TMEAS	ULID	DIR_CHG_CH	DEST	S_ORD_1	S_NO_1	Y1	X1	S_ORD_2	S_NO_2	Y2	X2	R_LENGTH
1	381	0	0.1187	381	0	SAMANPAZARI	1	2191a	37.86635	32.47722	2	1652a	37.86652	32.47838	0.118658
2	381	0.1187	0.4957	381	0	SAMANPAZARI	2	1652a	37.86652	32.47838	3	11a	37.86705	32.48261	0.377065
3	381	0.4957	0.8874	381	0	SAMANPAZARI	3	11a	37.86705	32.48261	4	26a	37.86818	32.48601	0.391688

(b)

ULID381_Event_Layer

OBJECTID*	RID	FMEAS	TMEAS	ULID	DIR_CHG_CH	DEST	S_ORD_1	S_NO_1	S_ORD_2	S_NO_2	R_LENGTH	LOC_ERROR	Shape*
1	381	0	0.1187	381	0	SAMANPAZARI	1	2191a	2	1652a	0.118658	NO ERROR	Polyline M
2	381	0.1187	0.4957	381	0	SAMANPAZARI	2	1652a	3	11a	0.377065	NO ERROR	Polyline M
3	381	0.4957	0.8874	381	0	SAMANPAZARI	3	11a	4	26a	0.391688	NO ERROR	Polyline M

(c)

OBJ	RID	FMEAS	TMEAS	ULID	DIR	DEST	S_ORD_1	S_NO_1	S_ORD_2	S_NO_2	R_LENGT	LOC_ERRO	Shape*	Shape_Leng
15	38	0	0.1187	381	0	SAMA	1	2191a	2	1652a	0.118658	NO ERROR	Polyline M	118.657544
5	38	0.1187	0.4957	381	0	SAMA	2	1652a	3	11a	0.377065	NO ERROR	Polyline M	377.064994
10	38	0.4957	0.8874	381	0	SAMA	3	11a	4	26a	0.391688	NO ERROR	Polyline M	391.68845
13	38	0.8874	1.3053	381	0	SAMA	4	26a	5	27a	0.417927	NO ERROR	Polyline M	417.926798
2	38	1.3053	1.7096	381	0	SAMA	5	27a	6	2418a	0.404277	NO ERROR	Polyline M	404.277221
6	38	1.7096	2.1623	381	0	SAMA	6	2418a	7	2411d	0.452715	NO ERROR	Polyline M	452.715486
4	38	2.1623	2.9089	381	0	SAMA	7	2411d	8	21a	0.74656	NO ERROR	Polyline M	746.560339
18	38	2.9089	3.3098	381	0	SAMA	8	21a	9	1014a	0.400878	NO ERROR	Polyline M	401.530357

Legend

⊙ BUS STOPS

— BUS ROUTE (ULID_381)

Figure 3.5. (a) Event table, (b) event layer and (c) resulting attribute table and route feature class after the LR & DynSeg processes for the generation of inter-stop segmented route feature class.

As can be seen in Figure 3.5c, now the "hatches" which are the markers in terms of kilometers, can be labeled at the beginning and at the end of each inter-stop segment cumulatively. This means that every smart card or BTD around an inter-stop segment can now be linearly referenced to the correct segment with their cumulative kilometers from the beginning of the route. The flowchart (Figure A.4.1) and the pseudocodes (Table A.4.1) for the route production explained until now was presented in Appendix A.4

Generation of Bus-Stop Feature Classes with Station Kilometers (Step I.2)

The provided bus-stop objects consist of point features class with stop order numbers and directions (see Figure 3.6.a). This process aims to obtain the station kilometers of bus-stops (see Figure 3.6.d) that will be necessary in the further part of this study while conducting the GIS preprocessing of the SCD/BTD part and estimating the bus

arrival times to the bus stops. The attribute descriptions of the initial bus-stop data and the resulting data are listed in Table 3.2

Table 3.2. Attribute descriptions for the initial data

Field Name	Attribute Description
Initial Data	
ULID	Unique Line ID
DIR_CHG_CH	Direction Values
DEST	Destination
S_ORD_1	Order Value of the Bus Stop
S_NO_1	Number of the Bus Stop
Y	Latitude
X	Longitude
Resulting Data	
MEAS	Cumulative Kilometer of the Bus Stop

The following LR and DynSeg tasks have been conducted respectively in the GIS environment to generate the bus-stop feature class with station kilometers;

- Creating the event table of bus-stop station kilometers by using the "Linear Referencing" tool with,

  Input: Bus-stop feature class, inter-stop segmented route feature class

  Output: Event Table of bus-stop station kilometers (see Figure 3.6.b)

- The DynSeg of the Event Table using the "Locate Features Along Route" tool with,

  Input: Event Table of bus-stop kilometers and inter-stop segmented route feature class of bus line

  Output: Event Layer of the bus-stop locations over the route with kilometer attributes (see Figure 3.6.c))

- Exporting Event Layer to the feature class with,

  Input: Event Layer of the bus-stop feature class

  Output: Bus-stop feature class with kilometer attributes (see Figure 3.6.d)

(a)

STOP381

FID*	Shape*	S_NO_1	DIR_CHG_CH	S_ORD_1	DEST	ULID	S_NAME_1	X	Y
1	Point	2191a	0	1	SAMANPAZARI	381	Tren Garı	32.47722	37.86635
2	Point	1652a	0	2	SAMANPAZARI	381	İstasyon	32.47838	37.86652
3	Point	11a	0	3	SAMANPAZARI	381	Eski Stat	32.48261	37.86705

(b)

STOP381_EventTable

OBJECTID*	RID	MEAS	S_NO_1	DIR_CHG_CH	S_ORD_1	DEST	ULID	S_NAME_1	X	Y
1	381	0	2191a	0	1	SAMANPAZARI	381	Tren Garı	32.47722	37.86635
2	381	0.1187	1652a	0	2	SAMANPAZARI	381	İstasyon	32.47838	37.86652
3	381	0.4957	11a	0	3	SAMANPAZARI	381	Eski Stat	32.48261	37.86705

(c)

STOP381_Event_Layer

OBJECTID*	RID	MEAS	S_NO_1	DIR_CHG_CH	S_ORD_1	DEST	ULID	S_NAME_1	X	Y	LOC_ERROR	Shape*
1	381	0	2191a	0	1	SAMANPAZARI	381	Tren Garı	32.47722	37.86635	NO ERROR	Point M
2	381	0.1187	1652a	0	2	SAMANPAZARI	381	İstasyon	32.47838	37.86652	NO ERROR	Point M
3	381	0.4957	11a	0	3	SAMANPAZARI	381	Eski Stat	32.48261	37.86705	NO ERROR	Point M

(d)

BUS STOPS

OBJECTID*	Shape*	RID	MEAS	S_NO_1	DIR_CHG_CH	S_ORD_1	DEST	ULID	S_NAME_1	X	Y	LOC_ERROR	F_TYPE	TS_TYPE
1	Point M	381	0	2191a	0	1	SAMANPAZARI	381	Tren Garı	32.47722	37.86635	NO ERROR	STOP	1
2	Point M	381	0.1187	1652a	0	2	SAMANPAZARI	381	İstasyon	32.47838	37.86652	NO ERROR	STOP	1
3	Point M	381	0.4957	11a	0	3	SAMANPAZARI	381	Eski Stat	32.48261	37.86705	NO ERROR	STOP	1

Figure 3.6. (a) Initial, (b) event table, (c) event layer and (d) resulting attribute tables the generation of bus-stop kilometers

After exporting to a feature class, the attribute table of the resulting event layer is as in Figure 3.6.d. Finally, we have "Point M" type of bus-stop feature class, and kilometer values increasing cumulatively from the beginning of the route are in the "*MEAS*" field. Additionally, "*F_TYPE*" and "*TS_TYPE*" fields need to be added final bus-stop feature class for the further processes in the arrival time estimation part of this study to separate bus stop data from BTD or SCD.

The detailed flowcharts (Figure A.5.1) and the pseudo-codes (Table A.5.1) for the generation of bus-stop kilometers explained in this section were presented in Appendix A.5.

3.3   Dynamic Bus Line Service and Performance Analyses

The aim of the steps explained in this section is to generate the kilometers of BTD/SCD by matching and referencing the correct inter-stop segments of the produced bus route feature class. Maybe the most challenging part and the core of the proposed methodology is takin place in this section. This challenging part has been accomplished by using the generated GIS tools for the preprocessing and the LR tasks.

The dynamic bus line service and performance analyses comprised of the following three steps in the framework (II.1, II.2 and II.3). Dynamic analyses can be performed with either "BTD" or "SCD" in almost the same way with minor differences. Initially, two-phase (DBMS and GIS) of preprocessing for BTD/SCD need to be conducted. While preprocessing of BTD/SCD in DBMS and GIS environments (Step II.1), input and output datasets are as follows;

- Inputs: Daily bus trajectory/smart card data and trip data
- Output: Single trip BTD feature classes

The flowchart, pseudocodes and the developed GIS tool to conduct these processes were presented in Appendix A.6 and subsections.

Step II.2 has LR and DynSeg actions for the trips in BTD/SCD with;

- Inputs: Single trip BTD/SCD feature classes
- Output: Single trip BTD/SCD "point m" type of feature classes with kilometers.

The flowchart pseudocodes and the developed GIS tool to conduct these processes were presented in Appendix A.7 and subsections. On the other hand, the details and results of this step will be discussed in section 3.3.1.

Postprocessing of the fields (cleaning and rearranging the fields) of BTD/SCD in GIS Environment (Step II.3) has;

- Inputs: Single trip BTD/SCD feature classes

- Output: Single trip BTD/SCD feature classes with the necessary fields

The flowchart and pseudocodes and the developed GIS tool to conduct these processes were presented in Appendix A.8.

3.3.1 Linear Referencing (LR) of SCD and BTD Feature Classes

After the preprocessing phase, smart card and bus trajectory feature classes are ready for LR tasks. Both for smart card and bus trajectory, the LR procedure is almost the same. The only difference is the input dataset type. While the input of the bus trajectory dataset has to be shift by shift, the input smart card dataset has to be in a daily set of data. The output dataset will be trip by trip for both kinds of datasets. Therefore, the following geoprocessing task have been conducted for the LR phase (Step II.2);

- Select the trip from the daily SCD to be processed by using the "Select Layer by Attribute" tool (The preprocessed feature class of a BTD is already in one trip format.)
- Use the "Locate Features Along Route" tool from "Linear Referencing" Toolbox

The input and output datasets in the LR part of this step are as follow;

Input: The feature class of BTD/SCD and inter-stop segmented bus route feature class.

Output: Event Table of the kilometers for the smart card or BTD.

After accomplishing the LR task, the resulting event tables generated from smart card (a) and bus trajectory (b) data were presented in Figure 3.7. The redundant CID values, as can be seen in these event tables, will be eliminated after the DynSeg process. This elimination will be done by filtering the rows that providing the "*SUBRID = S_ORD_1*" condition. This process was conducted under the "Filtering and Exporting" part of the generated GIS tool (see Figure A.7.2).

(a) SC443OCT30V211_1_EVENT_TABLE

OBJ	SUBRID	MEAS	DAYINDEX	VID	TT	CID	CT	TS	LID	SLID	INDEX	ULID	SHFTID	STripID	DULTripID	AsgnStopOrd	AsgnStat
1	1	0.0031	1504	211	BILET	DF45BAB5B29EDC53C349983B779C0	0	2/20/2010 19:31:00	44	3	2	443	443_211	1	104	1	1
2	129	62.2986	1504	211	BILET	DF45BAB5B29EDC53C349983B779C0	0	2/20/2010 19:31:00	44	3	2	443	443_211	1	104	1	1
3	124	60.0268	1524	211	BILET	2BDD4AEC92B2994B891F22AF1BF36	2	2/20/2010 19:34:00	44	3	3	443	443_211	1	104	7	1

AsgnStat	AsgnStopID	Ulid_Vid_Tr	DIR_CHG_CH	FMEAS	TMEAS	ATUS	DESTINATION	USID	S_ORD_1	S_NO_1	S_ORD_2	S_NO_2	LOC_ERROR
1	1022d	443_211_1	0	0.0000	0.2874	44D	BOSNA SANAYI	ULID443_1_2_1022d_1021d	1	1022d	2	1021d	NO ERROR
1	1022d	443_211_1	0	0.0000	0.2874	44D	BOSNA SANAYI	ULID443_1_2_1022d_1021d	1	1022d	2	1021d	NO ERROR
1	1016d	443_211_1	0	2.2749	2.7496	44D	BOSNA SANAYI	ULID443_8_7_1015d_1016d	7	1015d	8	1016d	NO ERROR

(b) BGPS443OCT30V211D224_1618_1900_EVENT_TABLE

OBJ	SUBRID	MEAS	SID	PID	VID	TS	SATNO	H	V	LID	DID	DAY_ID	ULID	MV_DIR_F	MV_DIR_T	ShftID	shft_intrvl	tripID
1	443	0.1335	27289253	81	211	10/30/2018 18:40:38	13	27	58	44	224	2503382	443	1	2	443_211_22	2018-10-30 18:18:03_ 2018-10-30 18:18:03	1
2	443	62.1679	27289253	81	211	10/30/2018 18:40:38	13	27	58	44	224	2503382	443	1	2	443_211_22	2018-10-30 18:18:03_ 2018-10-30 18:18:03	1
3	443	0.0000	27289231	59	211	10/30/2018 18:40:28	13	26	55	44	224	2503153	443	1	2	443_211_22	2018-10-30 18:18:03_ 2018-10-30 18:18:03	1

tripID	SEGTID	MEAS	F_TS	L_TS	DIR_CHG_CH	RID	FMEAS	TMEAS	ATUS	DEST	USID	S_ORD_1	S_NO_1	S_ORD_2	S_NO_2
1	ULID443_1022d_1021d_1_2	0.1335	10/30/2018 18:40:28	10/30/2018 18:40:48	0	443	0.0000	0.2874	44D	BOSNA S	ULID443_1_2_1022d_1021	1	1022d	2	1021d
1	ULID443_1022d_1021d_1_2	62.1679	10/30/2018 18:40:28	10/30/2018 18:40:48	0	443	0.0000	0.2874	44D	BOSNA S	ULID443_1_2_1022d_1021	1	1022d	2	1021d
1	ULID443_1022d_1021d_1_2	0.0000	10/30/2018 18:40:28	10/30/2018 18:40:48	0	443	0.0000	0.2874	44D	BOSNA S	ULID443_1_2_1022d_1021	1	1022d	2	1021d

Figure 3.7. Resulting event tables for (a) SCD and (b) BDT after LR

3.3.2   Dynamic Segmentation of Event Table

The resulting event table from the LR phase needs to be converted to a feature class for further operations. The following tasks are conducted to obtain these feature classes;

- Use "Make Event Layer" tool from "Linear Referencing" toolbox to locate smart card or BTD on the map.

Input: The event table of the SCD/BTD and inter-stop segmented bus route feature class

Output: Event Layer of the BTD/SCD (see Figure 3.8)

- Use the "Select Layer by Attribute" tool to filter redundant CID values by selecting data where the referenced root ids are equal to stop order values (*SUBRID = S_ORD_1"* condition).
- Export selected objects in the event layer to a feature class.
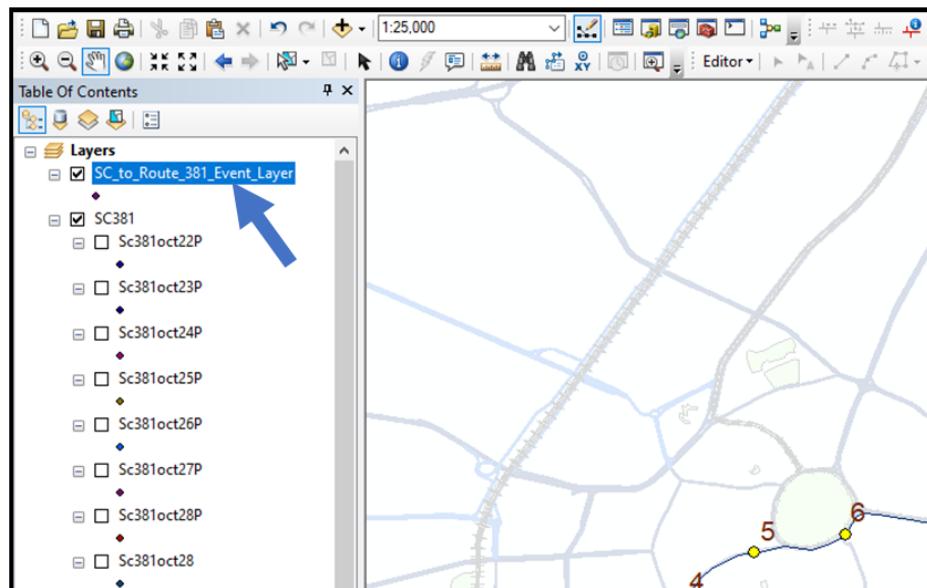
Figure 3.8 Resulting event layer in the table of contents
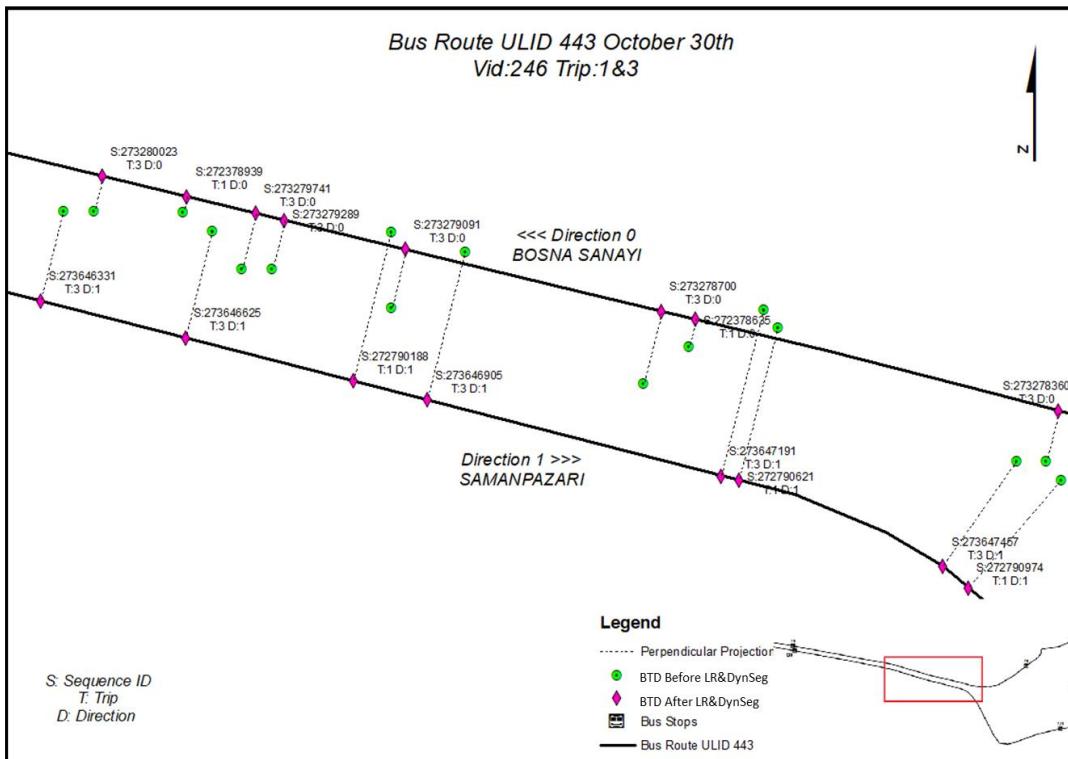
Figure 3.9 Before and After the Linear Referencing

After the LR and DynSeg of BTD or SCD conducted by the steps explained above, the perpendicular projections of these points to the correct bus line segments are obtained precisely. BTD before and after LR tasks are shown in Figure 3.9. The resulting points after the LR and DynSeg processes are shown by diamond-shaped points. These points are labeled by SequenceID, trip and direction values in the figure. Even the points close to the wrong side of the route segment are matched to the correct route segment, trip and direction after the process.
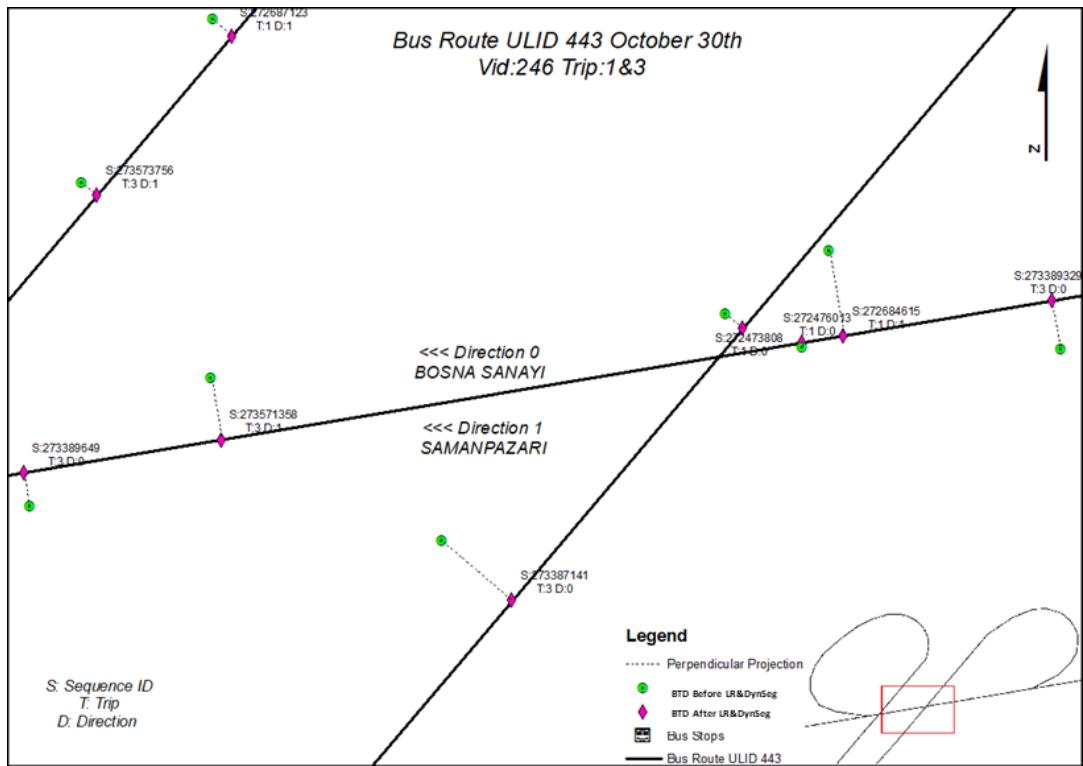
Figure 3.10. Before and after the Linear Referencing (LR) in a bi-directional route segment.

Figure 3.10 shows a part of the first intersection area from Route ULID 443. There is a coinciding bi-direction on the same segment, one is inbound another one is outbound. Even in this kind of complex route parts, the smart card or BTD points have matched the correct route segments and directions through the methodology used in this study. The resulting sample attribute table of a smart card and BTD after the LR and DynSeg are shown in Figure 3.11. In this attribute table, the kilometers beginning from the first bus stop of the bus route have been populated in the "MEAS" field.

The flowcharts of the LR and DynSeg for BTD (a) and SCD (b) are presented in Figure A.7.1. The pseudo-codes for BTD and SCD are presented in Table A.7.1 and Table A.7.2. The developed LR & DynSeg Tool in Model Builder is presented in Figure A.7.2.

Sc443oct22PT443_211_1

OBJ	Shape *	SUBRID	MEAS	DAYINDEX	VID	TT	CID	CT	TS	INDEX	ULID	ShiftID	STripID	DULTripID	AsgnStopOrd
1	Point M	1	0.0031	1504	211	BILET	DF45BAB5B29EDC53C349963B779C0	0	2/20/2010 19:31:00	2	443	443_211	1	104	1
2	Point M	7	2.2775	1524	211	BILET	2BDD4AEC92B2994B891F22AF1BF36	2	2/20/2010 19:34:00	3	443	443_211	1	104	7
3	Point M	9	3.4109	1526	211	BILET	FB6F34A2A06E56BC16FD4C75BA833	0	2/20/2010 19:37:00	5	443	443_211	1	104	
4	Point M	9	3.4109	1525	211	BILET	E92A31037B58050AA672FDA11D4EE	0	2/20/2010 19:37:00	4	443	443_211	1	104	
5	Point M	11	4.0427	1537	211	ABILE	D3250D11A0FB88461D53C52DB541F6	2	2/20/2010 19:39:00	6	443	443_211	1	104	11
6	Point M	11	4.0427	1538	211	BILET	9CE3D6DDC05E68A7F7255D15D3B85	2	2/20/2010 19:39:00	7	443	443_211	1	104	11
7	Point M	11	4.0427	1539	211	BILET	3449404D67C395DDE29603160721D8	37	2/20/2010 19:40:00	8	443	443_211	1	104	11

DULTripID	AsgnStopOrd	AsgnS	AsgnStopID	Ulid_Vid_Trip	DIR_CHG_CH	FMEAS	TMEAS	ATUS	DEST	USID	S_ORD_1	S_NO_1	S_ORD_2	S_NO_2
104	1	1	1022d	443_211_1	0	0.0000	0.2674	44D	BOSNA SANAYI	ULID443_1_2_1022d_1021d	1	1022d	2	1021d
104	7	1	1016d	443_211_1	0	2.2749	2.7496	44D	BOSNA SANAYI	ULID443_8_7_1015d_1016d	7	1015d	8	1016d
104	10	1	1130a	443_211_1	0	3.1077	3.4179	44D	BOSNA SANAYI	ULID443_9_10_1014d_1130a	9	1014d	10	1130a
	10	1	1130a	443_211_1	0	3.1077	3.4179	44D	BOSNA SANAYI	ULID443_9_10_1014d_1130a	9	1014d	10	1130a
104	11	1	2407a	443_211_1	0	4.0231	4.7114	44D	BOSNA SANAYI	ULID443_12_11_3d_2407a	11	3d	12	2407a
104	11	1	2407a	443_211_1	0	4.0231	4.7114	44D	BOSNA SANAYI	ULID443_12_11_3d_2407a	11	3d	12	2407a

bgpsu443o30v211d224_1618_1900_R_T443_211_224_1

OBJ	Shape *	MEAS	SID	PID	VID	TS	SATNO	H	V	LID	DID	DAY_ID	ULID	MV_DIR_F	MV_DIR	ShiftID	shft_intrvl	tripID	segtid
783	Point M	0.0000	27289231	59	211	10/30/2018 16:40:28	13	26	55	44	224	2503153	443	1	2	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1022d_1024...
806	Point M	0.0558	27289234	60	211	10/30/2018 16:40:33	13	26	52	44	224	2503189	443	1	2	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1022d_1021d...
444	Point M	0.1335	27289253	61	211	10/30/2018 16:40:38	13	27	58	44	224	2503362	443	1	2	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1022d_1021d_1...
841	Point M	0.2142	27289292	62	211	10/30/2018 16:40:43	13	27	56	44	224	2503712	443	1	2	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1022d_1021d...
161	Point M	0.2890	27289326	63	211	10/30/2018 16:40:48	13	27	51	44	224	2504034	443	1	2	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1021d_1020d_2_...
1404	Point M	0.3563	27289364	64	211	10/30/2018 16:40:53	13	27	45	44	224	2504398	443	2	3	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1021d_1020d_2_3
429	Point M	0.4224	27289400	65	211	10/30/2018 16:40:58	13	27	50	44	224	2504717	443	2	3	443_211_22	2018-10-30 16:18:03_2	1	ULID443_1021d_1020d_2_3

segtid	F_TS	L_TS	DIR_CHG_CH	FMEAS	TMEAS	ATUS	DEST	USID	S_ORD_1	S_NO_1	S_ORD_2	S_NO_2
ULID443_1022d_1021d_1_2	10/30/2018 16:40:28	10/30/2018 16:40:48	0	0.0000	0.2674	44D	BOSNA SA	ULID443_1_2_1022d_1021d	1	1022d	2	1021d
ULID443_1022d_1021d_1_2	10/30/2018 16:40:28	10/30/2018 16:40:48	0	0.0000	0.2674	44D	BOSNA SA	ULID443_1_2_1022d_1021d	1	1022d	2	1021d
..._1022d_1021d_1_2	10/30/2018 16:40:28	10/30/2018 16:40:48	0	0.0000	0.2674	44D	BOSNA SA	ULID443_1_2_1022d_1021d	1	1022d	2	1021d
...1022d_1021d_1_2	10/30/2018 16:40:28	10/30/2018 16:40:48	0	0.0000	0.2674	44D	BOSNA SA	ULID443_1_2_1022d_1021d	1	1022d	2	1021d
...1022d_1021d_1_2	10/30/2018 16:40:28	10/30/2018 16:40:48	0	0.2674	0.6242	44D	BOSNA SA	ULID443_2_3_1021d_1020d	2	1021d	3	1020d
...43_1021d_1020d_2_3	10/30/2018 16:40:43	10/30/2018 16:41:36	0	0.2674	0.6242	44D	BOSNA SA	ULID443_2_3_1021d_1020d	2	1021d	3	1020d
ULID443_1021d_1020d_2_3	10/30/2018 16:40:43	10/30/2018 16:41:36	0	0.2674	0.6242	44D	BOSNA SA	ULID443_2_3_1021d_1020d	2	1021d	3	1020d

Figure 3.11 Resulting SCD (a) and BTD (b) attribute table after DynSeg.

3.4 Performance Measures and Outputs

Evaluation of reliability can be examined by

  i. Visual Evaluation of Time-Space Diagrams

  ii. Numeric Evaluation of Travel Time

Bus trajectories for each trip and each day will be converted to time-space diagrams in order to visually evaluate the reliability of routes in the Public Transit Network. Arrival times of bus stops with or without smart card records (departure) need to be estimated to generate these diagrams. Inter-stop travel time diagrams can only be generated after the estimation of all bus-stop arrival times. For this purpose, the following outputs are generated.

3.4.1   Arrival Time Estimations (ATEs)

Initially, bus-stops with station kilometers and BTD/SCD with kilometers obtained as a result of previous steps have to be merged as pointed in the framework (Step III.1). This step aims to create a spreadsheet (Excel) table on which the arrival time estimations can be conducted. The input and output datasets used in this step is as follows;

Inputs: Single trip BTD/SCD feature classes with kilometers and bus-stops feature class with station kilometers

Outputs: Merged bus-stop and BTD/SCD spreadsheet (Excel) tables with station kilometers for every single trip in a day.

The flowchart, pseudocodes, resulting attribute tables and the developed GIS tool to conduct these processes were presented in Appendix A.9.

After the creation of the spreadsheet (Excel) tables needed for the estimation of arrival times to bus-stops, estimation of arrival times to the bus-stops in Excel (Step III.2) is achieved with,

Inputs: Merged bus-stop and BTD/SCD tables with station kilometers for every single trip in a day.

Output: Estimated arrival times (ATEs) of bus-stops for each bus service.

The flowchart and the pseudocodes developed to conduct these processes were presented in Appendix A.10.

The arrival times of bus-stops in a route, estimated by interpolating or extrapolating the timestamps according to the kilometers and timestamps of previous and subsequent smart card or BTD. The following interpolation formula have used if T stands for time and S stands for kilometer elapsed;

$$T_i = T_{i-1} + (S_i - S_{i-1}) * (T_{i+1} - T_{i-1}) / (S_{i+1} - S_{i-1})$$

Since this interpolation have to be calculated in linear order in terms of SCD/BTD and bus-stop kilometers (MEAS), a linear interpolation method has to be applied.

Even in the latest version, ArcGIS does not provide a linear interpolation tool. All interpolation tools in ArcGIS are in X, Y domain. This linearly referenced data is in the M domain. Although no need to use GIS software for the calculations in a tabular dataset like this. Therefore, the built-in function "FORECAST.LINEAR" of Microsoft Excel Software has been used for linear interpolation and extrapolation calculations. Other built-in functions, "MATCH" and "OFFSET" are used in combination with "FORECAST.LINEAR" to find and match the previous and subsequent closest records for the linear interpolation calculation.

The ATE of the bus-stops that takes place at the beginning of a route and does not have a preceding smart card or BTD have been calculated by extrapolation. Likewise, the ATE of the bus-stops that takes place at the end of a route and does not have a subsequent smart card or BTD have been calculated by extrapolation. The same workflow has conducted in the estimation process for both bus trajectory and smart card datasets, as seen in and Figure A.10.1 and Figure A.10.2. Pseudo codes of arrival time estimation for bus-stops were presented in Table A.10.1 and Table A.10.2.

After conducting thorough the processes presented in the flowcharts, the following resulting tables have been obtained for the smart card (a) and bus trajectory (b) dataset are presented in Figure 3.12. The data in the rows indicated by a polygon are belong to the bus-stops. The timestamps of these bus-stops are now estimated by linear interpolation and extrapolation. The timestamps belonging to the bus stops at the beginning of the tables are extrapolated timestamps. The ones in the rows between the smart card or BTD are interpolated timestamps. Besides, ISTTs have been calculated by using the differences between these resulting timestamps of bus-stops.

(a)

OBJECTID	MEAS	TS	STOP ATE	TS NRM	S ORD_1	S ORD_2	F_TYPE	TS_TYPE	DIR_CHG_CH	Min_Trip_TS	Max_Trip_TS	DEST	ShiftID	tripID
1	0.00	14:31:30	14:31:30	00:00:00	1	2	STOP	1	0					
2	0.00	14:31:30	#N/A	00:00:00	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
3	4.23	14:31:35	#N/A	00:00:05	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
4	5.64	14:31:37	#N/A	00:00:07	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
5	8.63	14:31:43	#N/A	00:00:13	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
6	35.08	14:31:47	#N/A	00:00:17	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
7	96.84	14:31:52	#N/A	00:00:22	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
8	174.73	14:31:57	#N/A	00:00:27	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
9	258.24	14:32:02	#N/A	00:00:32	1	2			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
10	267.35	14:32:03	14:32:03	00:00:33	2	3	STOP	1	0					
11	343.27	14:32:07	#N/A	00:00:37	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
12	406.52	14:32:12	#N/A	00:00:42	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
13	450.63	14:32:17	#N/A	00:00:47	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
14	475.64	14:32:22	#N/A	00:00:52	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
15	480.04	14:32:25	#N/A	00:00:55	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
16	484.43	14:33:08	#N/A	00:01:38	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
17	506.30	14:33:12	#N/A	00:01:42	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
18	556.47	14:33:17	#N/A	00:01:47	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
19	620.78	14:33:22	#N/A	00:01:52	2	3			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1
20	624.19	14:33:22	14:33:22	00:01:52	3	4	STOP	1	0					
21	695.74	14:33:27	#N/A	00:01:57	3	4			0	14:31:30	16:23:14	BOSNA SANAYI	443_246_261	1

(b)

OBJECTID	TS	STOP_ATE	TS_NRM	MEAS	S_ORD_1	S_ORD_2	AsgnStopOrd	F_TYPE	TS_TYPE	DIR_CHG_CH	DEST	Ulid_Vid_Trip	Max_TS	Min_TS
1	16:39:19	16:39:19	#N/A	0.00	1	2		STOP	1	0	BOSNA SANAYI			
2	16:40:09	16:40:09	#N/A	267.35	2	3		STOP	1	0	BOSNA SANAYI			
3	16:41:15	16:41:15	#N/A	624.19	3	4		STOP	1	0	BOSNA SANAYI			
4	16:41:16	#N/A	00:01:57	628.53	3	4	3			0	BOSNA SANAYI	443_211_1	18:49:47	16:41:16
5	16:41:19	#N/A	00:02:00	628.53	3	4	3			0	BOSNA SANAYI	443_211_1	18:49:47	16:41:16
6	16:41:35	#N/A	00:02:16	686.80	3	4	3			0	BOSNA SANAYI	443_211_1	18:49:47	16:41:16
7	16:43:06	16:43:06	#N/A	1201.25	4	5		STOP	1	0	BOSNA SANAYI			
8	16:44:11	16:44:11	#N/A	1568.53	5	6		STOP	1	0	BOSNA SANAYI			
9	16:44:12	#N/A	00:04:53	1572.40	5	6	5			0	BOSNA SANAYI	443_211_1	18:49:47	16:41:16
10	16:44:13	#N/A	00:04:54	1572.40	5	6	5			0	BOSNA SANAYI	443_211_1	18:49:47	16:41:16
11	16:44:48	16:44:48	#N/A	1953.91	6	7		STOP	1	0	BOSNA SANAYI			
12	16:45:18	16:45:18	#N/A	2274.87	7	8		STOP	1	0	BOSNA SANAYI			
13	16:46:02	16:46:02	#N/A	2749.55	8	9		STOP	1	0	BOSNA SANAYI			
14	16:46:02	#N/A	00:06:43	2751.00	8	9	8			0	BOSNA SANAYI	443_211_1	18:49:47	16:41:16

Figure 3.12. Resulting ATE tables from smart card (a) and bus trajectory (b) datasets.

As a result, using these resulting time-space tables, the performance measures are visualized via different graphical representations as:

➢ Time-Space diagrams (in Excel) (Step IV.1)

➢ ISTT diagrams (in Excel) (Step IV.2)

➢ Reliability estimations (Step V.1)

3.4.2   Numerical Evaluation Using Reliability Indicators

The numeric evaluation was made using the TT reliability indicators (Step IV.3) calculated by total travel times of daily and weekly bus trips. All ISTTs and station kilometers of bus stops are exported to an Excel spreadsheet to calculate the reliability indicators. Initially, average travel speeds are estimated for each segment of each trip. The trips with an average speed higher then 100km/h were excluded from the dataset of TT reliability. These trips, including unexpected average speeds due to the insufficient data or data distribution along the route, need to be excluded. A temporal analysis has been conducted to evaluate the variations of trip travel times (TT) according to their departure times. Furthermore, four indicators are selected for the evaluation of TT reliability among the most advised ones mentioned in chapter 2;

- Buffer Index (mean-based)
- Buffer Index (median-based)
- Planning Index
- Normalized Standard Deviation

First, some conventional statistics calculated to be used for calculation of the above indexes;

- Mean Travel Time
- Median Travel Time
- 95th Percentile Travel Time
- Standard Deviation of Travel Time
- Free − Flow Travel Time

Then four indexes were calculated for weekly, weekdays, and weekends. Resulting index values were used to calculate the correlations of these indexes to each other. After the definition of the indicator having the strongest correlation, a final evaluation has been made over these indexes.

3.4.3   Calculation of the Travel Time (TT) Reliability Indicators

TT reliability analysis has done based on three reliability indicators as recommended by the Federal Highway Administration (FHWA);

- Buffer Index
- Normalized Standard Deviation
- Planning Index

The formulas of these indexes are as follows;

Buffer Index = (95th Percentile TT − Mean TT) / Mean TT

Normalized SD = Standard Deviation of TT / Free_Flow TT

Planning Index = 95th Percentile TT / Free_Flow TT

Buffer Index is the most recommended one between ten indicators evaluated by the four different sources (guide, report, study) mentioned in the background chapter. The Buffer Index is generally calculated based on the mean travel time. On the other hand, the Median based Buffer Index, rather than the mean (average) based, is recommended in the study of (Pu, 2011) to avoid underestimating unreliability. Hence, the Mean Based Buffer Index, Median Based Buffer Index, Normalized Standard Deviation, and Planning Index has calculated to evaluate TT reliability.

Before calculating these indicators, some conventional statistical calculations have been made, which will be the basis for the calculation of the above-mentioned indicators.

- Standard Deviation of Travel Times,
- Mean Travel Times,
- 95th Percentile Travel Times and
- Free-Flow Travel Times

have been calculated in Microsoft Excel spreadsheet software. ISTTs that are generated from ATEs for bus stops have been used in these statistical calculations.

Total travel times of each trip have been inferred from the arrival time estimations. Mean travel times of each day or week are the average of these travel times that belonging to the trips.

Free-Flow Travel Times have been calculated by taking the maximum speed limit of 50km/h+10% (55km/h) for all inter-stop segments.

The standard deviation has not been calculated for the entire population. The standard deviation was calculated for a sample of the dataset since some of the trips were excluded.

3.5 Strength and Weakness of the Proposed Method

Probably the only weak part of the proposed methodology was the arrival times estimation by using insufficient smart card datasets. The estimation of arrival times to the bus stops need to be conducted on the trips with a dense number of passengers. On the other hand, smart card readings need to be well distributed all over the route. It is very hard to estimate arrival times for the trips that lack these conditions. As long as the interpolation or extrapolation distance is getting longer, the estimation health is getting lower. Hence the trips estimated under such conditions were mostly excluded from the time-space/inter-stop diagrams and from the sample dataset before calculating the reliability indicators. Nevertheless, especially stop based evaluations over inter-stop travel time diagrams can be misleading, for some ATEs calculated by using the smart card records in excessive distances.

The strength of the proposed method comes from the preprocessing phase of the datasets and LR method specifically produced for this type of poor datasets without a GTFS based schedule and GIS-based network dataset. All SCD successfully matched to the related route segment without using a map-matching methodology, which is hard to implement precisely. Thanks to the developed geoprocessing tools under model builder. This is quite important since the provided dataset is comprised of only the departure locations of passengers, so a single point of SCD is critical for

the estimation of arrival times. In the following chapter, 2 case study PT lines will be processed by using the methods explained here.

CHAPTER 4

CASE STUDIES FOR PUBLIC TRANSIT RELIABILITY IN KONYA

Konya is the largest city in Turkey in terms of its surface area and the seventh most populous one. Konya, which is composed of 31 districts, has a population of 2.205.609, according to TUIK's 2018 statistics. Konya Municipality was established in 1830 and Konya has the status of "Metropolitan City" in 1987. Since 1989, municipal services have been carried out according to this status. Konya is one of the most developed cities in Turkey in terms of the economy. Konya is also important with its natural and historical richness. The city was the capital of the Anatolian Seljuks and was the administrative capital of Karaman state during the rule of the Ottoman Empire. The Rumi's (Mevlana) tomb, a well-known Sufi, is in Konya. On the other hand, Konya is one of the most important industrial cities in Turkey. The historical old city is located in the Aladdin boulevard in the middle of Konya, and the history of Alaaddin Hill, located at the center of the city and the public transportation network, dates back to 3000 BC.

4.1   Bus Public Transit System in Konya

The city has continued to develop around the historic center over the years and reached the land-use state, as seen in Figure 4.2a in the year of 1984. In 1999, the land-use situation changed in accordance with the development of the city, as shown in Figure 4.2b (Yenice, 2012). The most recent map of the Konya taken from Google maps which represent today's Konya, can be seen in Figure 4.1. The city's primary development direction is running through the Selçuk University Campus, located in the city's uppermost north side. All new shopping malls, parks, and residential areas are located along this north ax as the city developing.
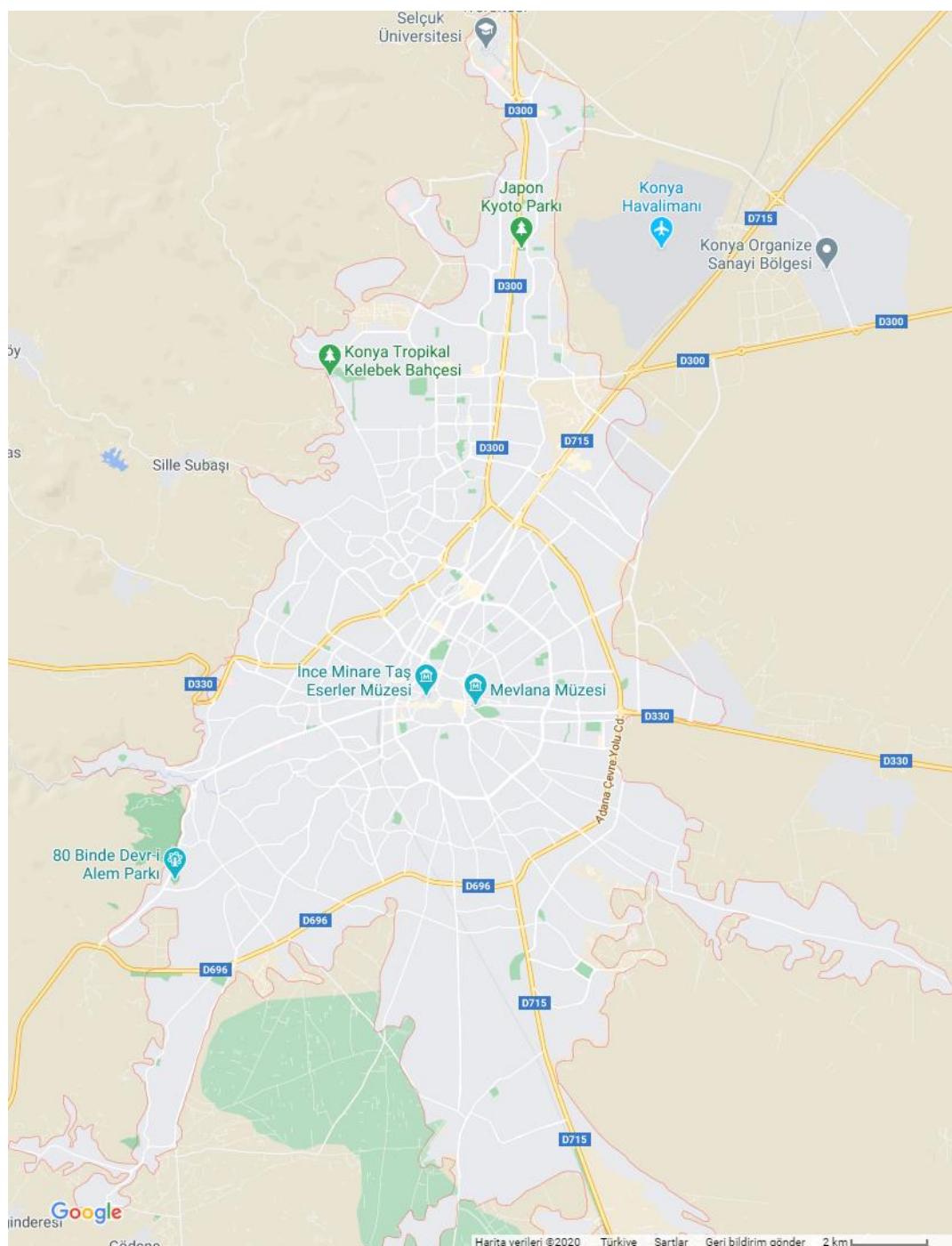
Figure 4.1. Current Google Map of Konya
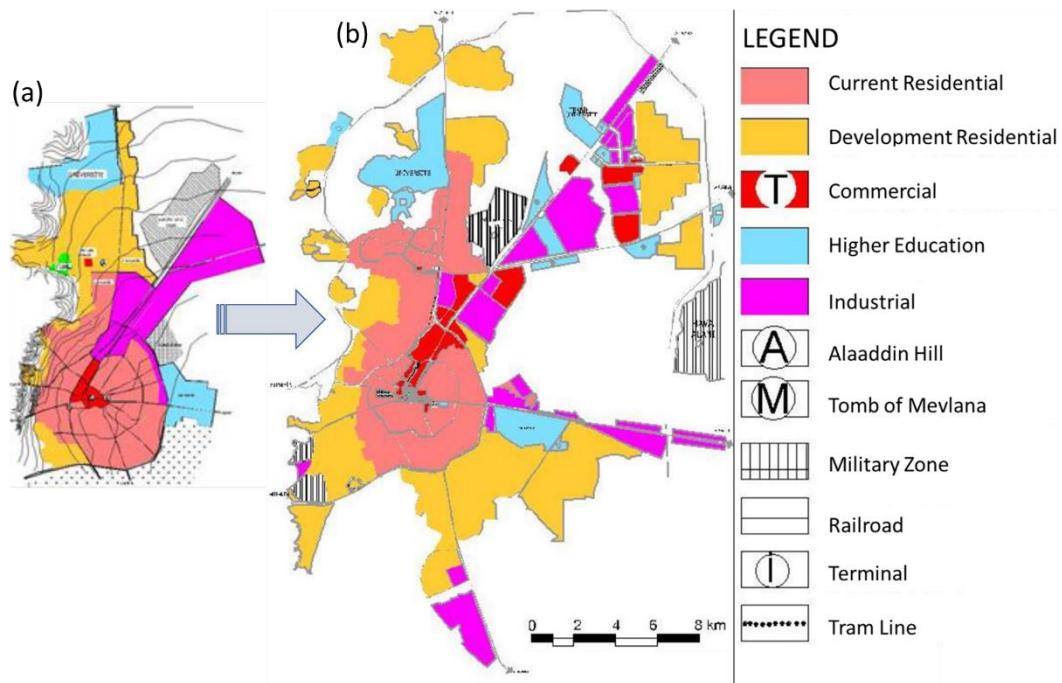
Figure 4.2. (a) 1984 and (b) 199 Land Use Schema of Konya City (Yenice, 2012).

In Konya, public transportation needs are met by tram, bus and minibus within the scope of Konya Metropolitan Municipality services. During the university education period, the north tram axis and southwest bus lines are used extensively. During the season, the number of daily passengers for the tram has been determined as 90,000 people. Efforts to provide detailed daily passenger numbers on the basis of stops continue with the smart card (Elkart) system.

In terms of PT infrastructure, tram use and minibus use are dominant for the north axis in Konya, and there are bus lines moving in the north. Bus and minibus mobility is also concentrated in the city center, and city center traffic is severely affected by this mobility in Konya. When the bus lines are examined, it is seen that they move parallel to some routes with the tram line, which is preferred to be used predominantly in the north axis. The most active axis of the city is the north axis in terms of city planning. It includes the areas where new high-density housing and

secondary center have been formed here. The dominant locations for the north axis are Konya Bus Station, Selçuklu Municipality, Small and Medium-Sized Industrial Zones, Selçuk University Alaeddin Keykubat Campus, Beyhekim State Hospital. In the north axis, where residential areas are also densely located, the average floor height is four floors and above, except for some specific areas. (Uyan et al., 2017)

Figure 4.3. Public Transit Network in Konya City

58

4.2   Public Transit Data for Konya

8.182 million smart card readings were recorded in October 2018 for all days. 547,624 of these records are from distinct (single), smart card IDs. Daily average passenger count was 272,755 for October 2018 under the light of these numbers.

The vector data of bus lines the stops used for this study were given by the Konya Municipality Public Transit Management Department. The department is providing some information of the bus lines and bus stop over a web page so-called ATUS (Smart Public Transit System). ATUS is providing bus departure times for the lines, but there is not a bus-stop based schedule information on their web page. The provided ATUS data consists of line and point geometry objects. There is not a GIS-based Network Dataset or a LRS containing the routable bus lines. On the other hand, a GTFS based bus-stop or schedule information has not been created for the management of this public transit network.
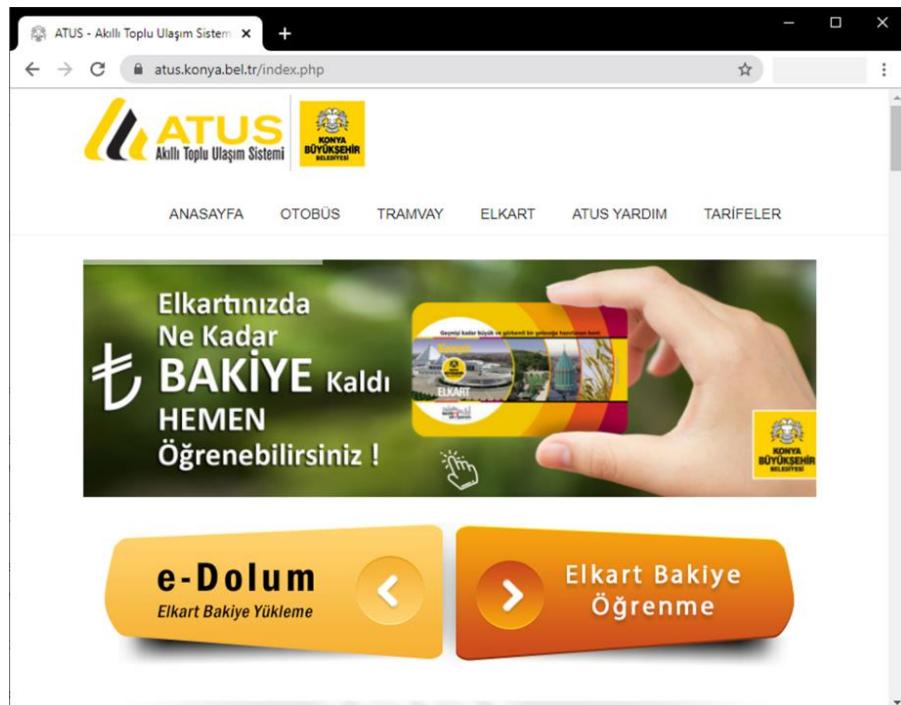

Figure 4.4. ATUS Web Page

4.3   Study PT Lines

There were two PT lines studied in detail for the depiction of the concepts and performance measures proposed in this study:

  i.   ULID 443 Samanpazarı Line - Bosna Sanayi
  ii.   ULID 50 Samanpazarı Line - Eski Kunduracılar

ULID 443 represents "Line 44 and Subline c (3)" and was selected for the initial application. Line 44-c starts from the Samanpazarı, east of the city center and reaches the city center Alaaddin Tepesi. After passing through Alaaddin Tepesi, the route proceeds to the Bosna Sanayi then turns back to the city center and Samanpazarı. This route is 62,3 km length and running from the city center through the almost all north axis of the city. ULID 443 has 130 bus-stops in two directions. (Figure 4.6). The inbound direction name is Bosna Sanayi and the outbound direction name is Samanpazarı. This PT line covers the north axis of the city, starting from the city center. Since the north axis is the most active axis with respect to the others, this line has an important place in terms of jobs-housing relationship and commuting passengers. ULID 50 (Line 5 with no subline) was selected for the second study PT line (see Figure 4.7). The assigned unique line ID The total length of the route is 15.6 km. The departure point of the route is east of the city center. After departure, buses head through the west direction where the city center takes place. Eski Kunduracılar – Samanpazarı Line line has 43 bus-stops in two directions. The bus line with ULID 50 starts from the Samanpazarı, east of the city center, and runs through the Eski Kunduracılar and turns back to the Samanpazarı. This line is outside of the city center and has less commuting passenger concerning the ULID 443. Most of the trips and commuters are concentrated in the morning hours. After mapping the bus trajectory and SCD to the bus route, by using the tools generated to conduct LR tasks and estimating the arrival times to the bus-stops, with the proposed methodology, necessary attribute tables have been generated to create time-space diagrams. October 30, 2018 Tuesday, was selected as the study day, as it is a midweek day with regular travel patterns.
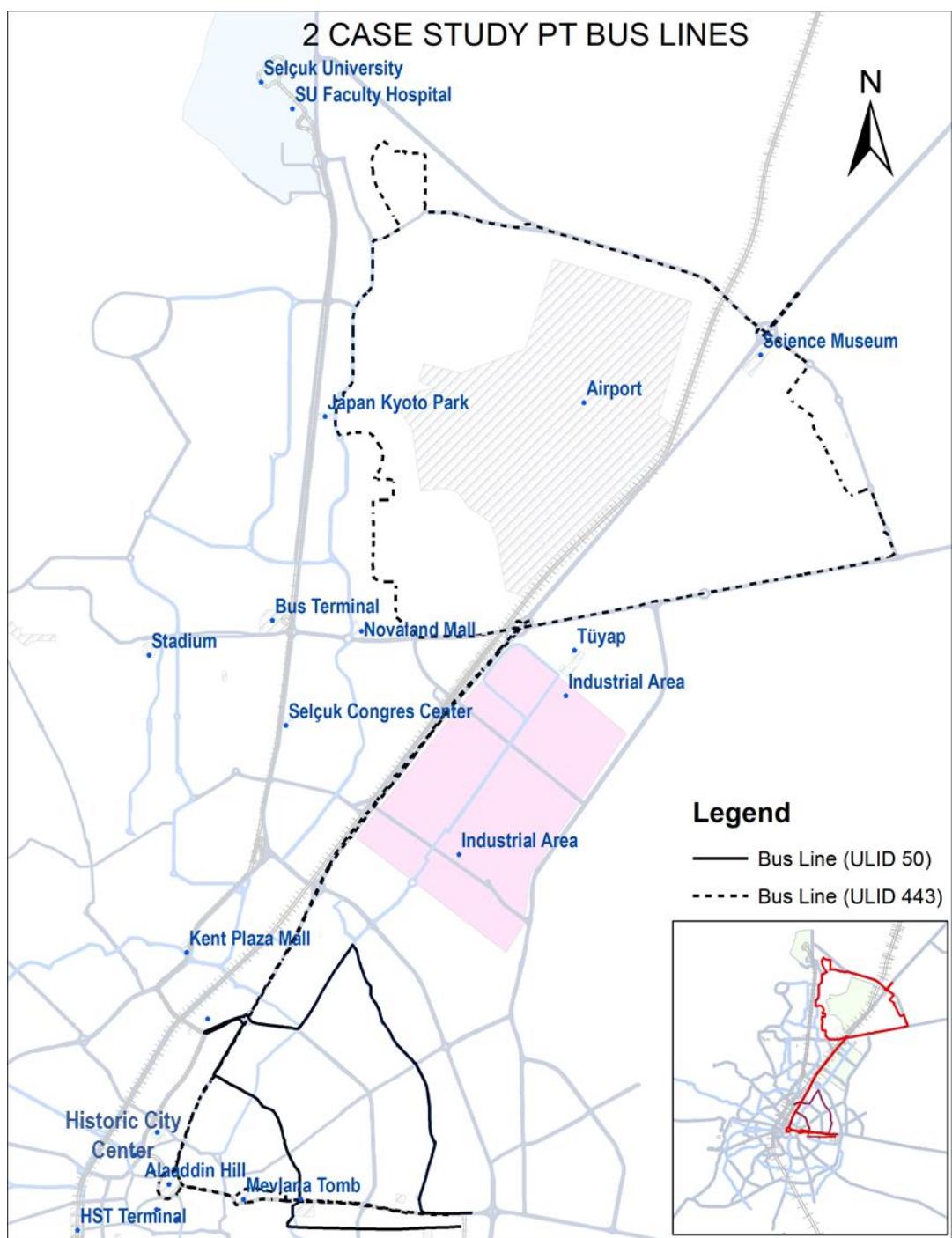
Figure 4.5. 2 Case Study PT Bus Lines
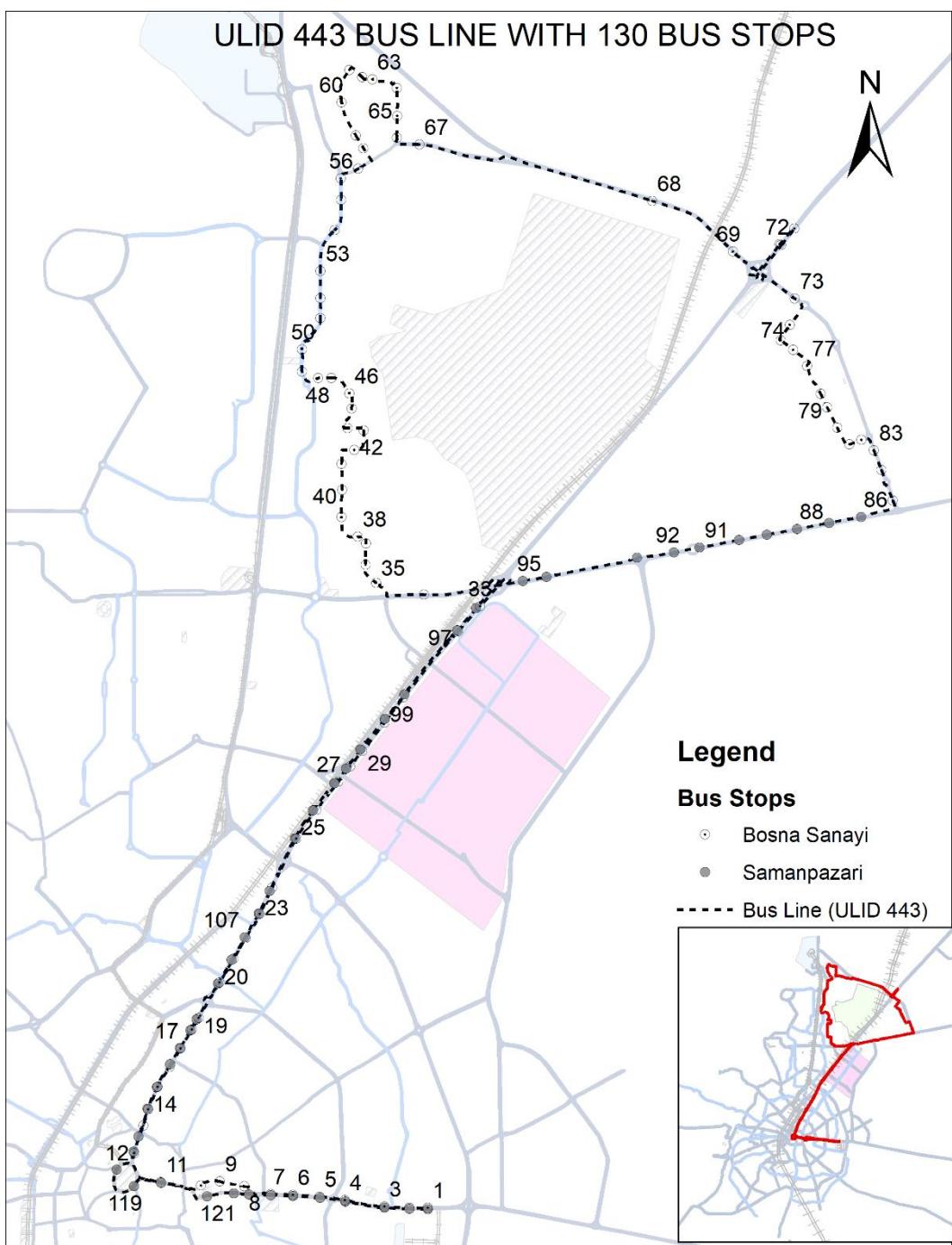
Figure 4.6. Study PT Line ULID 443
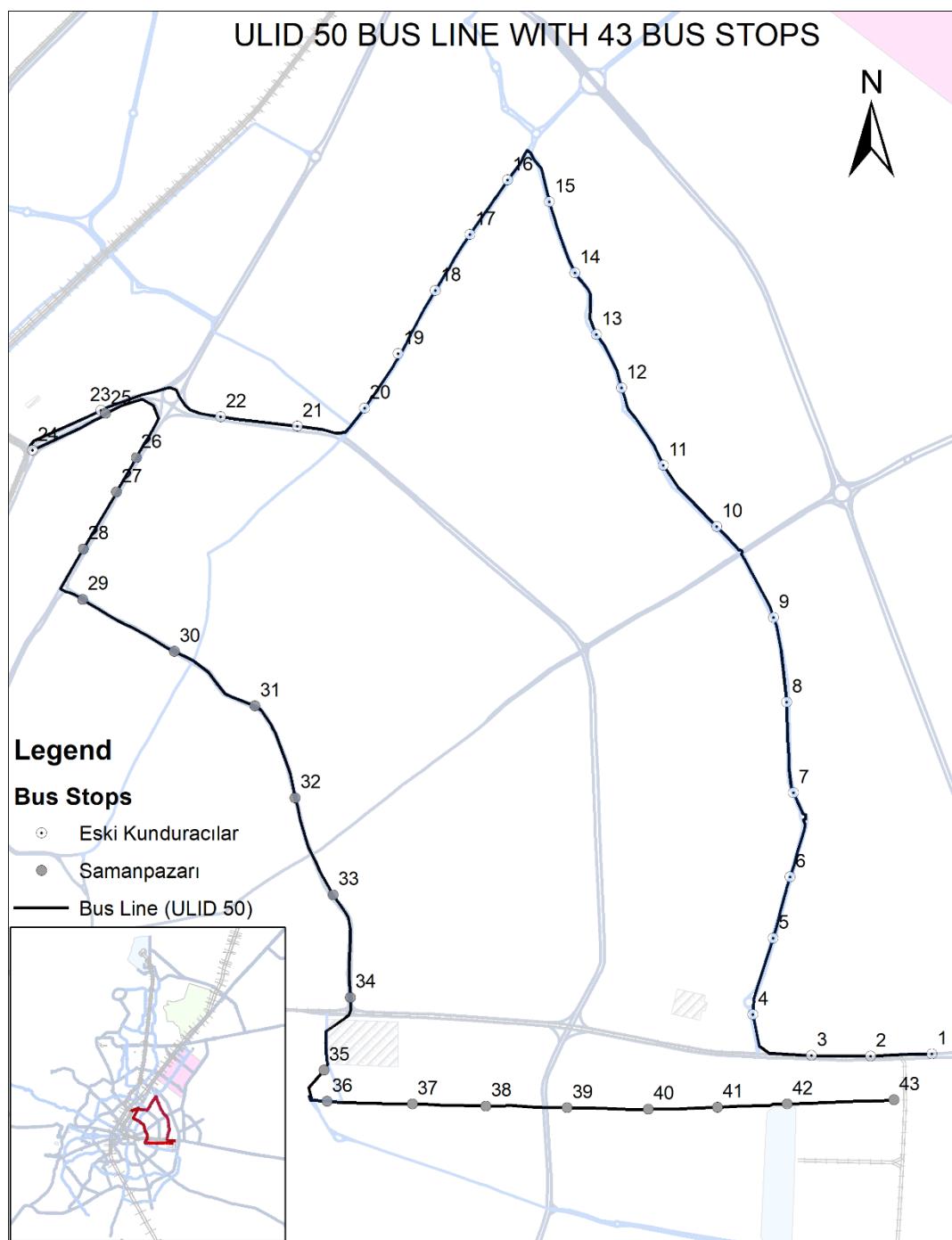
Figure 4.7. Study PT Line ULID 50

BTD of the same route ULID443 on October 30 is processed the same way as the smart card dataset. While there were 17 trips in the smart card dataset, there were only six trips in the bus trajectory dataset. Two of them have missing trips and direction field values. This situation consists of missing data storage in the database, storing BTD. Since GPS datasets are occupying an enormous amount of data in the database, some of them are deleted somehow in time. Those who manage PT systems do not store historical trajectory data in their archives for this kind of reason. Another critical reason is that these database management systems may not primarily be designed to store trajectory data for further analysis. Eventually, only 4 of 6 trips fetched from the dataset of October 30 were suitable to be processed.

4.4   Time-Space Diagrams

Typical time-space diagram of a trip for the route ULID 443 and ULID 50 are presented in Figure 4.8a and Figure 4.8b. A typical bus trajectory belonging to a single trip on the time-space diagram can be seen in Figure 4.9. This diagram is reflecting ground truth better than the one in the SCD based time-space diagram. Especially for bus stops where there are no smart card readings, arrival time estimations are not as precise as the ones estimated by bus trajectory datasets. So the departure and arrival times for the whole trip or inter-stops are estimated more precisely by using BTD datasets. On the other hand, since there are a limited number of trips due to the data storage and management problems mentioned before, the analysis conducted based on a bus trajectory dataset will be limited with the number and departure times of the trips takes place in the dataset. An analysis based on seventeen SCD trips can be more reliable than the four trips of BTD.
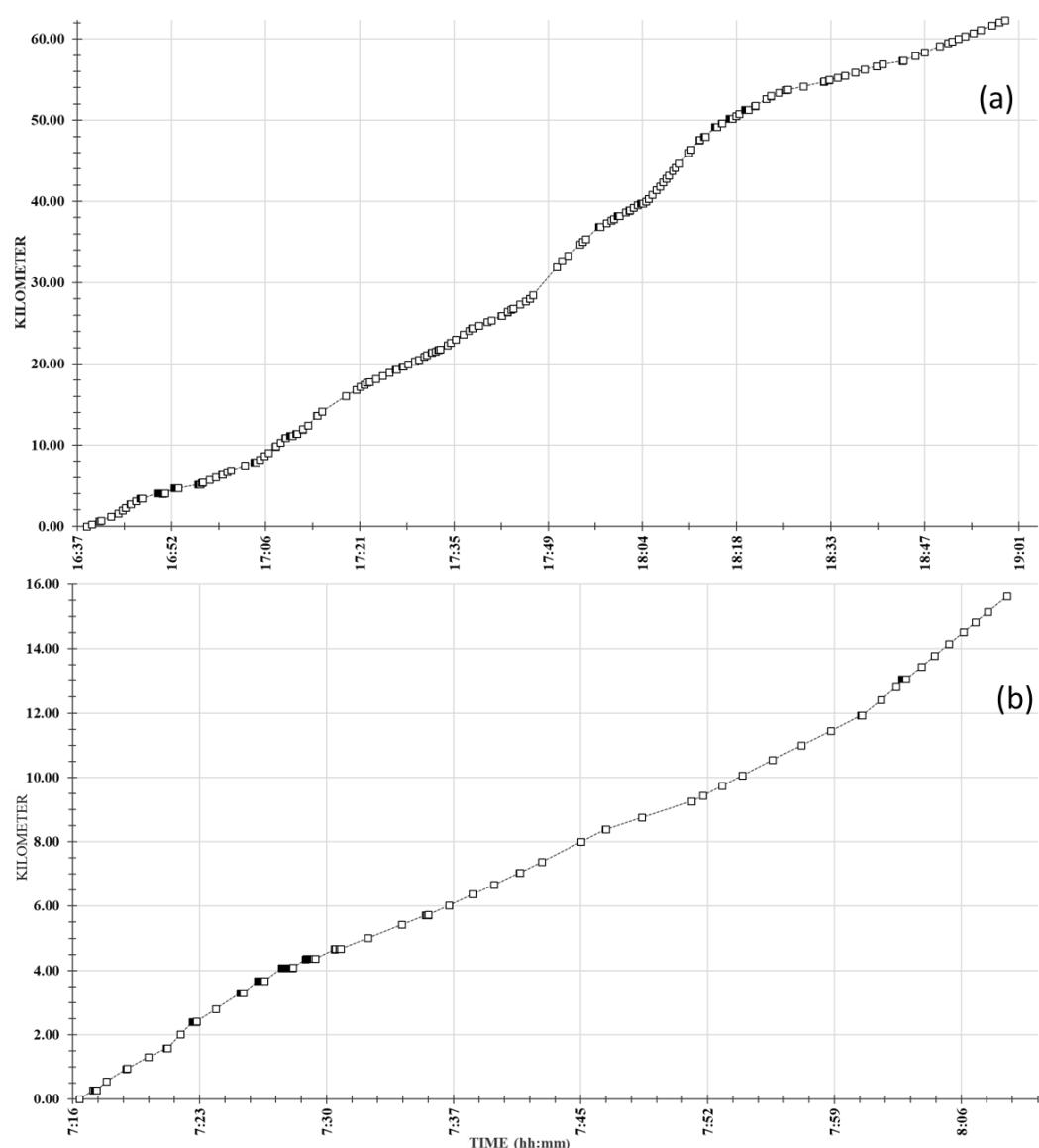
Figure 4.8. A sample time-space diagram from SCD for (a) ULID 443 and (b) ULID 50 on the study day
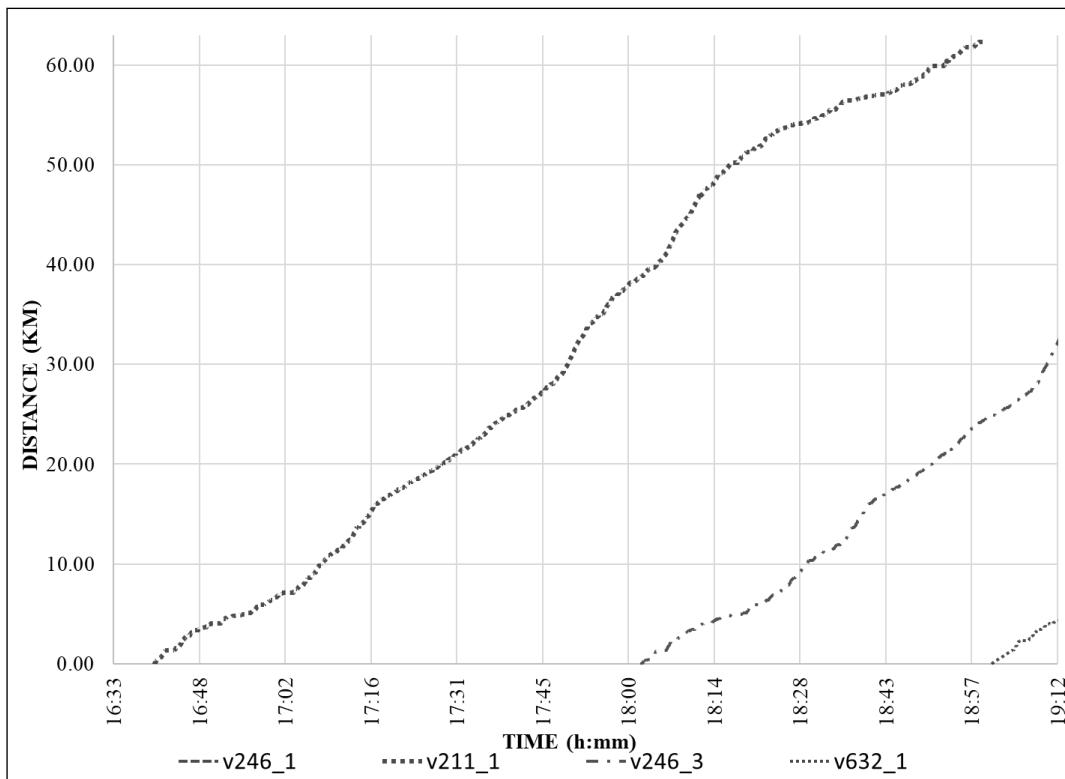
Figure 4.9 A sample-time-space from BTD for the study day of ULID 443

4.5   ATE for Study PT Line Bus Stops

Time-Space diagram of each trip in a day for both study PT lines have produced as a scatter plot diagram, as seen in Figure 4.10. The trip v608_1 of ULID 443 with the departure time 07:20 in Figure 4.10a has a linear pattern between the kilometers 12-38 and 38-50. Since there is only one smart card reading between kilometers 12 and 50 (bus stop orders 30 and 102), the diagram follows a linear pattern in time until the occurring of SCD. The only smart card reading is the one on the kilometer 38 (bus stop 77) in between kilometers 12 and 50. Those markers are only bus stops connected with lines between kilometers 12 - 38 and 38 - 50 in the diagram.

66
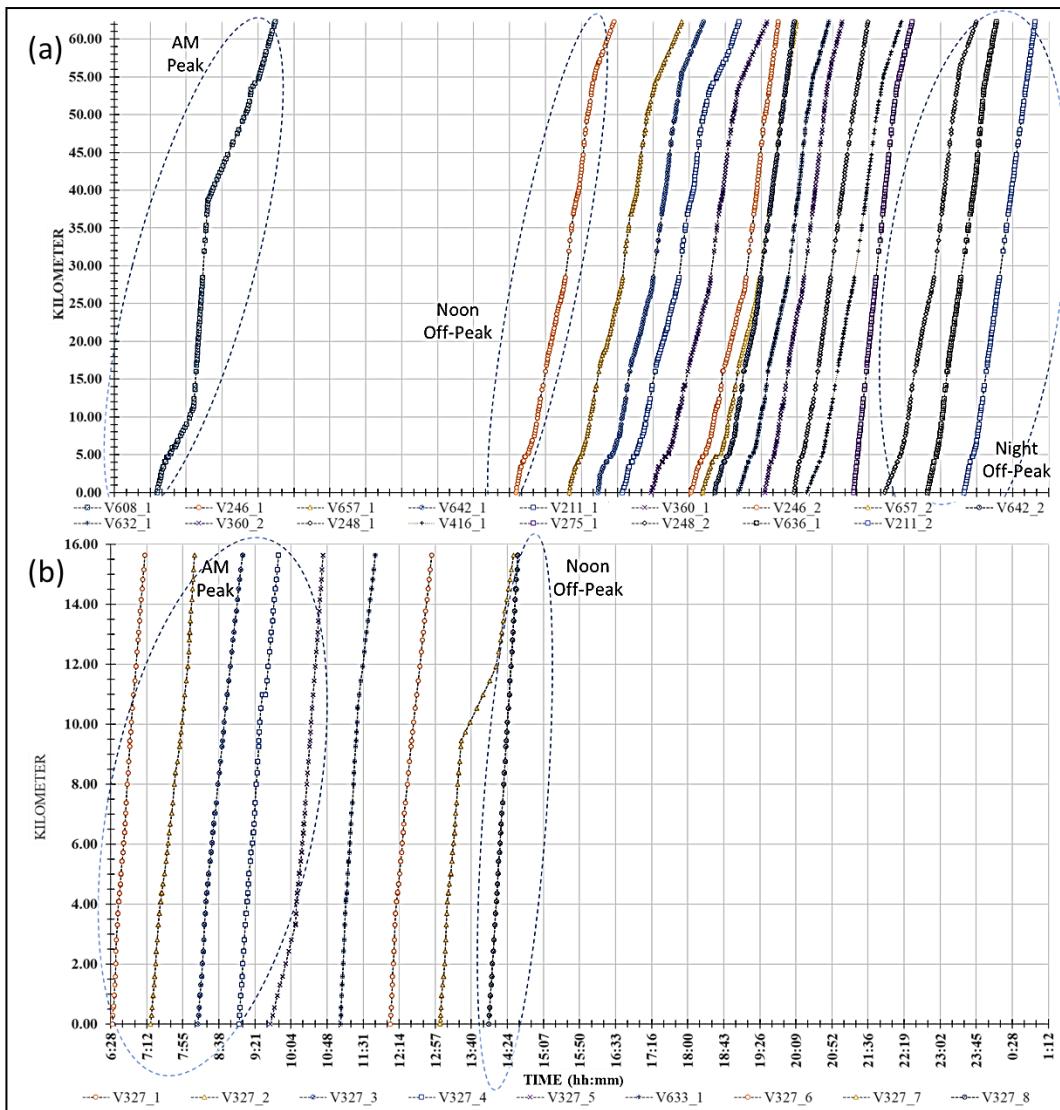
Figure 4.10 SCD Based Daily Time-Space Profiles of (a) ULID 443 and (b) ULID 50 PT Lines on Oct. 30th.

There are 17 trips that take place in the study PT line with ULID 443. The trip v608_1 is a morning trip with a departure time at 07:20. The other 16 trips belonging to 10 other vehicles take place in PM time periods. The first noticeable pattern is a bunch on trips v657_2 and v642_2. Since there are 15 minutes between departure times, it may be a possible condition for an evening peak hour period. The PT line with ULID 50 consists of 9 trips and there is not any trip after 14:00. Only morning

and noon trips can be evaluated in Figure 4.10b. The trip v327_7 with a departure time 14:00 has a considerable slowdown between the 9th and 12th kilometer of the route.

All timestamps of smart card readings have been normalized (reduced) to zero for better viewing of the scattering belonging to the trips in the same day. The variability of travel time between trips for total travel time and stop based travel times can be seen in this time-space diagram (Figure 4.11 and Figure 4.12). The first time-space diagram in Figure 4.11a and Figure 4.12a has been generated based on raw SCD before arrival times estimation of bus stops. The second time-space diagram in Figure 4.11b and Figure 4.12b has been generated based on both ATE of bus-stops and the smart card datasets. Datasets belonging to the trips like v275_1 in Figure 4.11a and v327_5 in Figure 4.12a can be excluded due to the lack of enough and well-distributed SCD.

The same kind of time-space diagram generated based on BTD, after timestamps normalized (reduced) to zero in Figure 4.10. Again, due to the limited number of trips are being stored in the bus GPS dataset, the variability in time and space for the trips could not be evaluated sufficiently, depending on the different time periods of a day. Since the direction table or the data itself was not stored in the dataset of the bus trajectory, morning and noon trips are not available. Although BTD is the best option to estimate bus arrival times of the bus stops precisely by using the station kilometers in the bus stop dataset, the boarding location of the passengers may vary from the stop location in the bus stop dataset. Even if a dwell location in bus trajectory identified other than the one in the bus-stop dataset, without comparing the SCD, it is hard to be determined whether this dwell caused by boarding/alighting or due to traffic congestion.
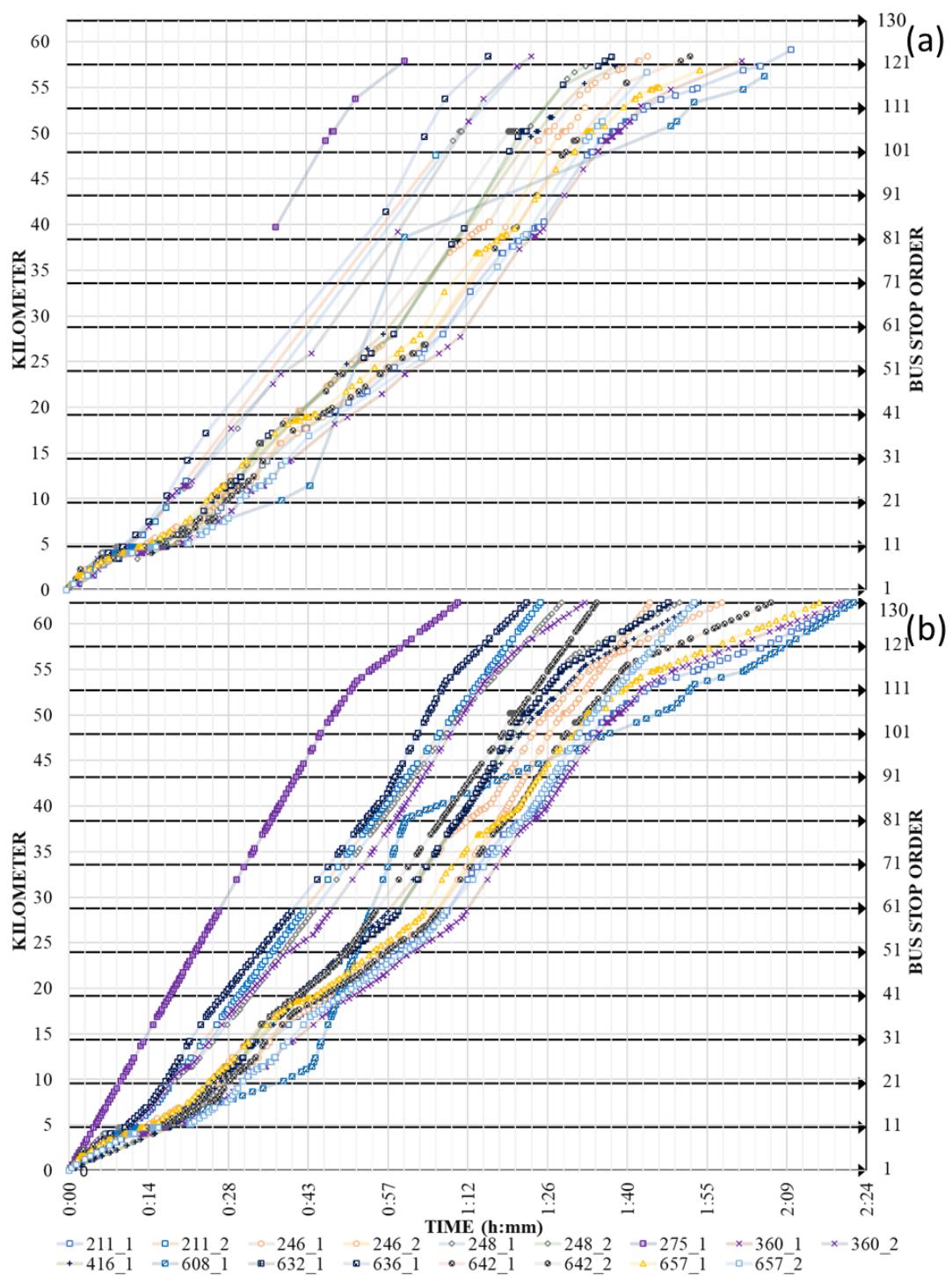
Figure 4.11 Time-Space Diagrams for ULID 443 (a) only SCD, and (b) SCD with ATE Interpolations and Extrapolations
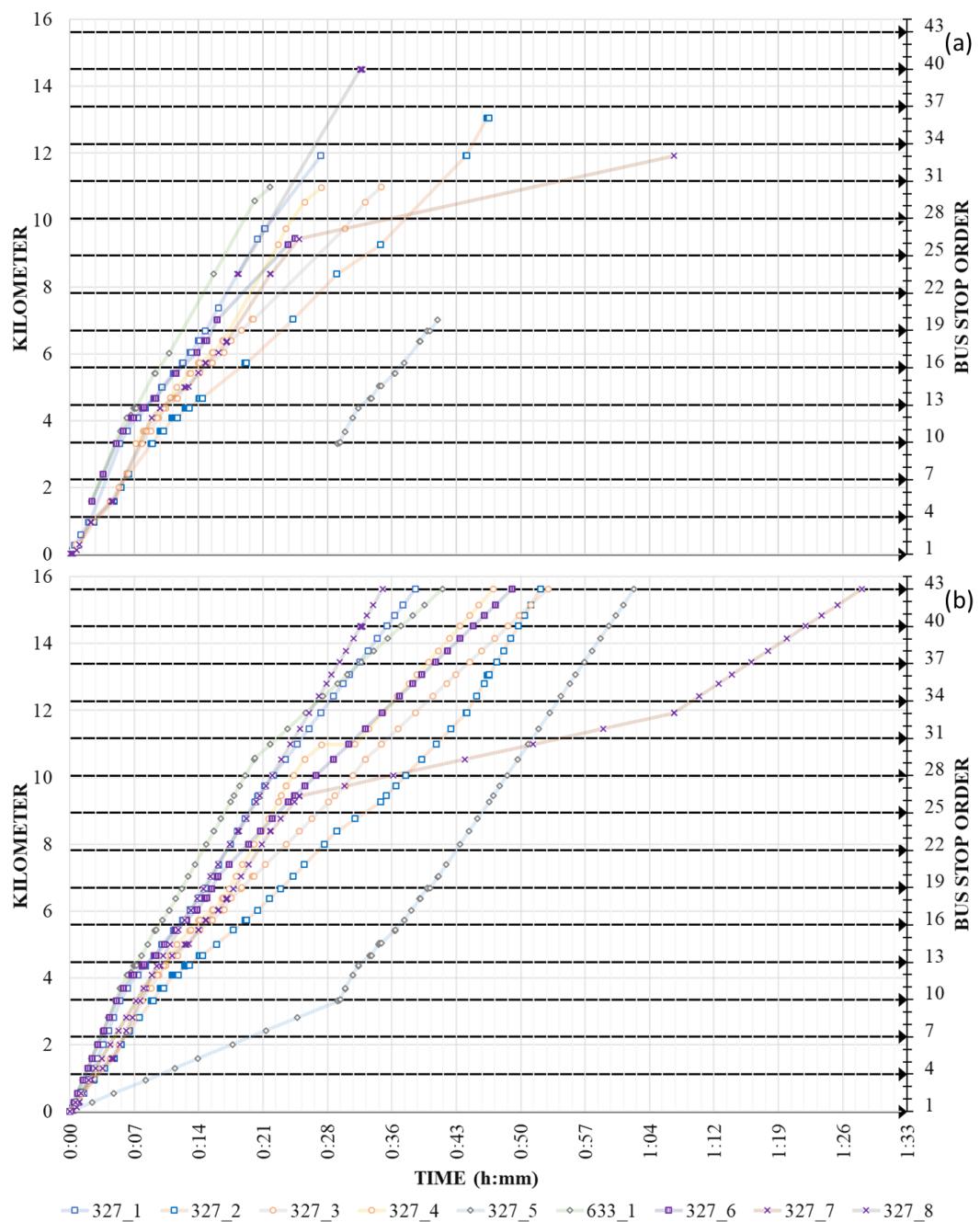
Figure 4.12. Time-Space Diagrams for ULID 50 (a) only SCD, and (b) SCD with ATE Interpolations and Extrapolations
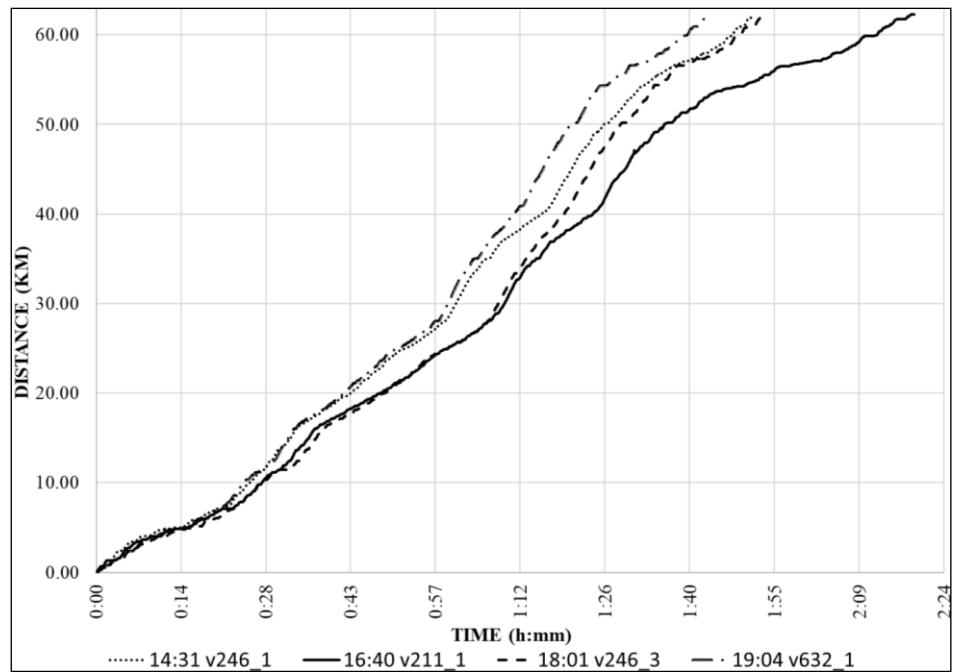
Figure 4.13 Time-Space Graph (ULID 443/ Oct.30) from BTD

As long as there is enough and well-distributed SCD in a trip, smart card-based ATE of bus stops can be more reliable with a large number of trips in a daily dataset. SCD based ATE of bus stops 99, 101, 102, 103, 106 are almost identical to the bus trajectory based ATEs in Figure 4.14a. Bus-stops with orders 98, 99, 100, 101 and 102 are zoomed-in and further examined to see the ATE differences between BTD and SCD. As can be seen in Figure 4.14b, differences are under 3 seconds except bus stop order 98 with a difference of 23 seconds.

71
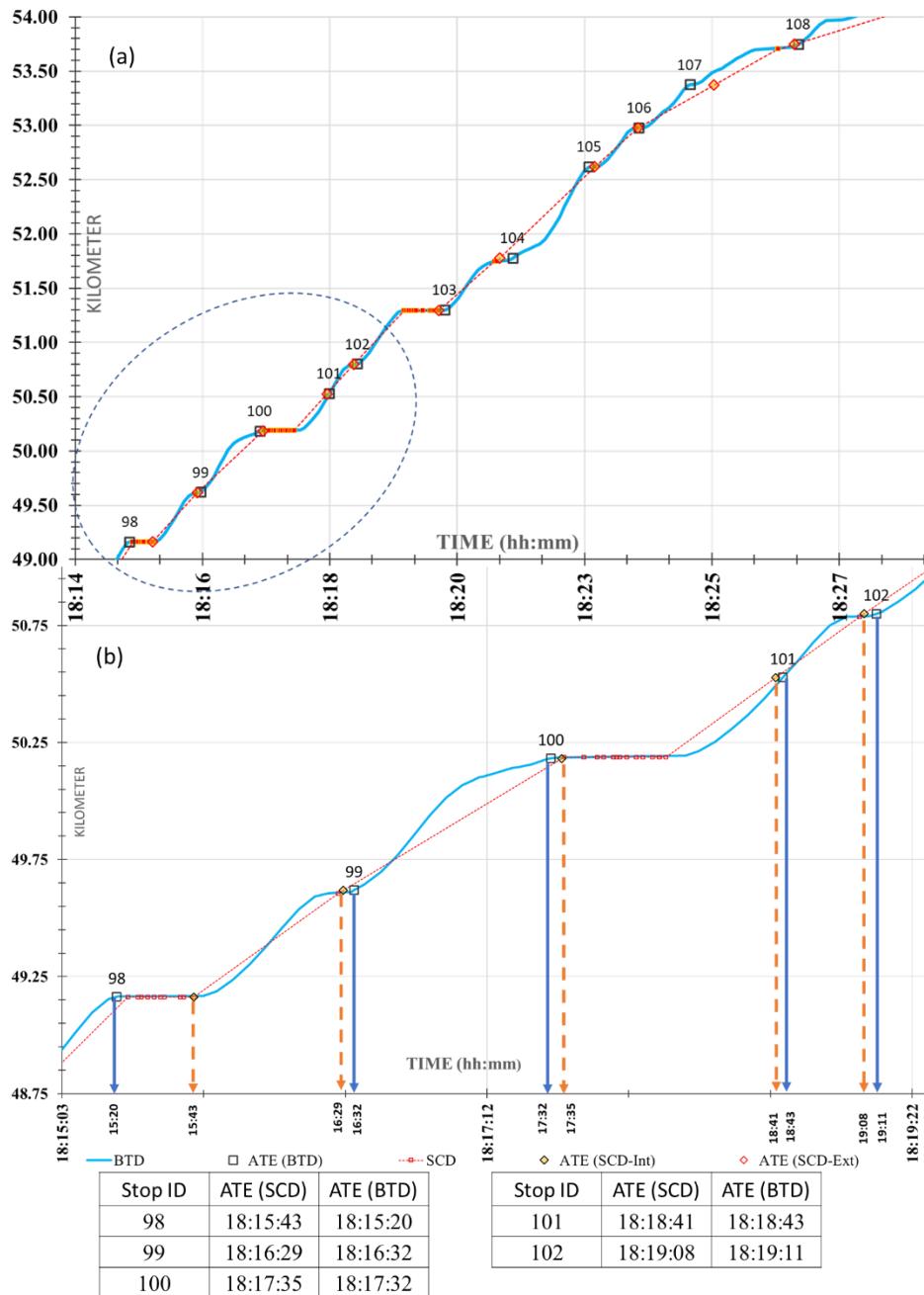
Figure 4.14. Smart card Locations and ATE

4.6   Inter-Stop Travel Time (ISTT) Diagrams

Inter-stop travel time diagrams for the bus line with ULID 443 on October 30 have been generated for the stop based evaluation of the reliability. To generate a travel

72

time diagram between bus-stops, kilometers (*MEAS)* of the bus-stops and the elapsed time between bus-stops (DT) are calculated. Inter-stop time difference (DT) columns have been generated by simply subtracting the estimated bus-stop arrival time values from previous ones. Inter-stop travel times for a single trip have shown in Figure 4.15. Each inter-stop travel time can be evaluated separately by taking the inter-stop distances into consideration in this diagram. All inter-stop travel times averaged in a diagram in Figure 4.15a.

4.6.1   Average Travel Speed Estimation

The average speed for each segment belonging to each trip of a day has been calculated by using inter-stop travel time differences and estimated arrival time differences. The trips with an average speed higher than 100km/h can be excluded from the TT reliability evaluation. When these trips are examined, it can be seen that cause of the unexpected speed values are insufficient smart card records and the insufficient distribution of these records along the route.

The scatter-plot diagrams of the average speeds for all trips on the bus route with ULID 443 has been presented in Figure 4.16. Four trips can be clearly identified as having an average speed above 100 km/h. On the other hand, it can be seen in Figure 4.17 that there is not any trip with an average speed higher than 100km/h. The maximum speed of a trip is about 60 km/h. Although the density (frequency) of smart card readings are considerably low in this route with ULID 50, there are not any speed values above expected. If we examine the dataset, the distribution of smart card readings along the route is better than the route with ULID 443. On the other hand, this route is a considerably shorter route than the route with ULID 443.

As a result, 4 trips with an average speed above 100km/h are excluded from the dataset of the study PT line with ULID 443. No trips are excluded due the average above 100 km/h from the study PT line with ULID 50.
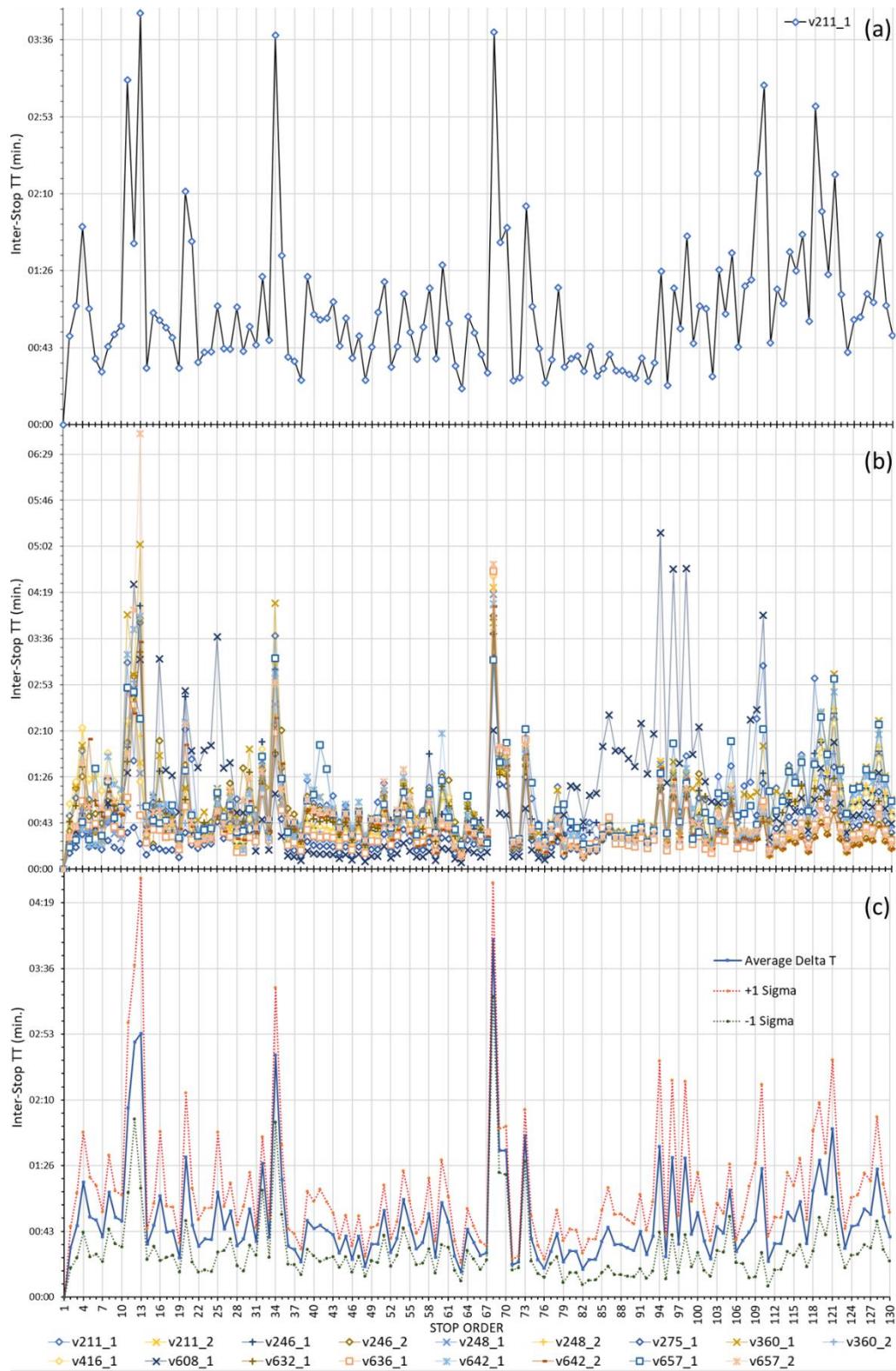
Figure 4.15 Inter-Stop Travel Time Diagrams of a single trip (a), daily trips (b) and the average of daily trips (c) for ULID 443/ Oct.30
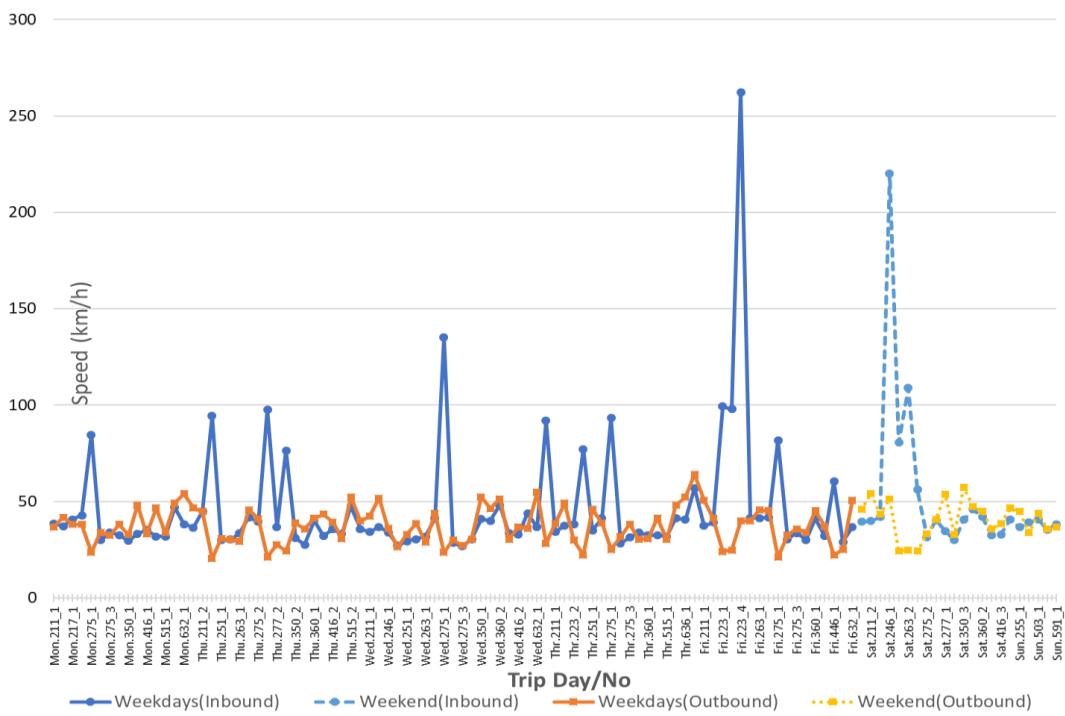
Figure 4.16. Average speeds per trip for ULID 443 Oct. 22-28
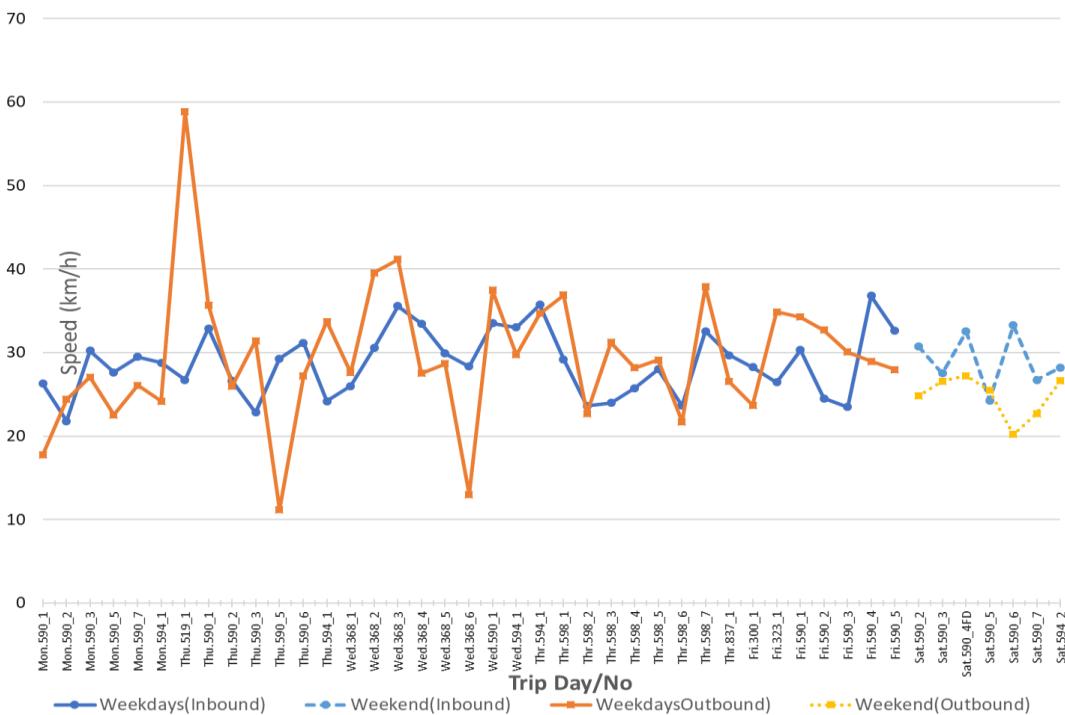
Figure 4.17. Average speeds per trip for ULID 50 Oct. 22-27

75

4.6.2   Inter-Stop Travel Times (ISTT) Heat Maps

ISTTs for all trips on October 30 have been represented as a heatmap for the ULID 443 in Figure 4.18 and ULID 50 in Figure 4.19. Trips are in ascending format in the vertical axis from upside to down by their departure time. The darker cells in the figures represent longer travel times. The first trip with the departure time 07:20 is in AM and the other trips below are in the PM time period (see Figure 4.18).
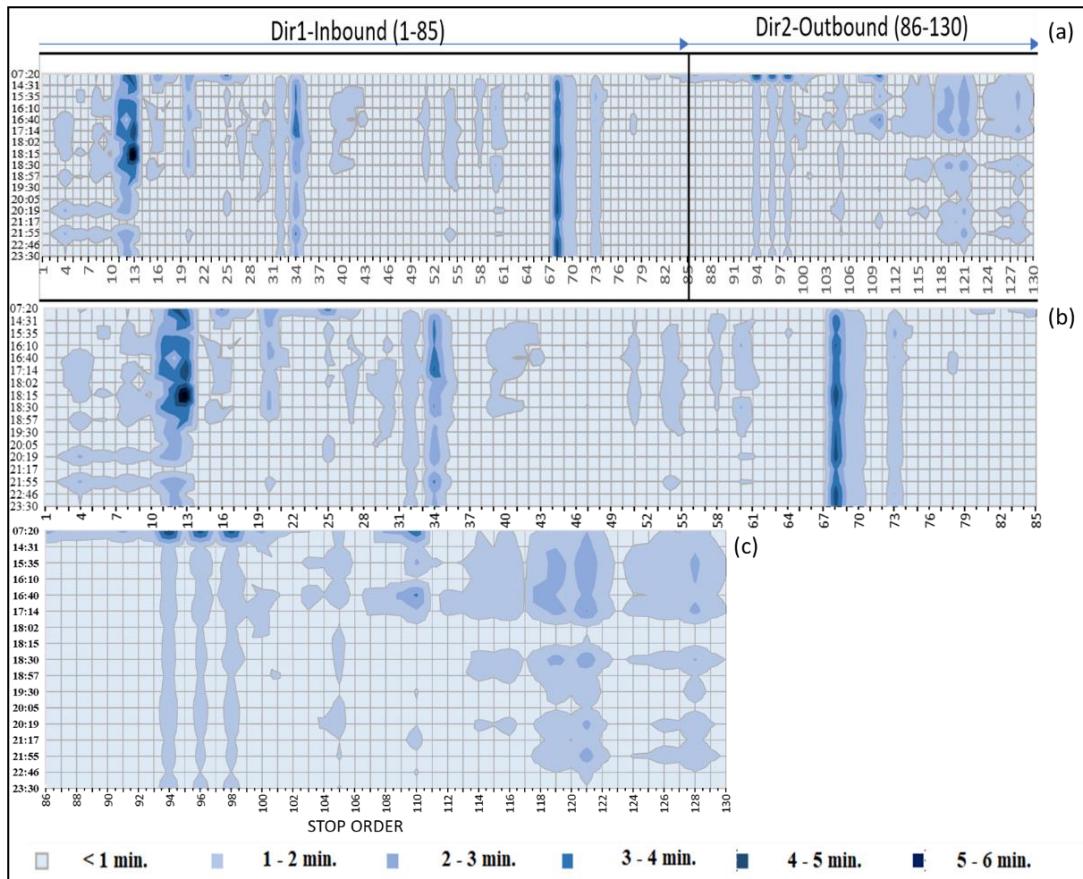
Figure 4.18 Inter-Stop Travel Time Heatmap (ULID 443 Oct30)
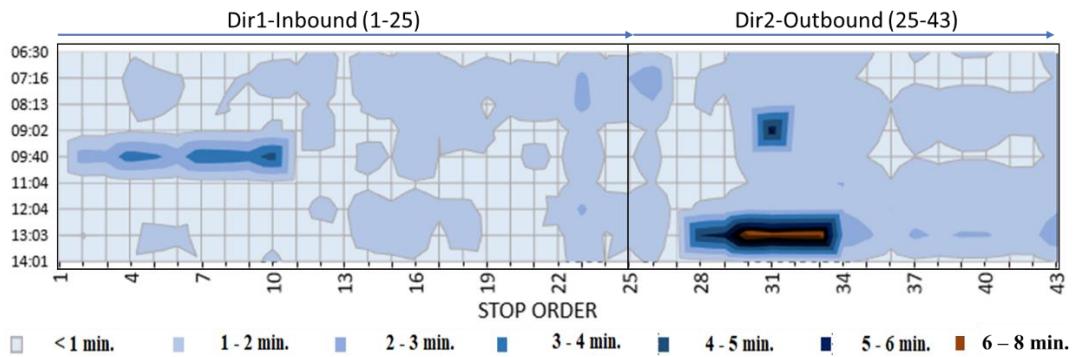
Figure 4.19. Inter-Stop Travel Time Heatmap (ULID 50 Oct30)

The ISTT heatmap in Figure 4.18a has been divided into two separate directions for deeper evaluation of travel times according to their trip departure and arrival times on the vertical axis, as seen in Figure 4.18b and Figure 4.18c. It is obvious that travel times are longer at inbound bus stops (direction 1). Morning and evening peak hours are darker cells, evening and night hours are lighter cells as expected. Travel times between bus stops 67, 68 and 69 are remarkably long in all trips. It is observed in Figure 4.19 that darker cells are concentrated between bus stops 4 and 10 for inbound direction. On the other hand, the darkest cells are in the bus line's outbound direction with ULID 50, indicating that inter-stop travel times are longer between bus stop 28 and 33 on noon off-peak hours.

4.6.3 Findings based on ISTT Evaluations

After generating and evaluating the inter-stop travel time diagrams for daily trips, it has been observed that there could be major differences in terms of ISTT between both trips and bus stops. When the dataset is examined, it is seen that some of these travel time differences may not reflect the ground truth, since the distribution problem of the smart card readings along the line. It has been observed that especially the differences in travel times between bus stops in the trips v608_1 and v657_2 are caused by the lack of proper distribution of smart card readings along the line. Although there are sufficient smart card readings on these lines, it has been observed

that most of them are concentrated at some bus stops and there is no smart card data at some other bus stops. Excluding the trips with an unexpected average speed may not be useful in some trips since average speeds can be within expected values due to the other random reasons. For these reasons, it may not always be possible to make a detailed assessment of the inter-stop travel times. On the other hand, inter-stop heat maps can give a healthy idea about inter-stop travel times in terms of showing the average delays in the segments between stops more clearly.

4.7   Reliability of PT Bus Services

The fourth week of October and the bus line with ULID 443 has been selected for the first weekly study case of daily time-space profile diagrams. All smart card datasets belonging to days between the 22nd and 28th of October have processed, as mentioned in the methodology chapter. The time-space daily profile diagrams belonging to all weekdays have evaluated day by day in terms of public transit reliability of the trips. Monday and Thursday trips seem more reliable if we consider the bunching trips on Tuesday, Wednesday and Friday in Figure 4.20 and Figure 4.21. Saturday and Sunday trips are taking place in Figure 4.22. Saturday trips present a similar pattern considering weekday trips but still seem more reliable than the weekdays like Tuesday, Wednesday and Friday. Sunday trips look like the most reliable ones considering all other days, but the low number of trips also has an effect on this situation. When we consider peak times, Monday trips can not be evaluated for morning peak hours, since there is not any trip occurring on the smart card dataset for this time period. On the other hand, the Monday trips taking place in the evening peak hours and the Thursday trips still look more reliable than the peak hours of the Tuesday, Wednesday and Friday. The trips with the departure time in off-peak time periods mostly look more reliable than the peak hours as expected. It is not that different for the Sunday trips for peak and off-peak hours. The travel times of the trips with the departure times taking place in noon-peak and evening-peak are longer than the ones taking place in off-peak time periods as expected.

The same evaluation has been made over the second weekly study case for the bus line with ULID 50 on the same week of October. There is no trip taking place on Sunday in this bus line, and there are also no trips running after 15:00 AM for any day in this bus line. It is observed in Figure 4.23 that the longest trip in terms of total TT on Monday is taking place in morning peak hours as expected. Other trips on Monday seem reliable. The trips are having the longest total TT taking place in the off-peak noon hours on Tuesday and Wednesday. The only trip having a considerable long total TT is taking place in morning peak hours as expected on Thursday (see Figure 4.24). Friday and Saturday seem like the most reliable days in terms of total travel time considering the trips taking place these days.

As a result, due to the fact that the bus line with ULID 50 is much shorter and the boarding time intervals between trips are longer than ULID 443, there has been no bunching observed in the time-space diagrams as in ULID 443. However, the trips having a relatively longer total TT could be identified to evaluate the TT reliability of a day.
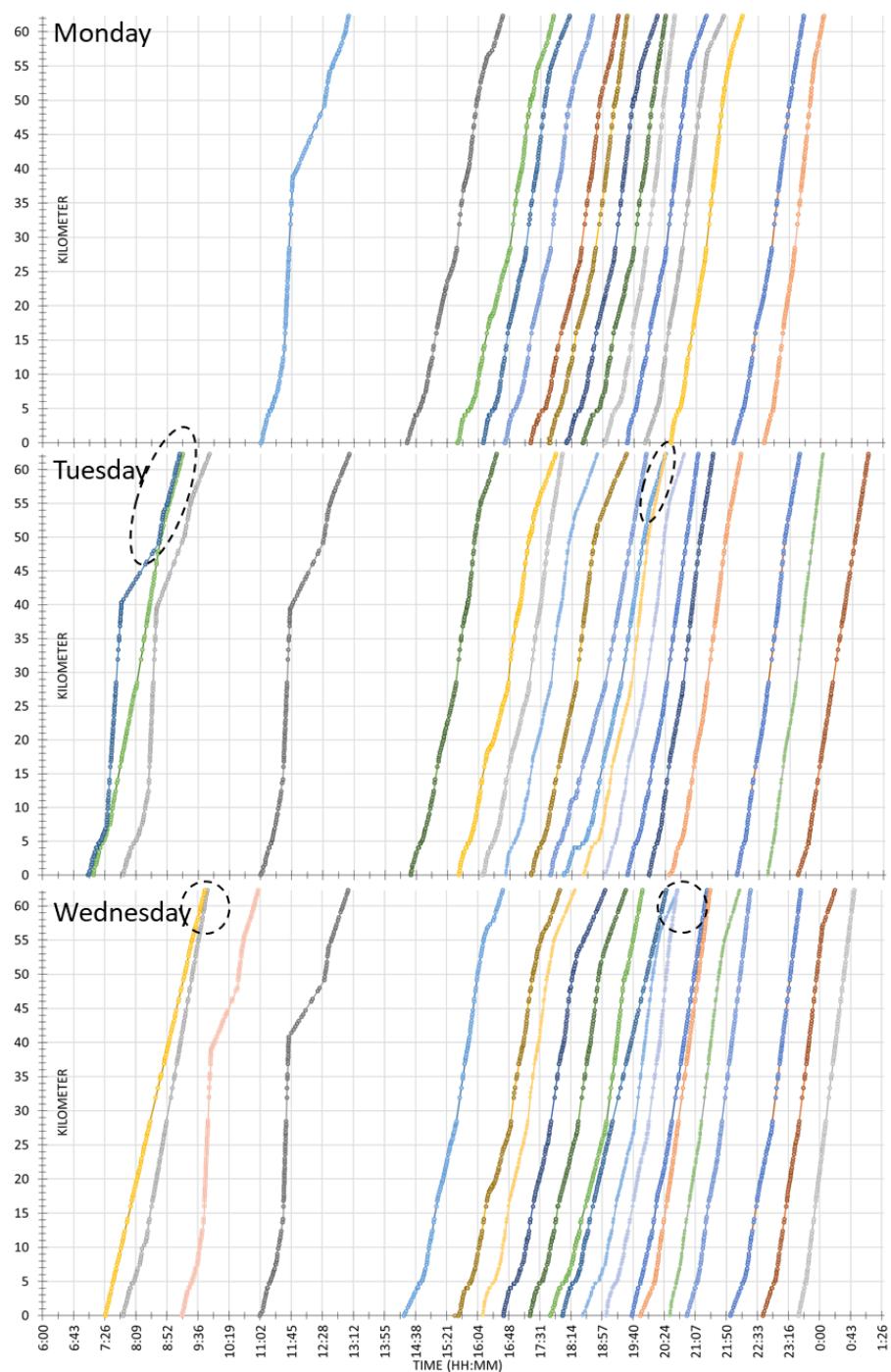
Figure 4.20. Time-Space Daily Profile on Monday, Tuesday, Wednesday for ULID 443
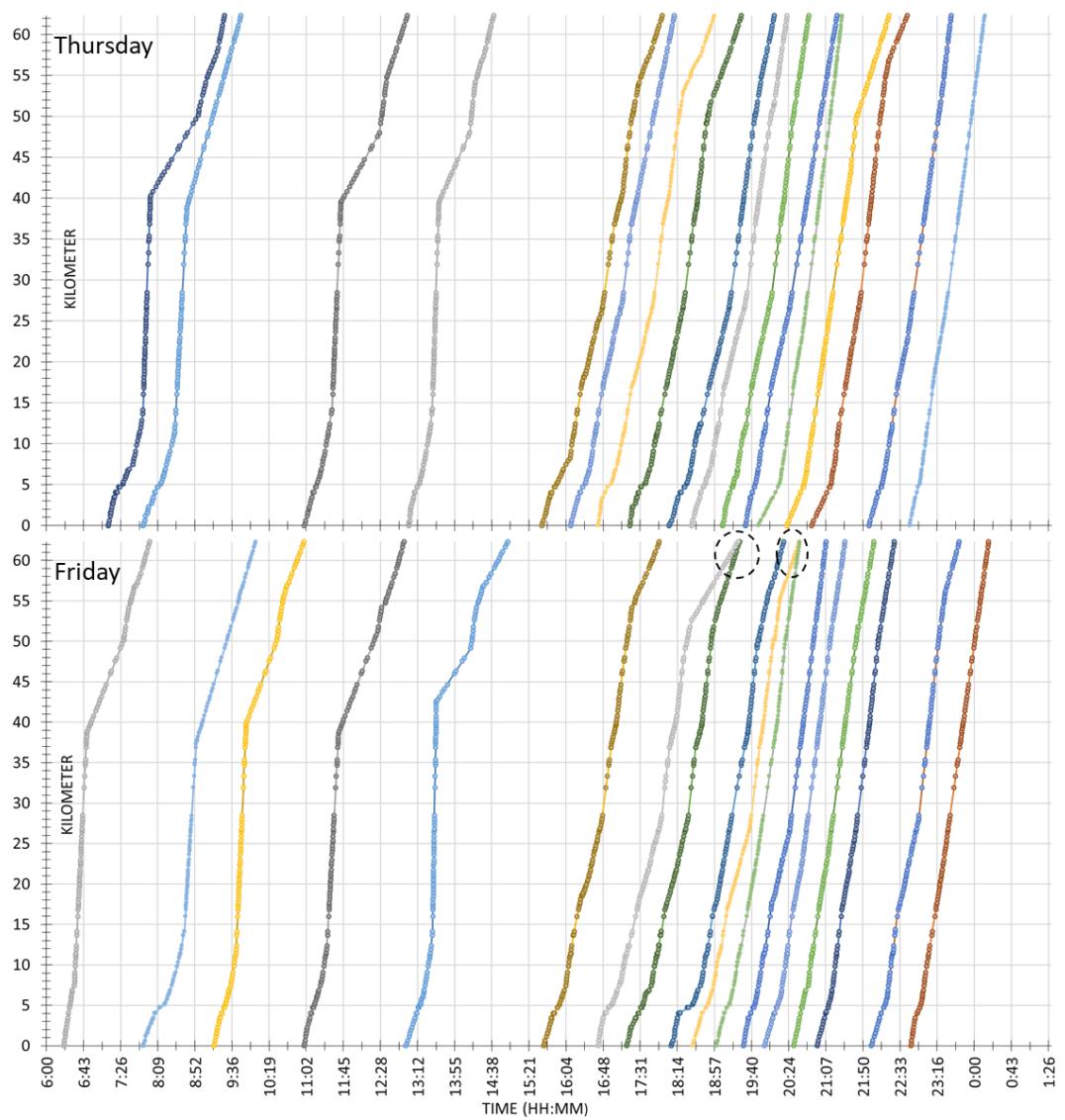
Figure 4.21. Time-Space Daily Profile – Thursday and Friday for ULID 443
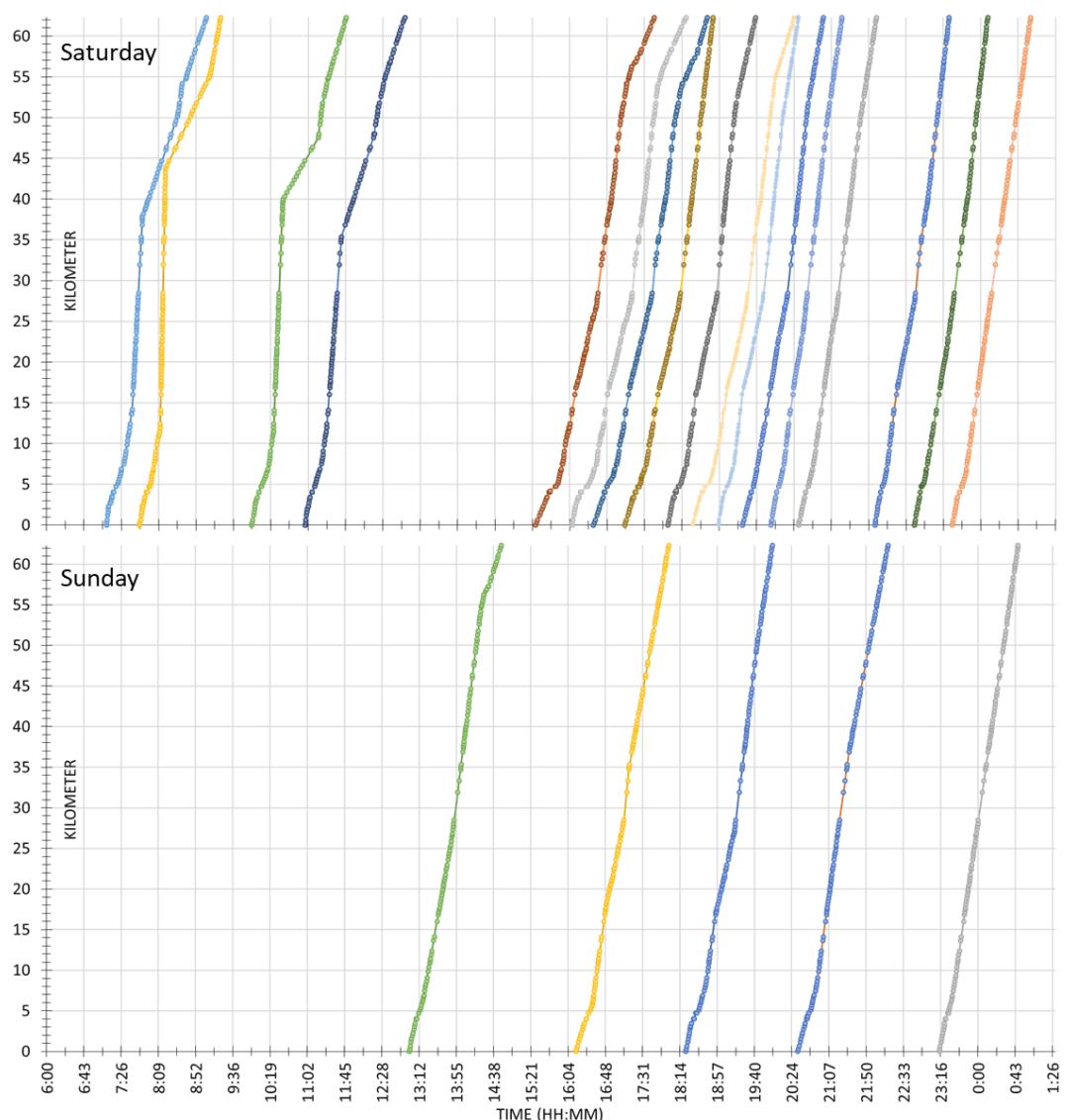
Figure 4.22. Time-Space Daily Profile – Saturday and Sunday for ULID 443 Trips
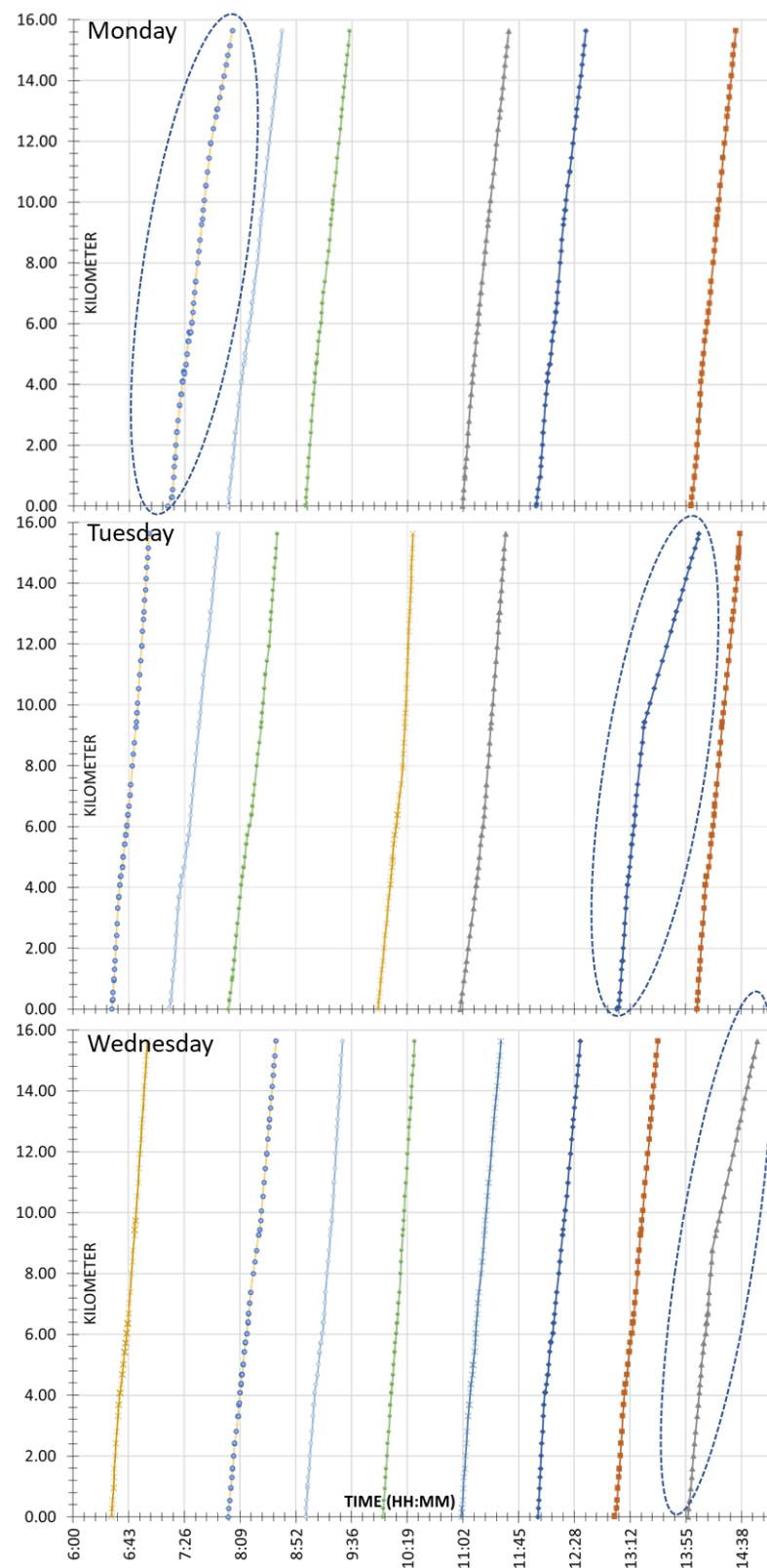
Figure 4.23. Time-Space Daily Profile on Monday, Tuesday and Wednesday for ULID 50
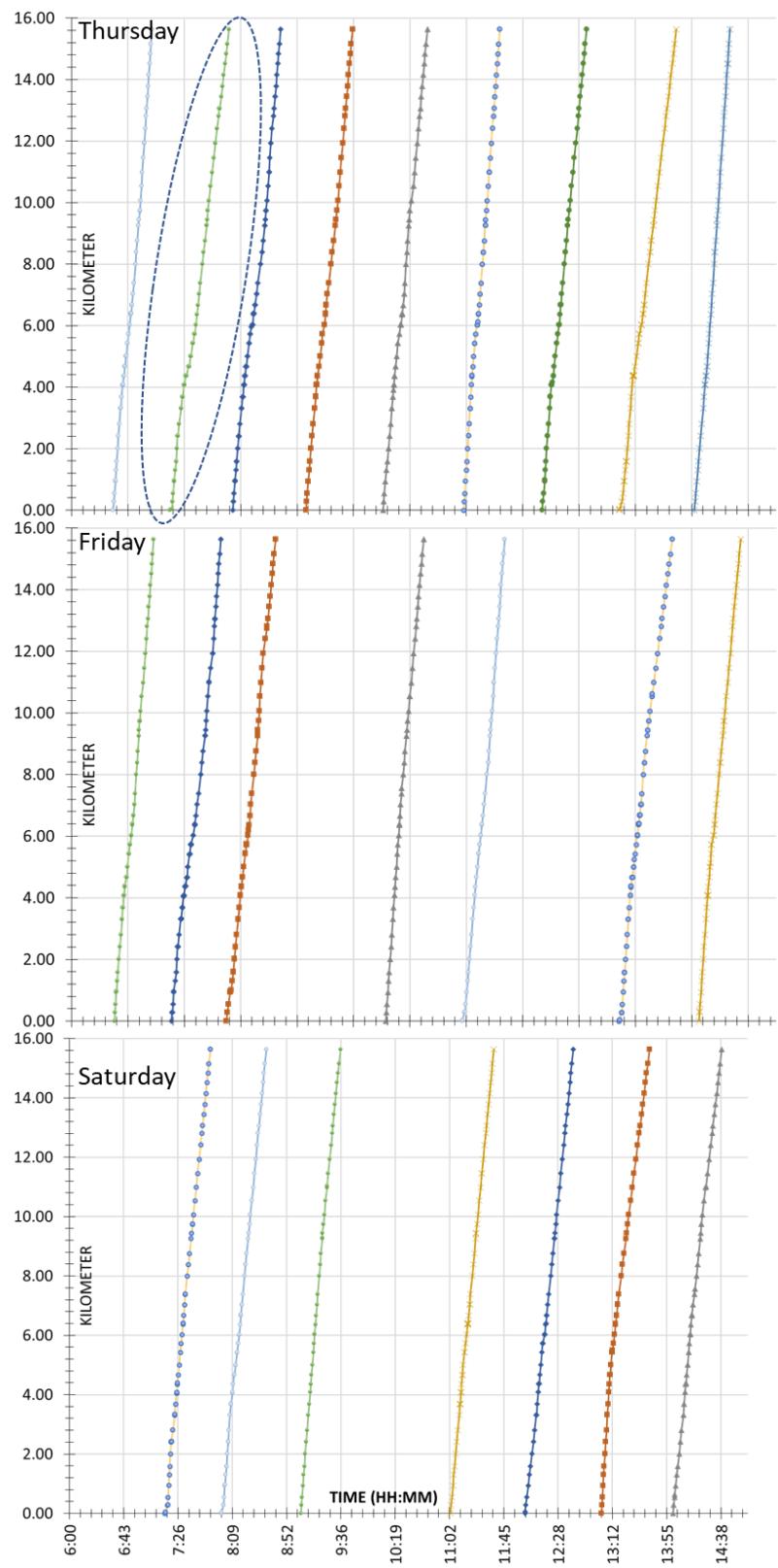
Figure 4.24. Time-Space Daily Profile on Thursday, Friday, Saturday for ULID 50

4.7.1   Weekly SC Based Variability Diagrams
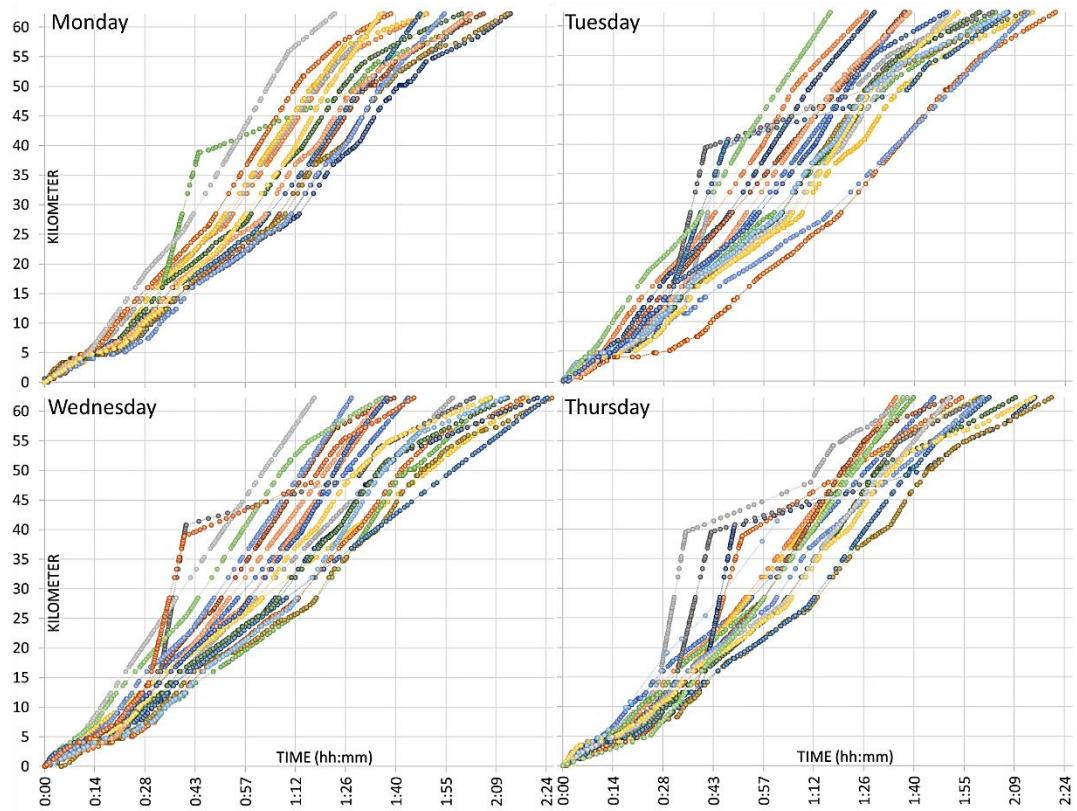
Figure 4.25. Weekly SCD based variability diagrams for Monday-Thursday for ULID 443
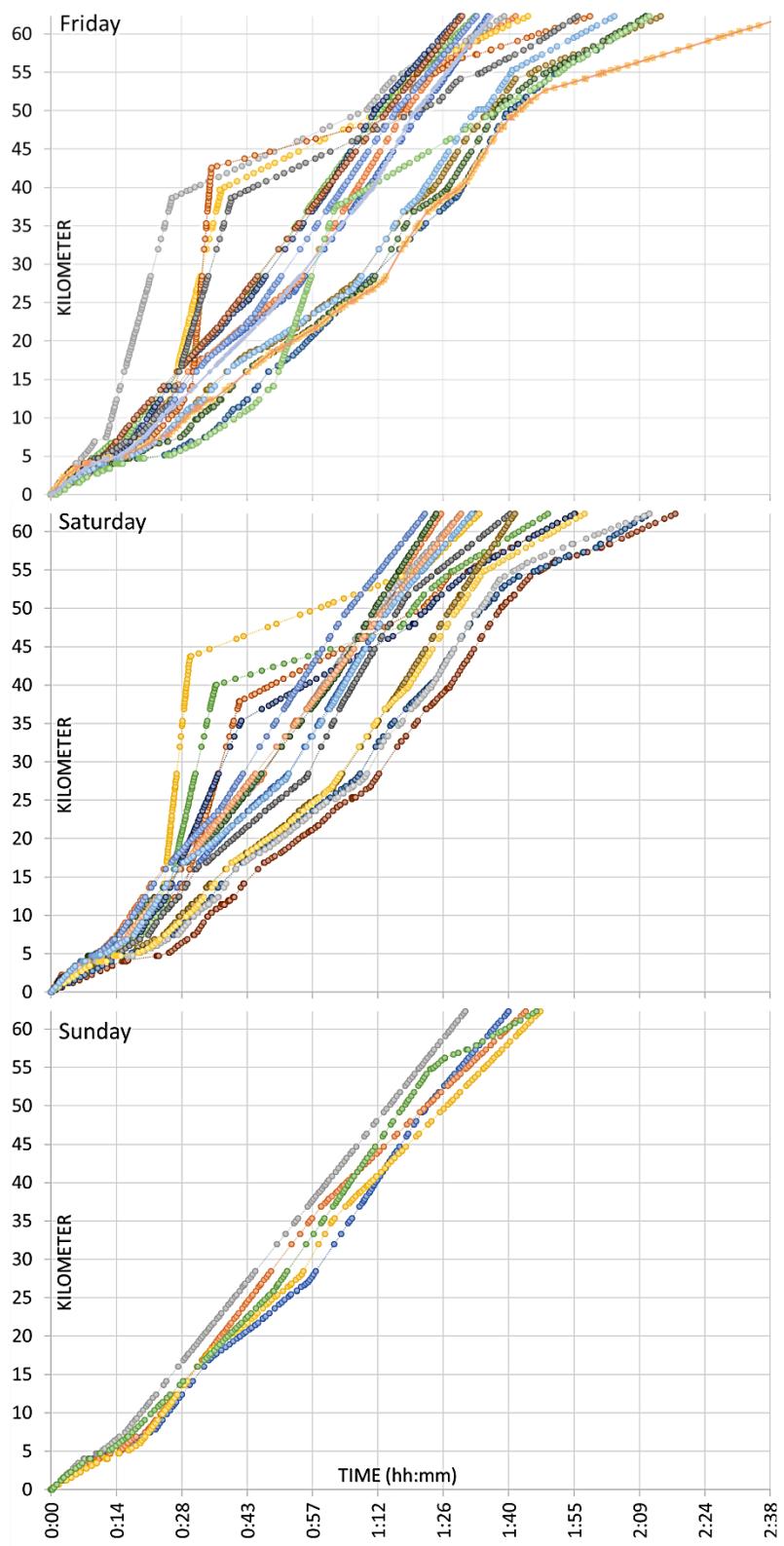
Figure 4.26. Weekly SCD based variability diagrams for Friday-Sunday for ULID 443

A reliability evaluation has been made over the weekly time-space diagrams in Figure 4.25 and Figure 4.26. All timestamps are reduced to zero for the evaluation of the daily TT reliability in terms of the variability of total travel times in a day. The scattering at the top of the diagrams indicates the TT reliability of that day in terms of total travel times. The higher scattering indicates lower TT reliability for the day. The fourth week of October and the bus line with ULID 443 has been selected for the first weekly study case in these diagrams. All smart card datasets belonging to days between the 22nd and 28th of October have processed, as mentioned in the methodology chapter. It has been observed that Monday and Thursday trips are the most reliable days among weekdays as expected since the evaluation based on time-space profile diagrams in the previous section was the same for these days. On the other hand, an excessive slowdown in the trip v515_1 after 50th kilometer affected the TT reliability of Friday considerably bad.
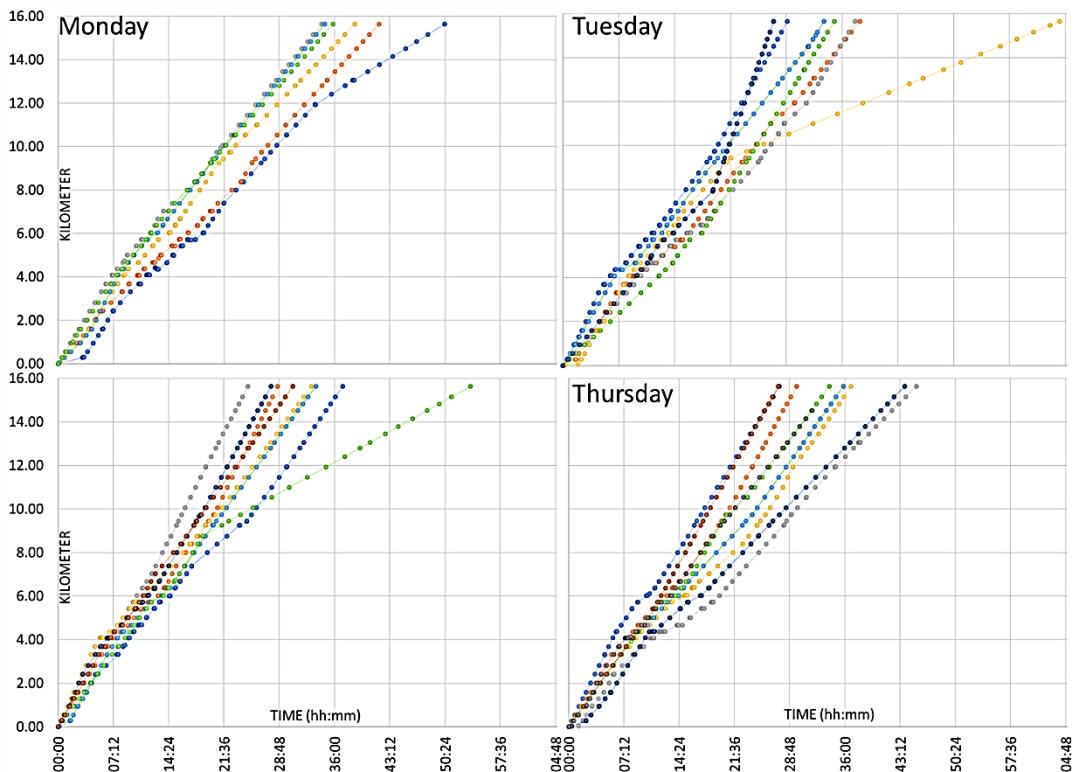
Figure 4.27. Weekly SCD based variability diagrams for Monday-Thursday for ULID 50
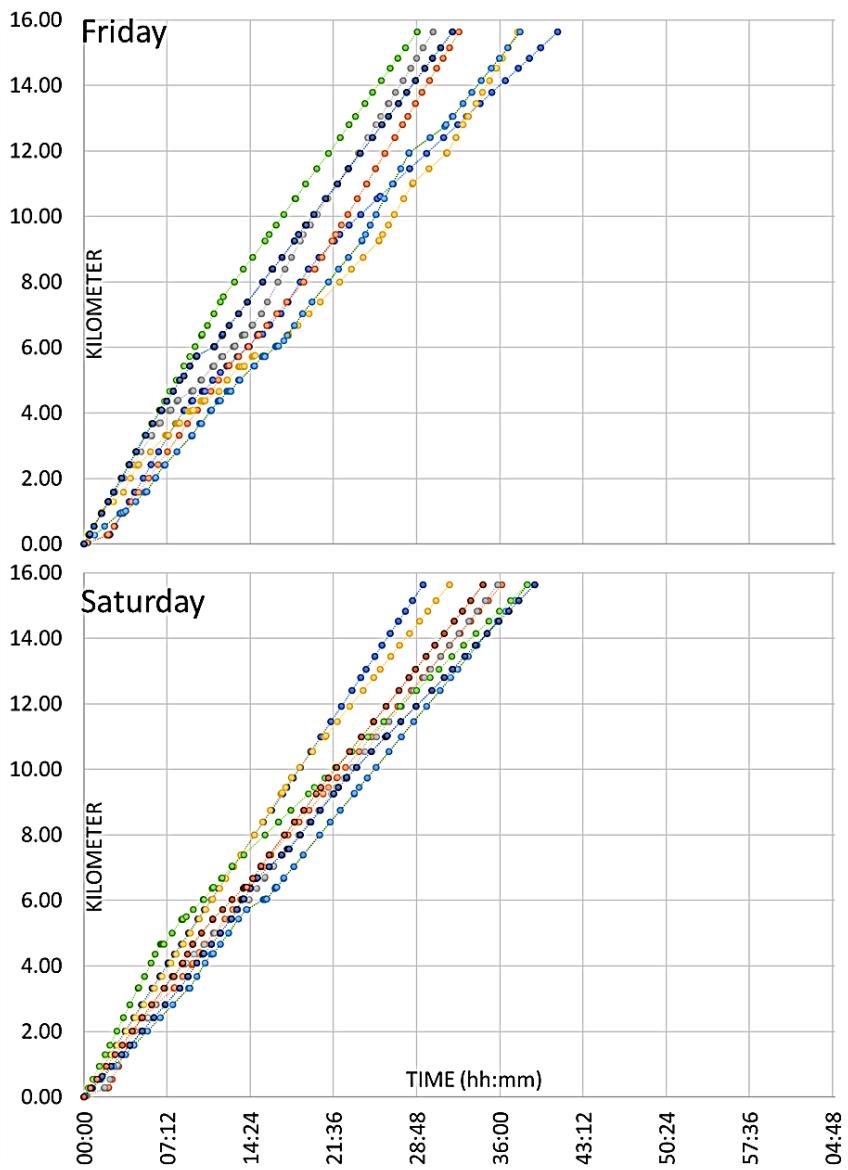
Figure 4.28. Weekly SCD based variability diagrams for Friday, Sunday for ULID 50

Another TT reliability evaluation has been made over the weekly time-space diagrams with reduced timestamps in Figure 4.27 and Figure 4.28. The fourth week of October and the bus line with ULID 50 has been selected for the second weekly study case in these diagrams. All smart card datasets belonging to days between the 22nd and 27th of October have processed, as mentioned in the methodology chapter.

There is not any trip taking place on Sunday in this line. It has been observed that Monday and Friday's trips are the most reliable days among weekdays. On the other hand, the Tuesday trip also seems reliable if the trip v590_5 ignored. However, when we examine the dataset and estimated timestamps of this trip, it is observed that an excessive slowdown after 9th km negatively affected the TT reliability of this day. Saturday trip as a weekend trip seems quite reliable considering the same day in the first weekly study case for ULID 443.

4.7.2   Variability of the PT Service TTs within a Day

Temporal analysis of the trip total travel times according to the departure times in a day has been conducted to see the time-dependent variability throughout the day.
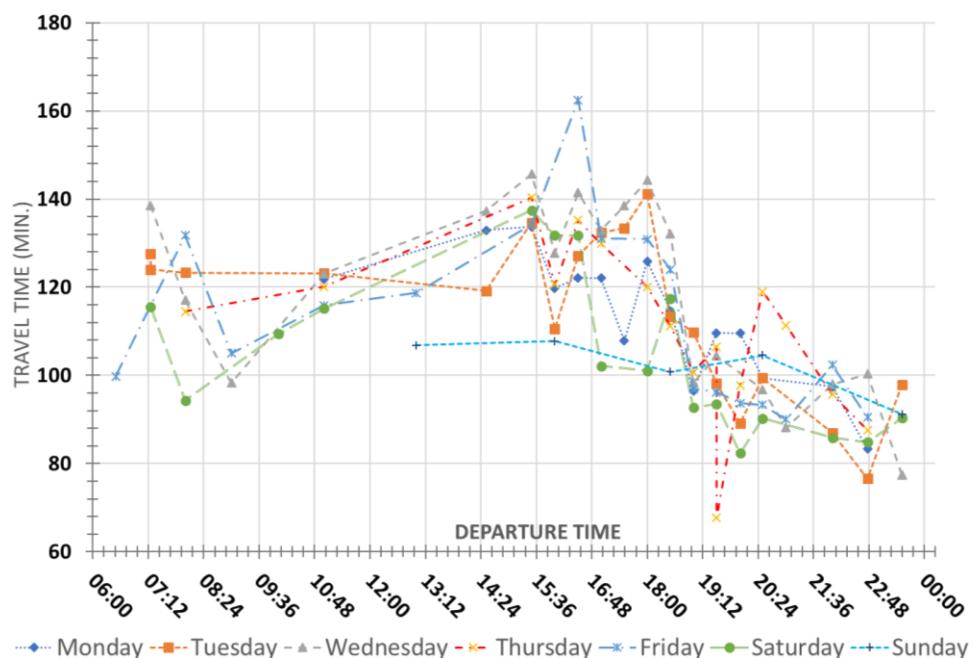
Figure 4.29. Travel times according to their departure times for ULID 443

Figure 4.29 has been evaluated to infer the reliability of travel tome according to the hours of the day. Travel times are high as expected on morning peak hours between 7 and 9 AM except for a Saturday trip with departure time at 8 AM. There is a considerable increase from 9 AM until 3 PM for almost all trips except for Sunday.

After 6 PM, which is the end of the evening peak hours (5 PM - 6 PM), travel times decreasing in almost all trips as expected. On the other hand, travel times for Sunday trips in the evening peak hours are not as low as other days.
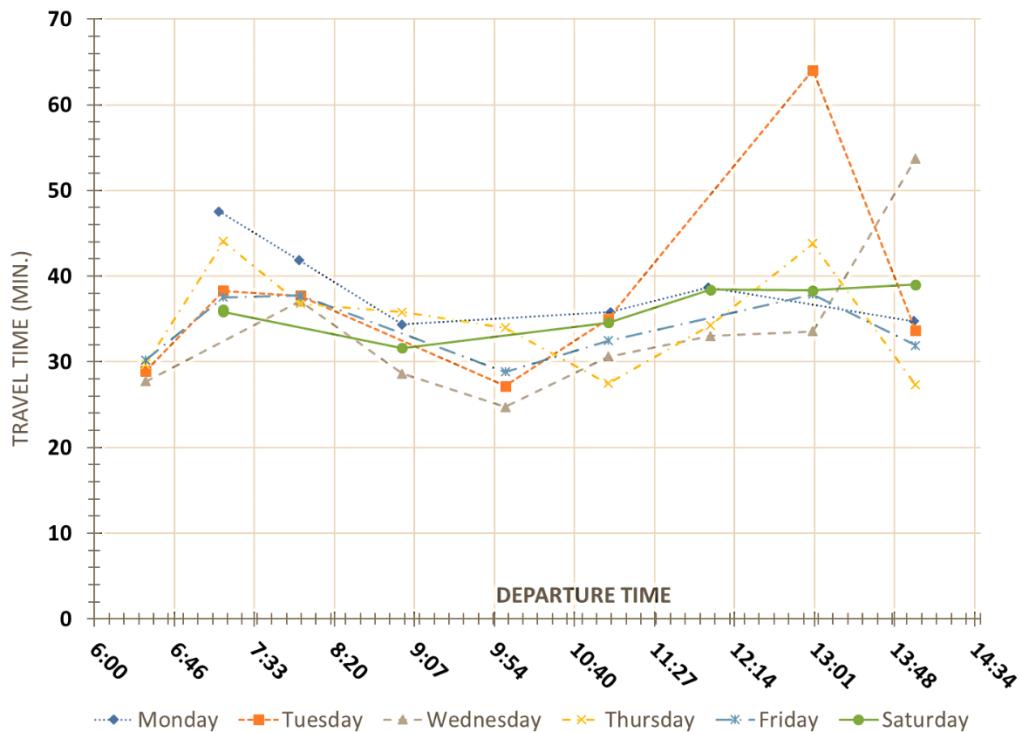
Figure 4.30. Travel times according to their departure times for ULID 50

There can be seen in Figure 4.30 that the trips on the route with ULID 50, takes place between 6:30 AM and 2 AM. There are high travel times at 7 AM trips, but travel times are decreasing until the end of the morning peak hours, 9 AM. The travel times of afternoon trips that take place on Tuesday, Wednesday and Thursday are considerably high as expected from the evaluation made in the previous sections of this study. The specific trip numbers which increase the mean travel times can be detected by using the time-space diagrams mentioned before in this chapter.

4.7.3 PT Travel Time (TT) Reliability Indicators

The total travel times of each trip has taken in to count for the time-dependent variability. The conventional travel time statistics, which will be the base for the calculation of the TT reliability indicators, have been calculated (see Table 4.1.). The lower travel times statistics indicate shorter travel times and early arrivals on weekends concerning the weekdays as expected for the PT line with ULID 443. On the other hand, while the mean and median travel times on weekends for the PT line with ULID 50 indicates longer travel times considering the weekdays, the standard deviation and 95th percentile travel time statistics indicate shorter travel times on weekend as expected.

The reliability indicators for all days, weekdays and weekends are calculated for each PT line and each direction in Table 4.2. The higher the value, the lower the TT reliability, according to this table. As we can see, the outbound direction of the selected PT line with ULID 50 has the lowest reliability, especially for weekdays. The weekend trips of the ULID 50 have higher reliability concerning the weekday trips as expected. The inbound direction of both PT lines has marginally higher reliability compared to their outbound counterparts. The inbound direction of the ULID 50 has the highest reliability among all PT lines, especially for weekends. The outbound direction of the ULID 443 has the lowest reliability on weekends considering all PT lines and directions.

All four indicators have a positive correlation in PT line with ULID 50 for both directions. Contrary, the ULID 443 weekend trips have slightly lower reliability as not expected according to the weekday trips based on the buffer index values. If we look at the Normalized Standard Deviation and Planning Index values, weekend trips are more reliable, according to the Normalized index and Planning index values.

Table 4.1. Travel time (TT) reliability Statistics

Weekday status	Mean TT (mins)	Median TT (mins)	STD (mins)	95th Percentile TT (min)	Free-Flow TT (min)
PT Line ULID 443 – Inbound					
All days	72.71	73.09	13.58	94.56	43.98
Weekdays	73.56	73.75	14.19	96.58	43.98
Weekends	69.11	66.69	10.13	89.33	43.98
PT Line ULID 443 – Outbound					
All days	39.09	36.43	13.22	67.02	23.98
Weekdays	40.03	38.34	13.51	69.38	23.98
Weekends	35.06	30.87	11.33	61.93	23.98
PT Line ULID 50 – Inbound					
All days	19.75	19.10	2.73	24.66	8.72
Weekdays	19.82	18.95	2.88	24.80	8.72
Weekends	19.37	19.50	1.95	22.19	8.72
PT Line ULID 50 – Outbound					
All days	15.95	15.08	6.14	32.47	8.33
Weekdays	15.77	14.77	6.65	35.79	8.33
Weekends	16.91	16.20	1.89	20.41	8.33

Table 4.2. Travel time (TT) reliability Indicators

Weekday status	Buffer Index (Mean Based)	Buffer Index (Median based)	Normalized SD	Planning Index
PT Line ULID 443 – Inbound				
All days	0.30	0.29	0.31	2.15
Weekdays	0.31	0.31	0.32	2.20
Weekends	0.29	0.34	0.23	2.03
PT Line ULID 443 – Outbound				
All days	0.71	0.84	0.55	2.79
Weekdays	0.73	0.81	0.56	2.89
Weekends	0.77	1.01	0.47	2.58
PT Line ULID 50 – Inbound				
All days	0.25	0.29	0.31	2.83
Weekdays	0.25	0.31	0.33	2.84
Weekends	0.15	0.14	0.22	2.54
PT Line ULID 50 – Outbound				
All days	1.04	1.15	0.74	3.90
Weekdays	1.27	1.42	0.80	4.30
Weekends	0.21	0.26	0.23	2.45

Correlation Between Reliability Indicators

After the calculation of TT reliability by using these four most used and advised indicators, we can evaluate and chose some of them for further evaluations by using more sample datasets. As presented in Table 4.3, Median based Buffer Index has the strongest positive correlation with all other indicators overall. However, Mean based Buffer Index and Normalized Standard Deviation have close values to the Median Based Buffer Index. Planning Index has the weakest positive correlation among all others.

The correlation between weekday indicators are shown in Table 4.3. Again, the Median Based Buffer Index has the most positive correlation with other indicators. Mean based Buffer Index and Normalized Standard Deviation have closer values to the Median Based Buffer Index this time. Planning Index has the weakest positive

correlation among all others once again. Weekend correlations are not much different from the ones on weekdays and overall tables. The planning Index has the weakest positive correlation again. Mean based Buffer Index has a slightly better positive correlation then Median Based Buffer Index and Normalized Standard Deviation among weekend indicators.

Consequently, the Median Based Buffer Index has a strong positive correlation, considering the other three indicators. Hence considering the Median Based Buffer Index values, the inbound direction of ULID 50 has the highest TT reliability for overall, weekdays and weekend trips. On the other hand, outbound of the same route, ULID 50 has the lowest TT reliability for overall and weekdays trips. The outbound direction of ULID 443 has the lowest TT reliability between weekend trips.

Table 4.3. Correlation Matrix for Weekly Indicators

Indicator	BI(mean)	NSD	PI	BI(median)
(All days)				
BI(mean)	1.0000	0.9970	0.8469	0.9943
NSD	0.9970	1.0000	0.8389	0.9993
PI	0.8469	0.8389	1.0000	0.8480
BI(median)	0.9943	0.9993	0.8480	1.0000
Weekdays				
BI(mean)	1.0000	0.9908	0.9102	0.9980
NSD	0.9908	1.0000	0.8916	0.9927
PI	0.9102	0.8916	1.0000	0.9292
BI(median)	0.9980	0.9927	0.9292	1.0000
Weekdays				
BI(mean)	1.0000	0.9940	0.3750	0.9945
NSD	0.9940	1.0000	0.4459	0.9901
PI	1.0000	0.4459	1.0000	0.4485
BI(median)	0.9945	0.9901	0.4485	1.0000
BI: Buffer Index, **NSD**: Normalized Standard Deviation, **PI**: Planning Index				

4.7.4 Findings Regarding the Reliability of PT using SDC

Consequently, total travel time reliability and inter-stop travel time reliability measures could have been derived from both SCD and BTD. Then the results have been evaluated and compared visually and numerically in this chapter. The results indicate that total travel time reliability measures can be derived from SCD instead of BTD for a bus line and the results can give enough idea about the total travel time reliability of a trip or line. On the other hand, the health of the inter-stop travel time reliability measures derived from SCD, depends on the distribution of smart card readings along the PT line. Even on the ULD443 line, which is one of the busiest lines, it has been observed that there is not enough SCD distribution in many trips for a healthy inter-stop reliability assessment. Therefore, smart card based total travel time reliability measures are healthier than inter-stop travel time reliability measures. Even so, a healthy inter-stop travel time reliability evaluation can be made over some selected PT lines with well distributed smart card reading by using the average dataset of a large time span.

4.8 Comparisons of SCD and BTD based Evaluations

Bus trajectory and SCD based time-space diagrams have been combined for the comparison of two datasets in terms of arrival time estimations (ATE). The ATE (BTD) stands for the bus-stops estimated based on BTD. The ATE (SCD-Int.) stands for the bus-stops estimated based on the interpolation of SCD and the ATE (SCD-Ext.) stands for the bus-stops estimated based on the extrapolation of SCD.

It can be seen that smart card locations in time, matches the bus trajectory in Figure 4.31a. As long as there are no smart card readings, the smart card based trajectory has a linear flow according to the bus trajectory. The smart card based ATE of bus-stop locations over bus trajectory for the trip v246_1 can be seen in Figure 4.31b. The arrival time of bus-stops estimated by interpolation calculations ATE (SCD-Int.) mostly flows over the trajectory. On the other hand, bus-stops' arrival time estimated
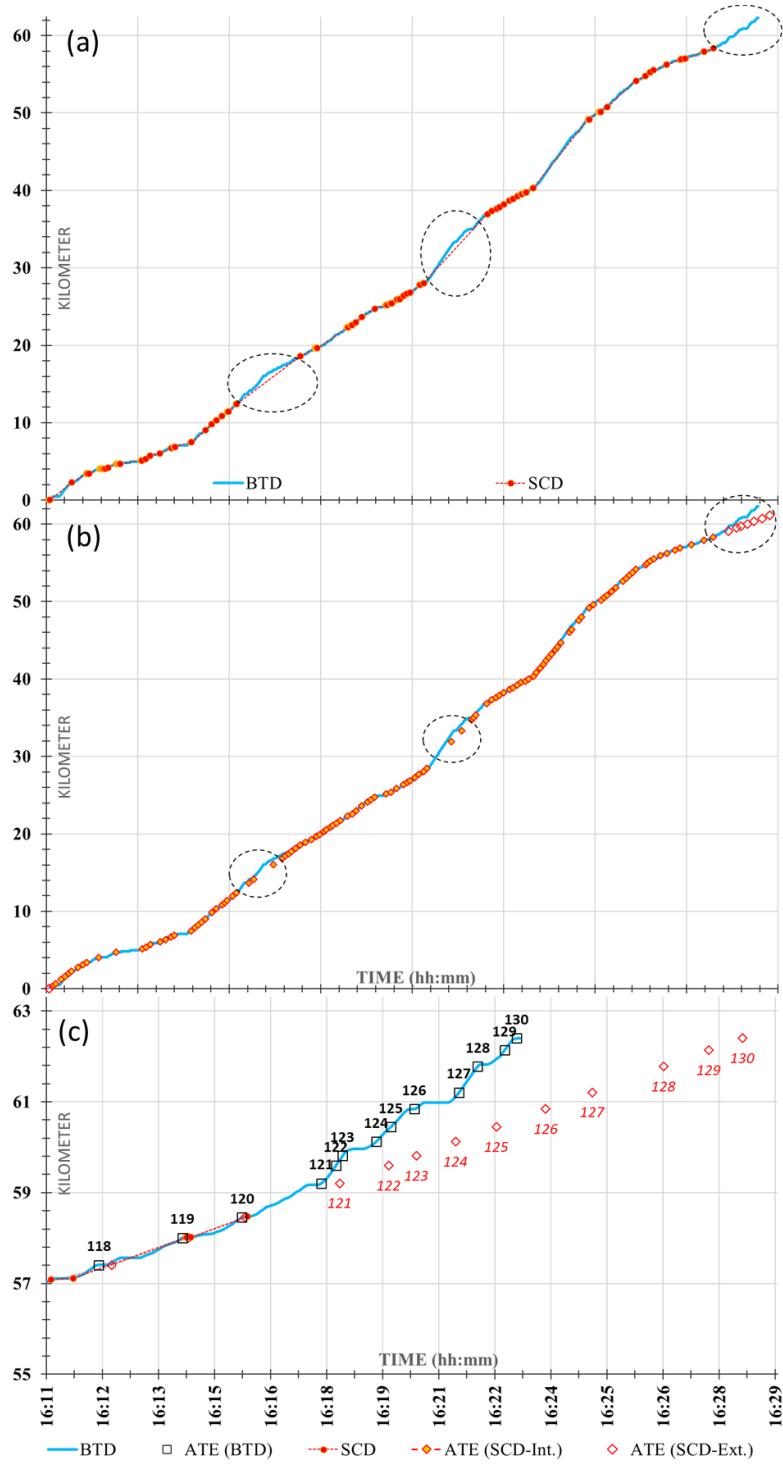
Figure 4.31. (a) raw SCD over BTD, (b) SCD based ATE and (c) ATE from SCD Extrapolation for trip v246_1

by extrapolation calculations ATE (SCD-Ext.) is diverting from the bus trajectory. It can be seen that, arrival time estimations with interpolation (ATE (SCD-Int.)) in the parts of the bus trajectory with sparse SCD also differ from the bus trajectory. In these areas, the smart card data based trajectory follows a linear pattern. If we look closer to the part of ATE (SCD-Ext.) in in Figure 4.31c, there is almost 6 minutes difference for the ATE of the last bus-stop in terms of total travel time between smart card based and bus trajectory based dataset. The differences in ATE (BTD) and ATE (SCD-Ext.) varies between 1 minute to 6 minutes.

Possible alighting or bus stops with congestions can be detected by the help of the smart card dataset. It is obvious in Figure 4.32 that the bus trajectory stands still in time for about 30 seconds, even there are no smart card readings on the bus stops 69, 70, 71. Smart card data can not catch this kind of slowdown in vehicle movements. On the other hand, we can observe and evaluate the possibilities that causing the slowdown of the vehicle with the help of a SCD.
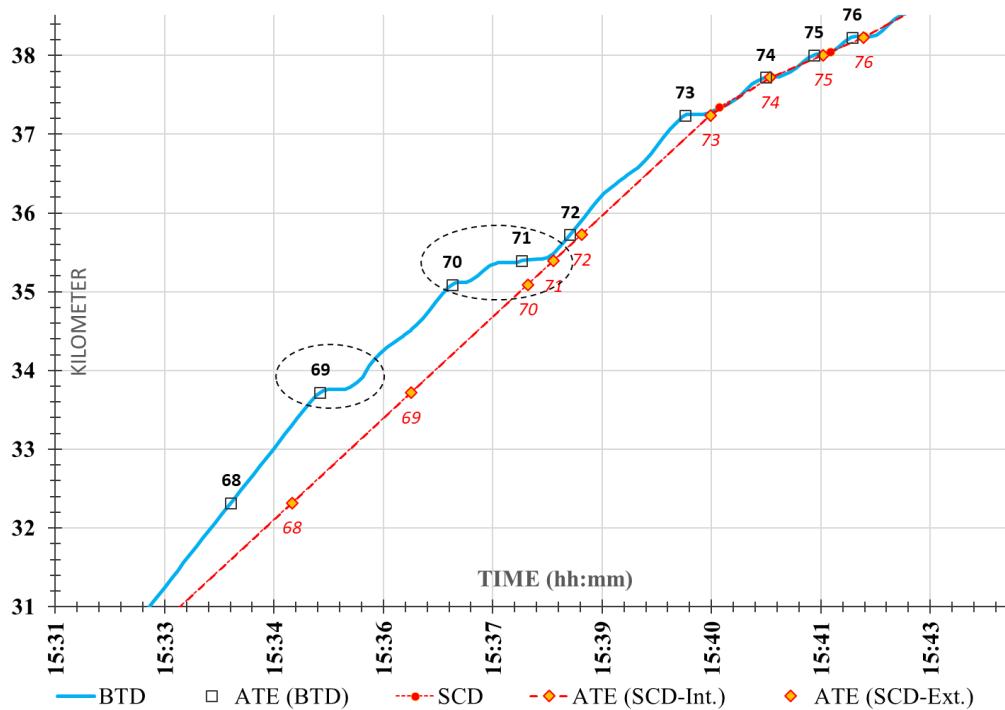
Figure 4.32. Possible alighting or congestions in trip v246_1

These two trips in Figure 4.33a, bus trajectory based v246_3 and smart card based v246_2, belong to the same trip concerning their time interval. Even if most card readings take place in the first 15 kilometers of the route, there is still sufficient distributed card readings until the 52. Km. The reason for the divergence between smart card based ATE and bus trajectory is the lack of the smart card record at the end segments of the route (Figure 4.33b). There is almost 8 minutes difference in terms of ATE of the last bus stops between BTD and SCD.
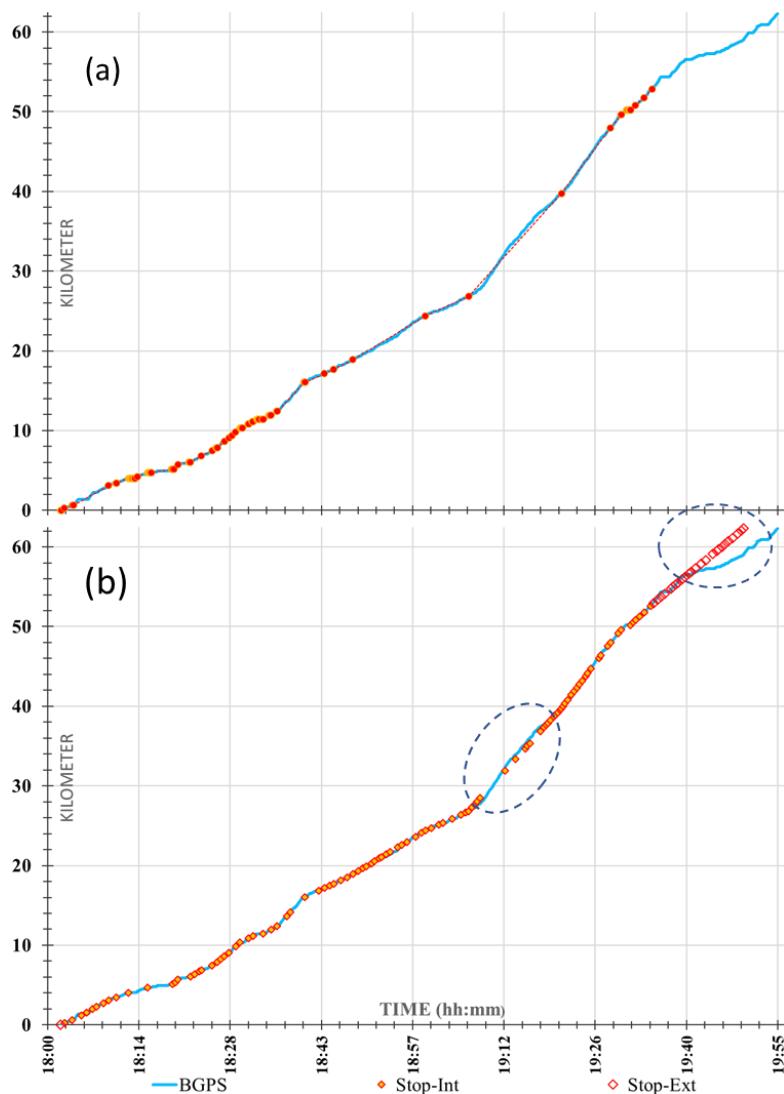
Figure 4.33. (a) raw SCD over BTD, (b) SCD based ATE for trip v246_2/3

There is speedup in ATE of bus stops for the trip v246_1 in Figure 4.31b, while there is slowdown in the ATE of bus stops for the trip v246_2/3 in Figure 4.33b. A more precise ATE can be done for the bus stops without any smart card record, by using or developing an improved extrapolation method. For instance, the chosen SCD pairs for the extrapolation bus stop ATEs are critical for a precise estimation process. Selecting too close or too far SCD pairs affect the extrapolation results marginally positive or negative. Smart card based ATE of bus stop locations over bus trajectory in trip v211_1 is presented in Figure 4.34. The arrival time of bus-stops estimated by extrapolation calculations ATE (SCD-Ext.) is not much diverting from the bus trajectory as in trip v246_1 in Figure 4.31. The difference in ATE of ATE (BTD) and ATE (SCD-Ext.) varies between 15 seconds to 1 minute.

As can be seen in the Figure 4.35, the dwell location of the bus is approximately 180 meters ahead of the bus stop location according to the bus stop dataset. No SCD was recorded at this location, where the bus waited for approximately 90 seconds. In addition to the fact that this location is a possible alighting or congestion zone, the differences between the actual stop location and the used stop location can be determined in such time-space diagrams by using BTD. On the other hand, SCD based bus stop ATE can be more reliable in such cases. SCD based ATE of the 4th bus stop reflects the ground truth better than BCD based ATE in Figure 4.35.

The trip v632_1 is the one with the least smart card readings on all trips (Figure 4.36a). There are only 100 smart card readings between the 9th and 121st bus stops. However, the distribution of smart card reading along the route is not that bad. As long as the smart card record has proportionally distributed, a more precise interpolation of bus stops can be achieved. Bus-GPS and Smart card based inter-stop travel time diagrams have combined to compare each other in Figure 4.37a, Figure 4.37b, Figure 4.38a and Figure 4.38b. It was observed that the most difference in terms of inter-stop travel times between SCD and BTD was occurring at the bus-stops with orders 68, 122 and 128 in all four trips. For a more in-depth evaluation of the ATE differences in terms of inter-stop travel times, some statistics are calculated.

The absolute differences have been used in the calculations since the importance of negative and positive differences are identical in this case.
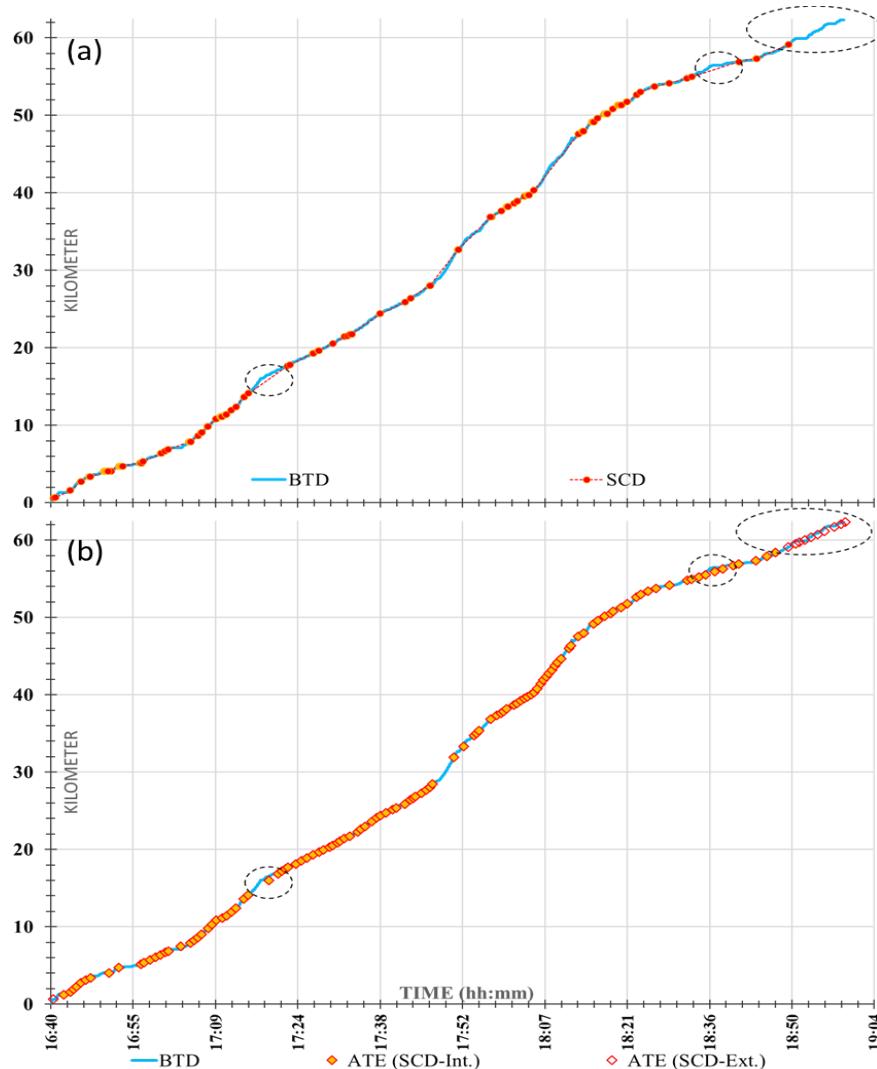
Figure 4.34. (a) raw SCD over BTD, (b) SCD based ATE for trip v211_1
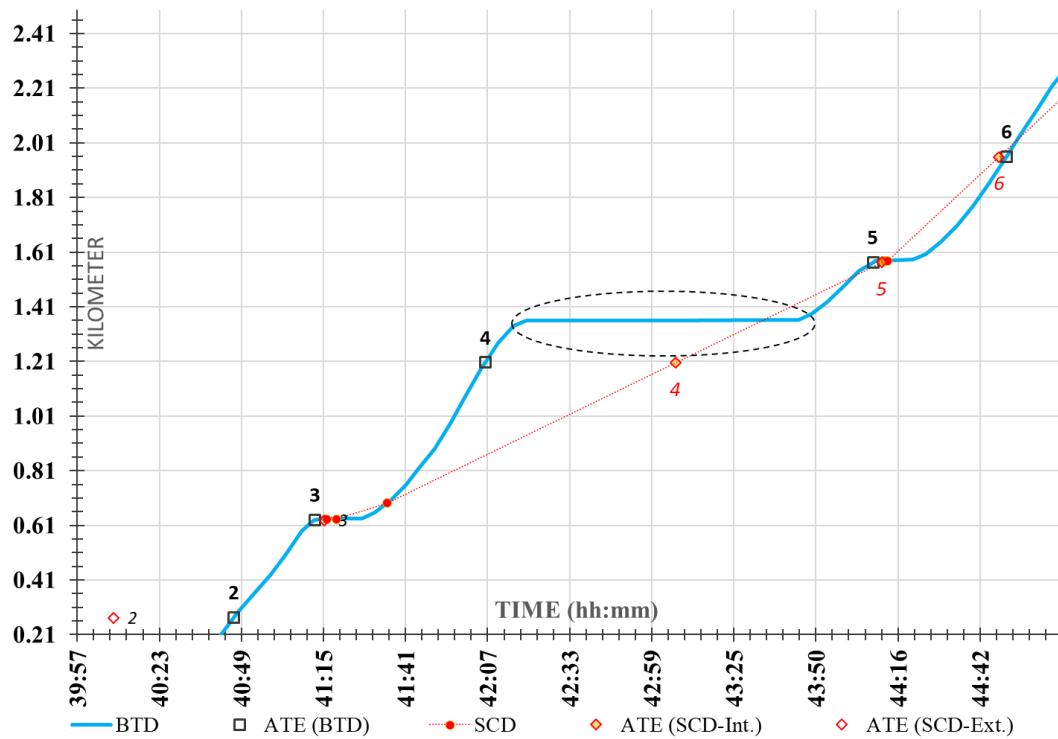
Figure 4.35. Changes in Bus Stop Location
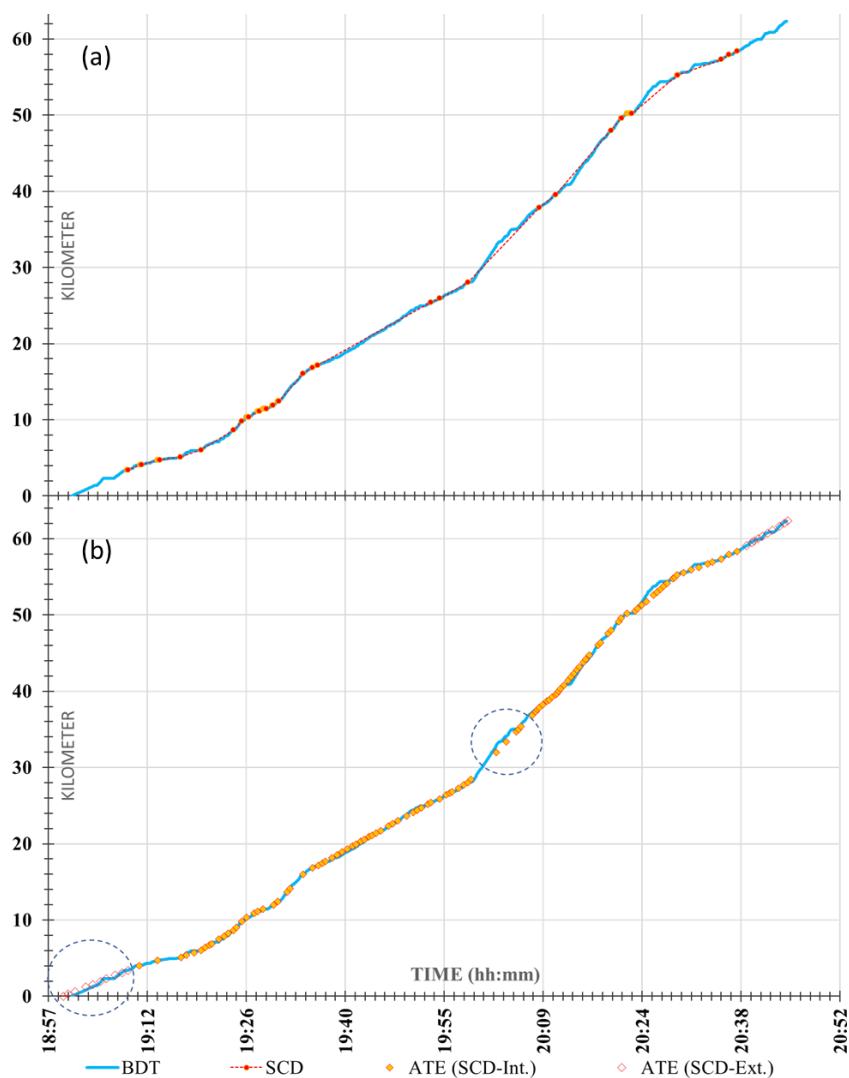
Figure 4.36. Bus-GPS and Smart card data combined trip v632_1

Table 4.4. Inter-Stop Travel Time Differences Between BTD and SCD

	v246_1	v211_1	v246_2/3	v632_1
Standard Deviation	00:00:11	00:00:17	00:00:17	00:00:13
Max. Difference	00:01:22	00:01:33	00:01:40	00:01:16
95th Percentile	00:00:40	00:00:59	00:01:07	00:00:41

Though standard deviations seem positive since all values are limited under 20 seconds for all four trips in Table 4.4, the maximum differences are still quite high for some bus-stops. The 95th percentile is the most reliable statistics for travel times type of datasets since higher time differences are more critical than lower ones. According to 95th percentile values, v246_1 and the v632_1 trips are more reliable considering to the other two trips in terms of the smart card based inter-stop travel time differences with BTD. The 95th percentile values of v211_1 and v246_2/3 considerably higher about one minute, which is not good. A 95th percentile value indicates that %95 of the samples (inter-stop travel time differences between SCD and BTD) are below that value. So, the lower the values, the higher the reliability. We have to be careful while making a stop-based evaluation of the reliability by using SCD based inter-stop travel times like the trips v211_1 and v246_2/3. Hence, evaluating the stop based reliability using SCD based inter-stop travel time is not as healthy as evaluating the total TT reliability based on SCD due to the ATE with insufficient and undistributed smart card dataset in bus stops.
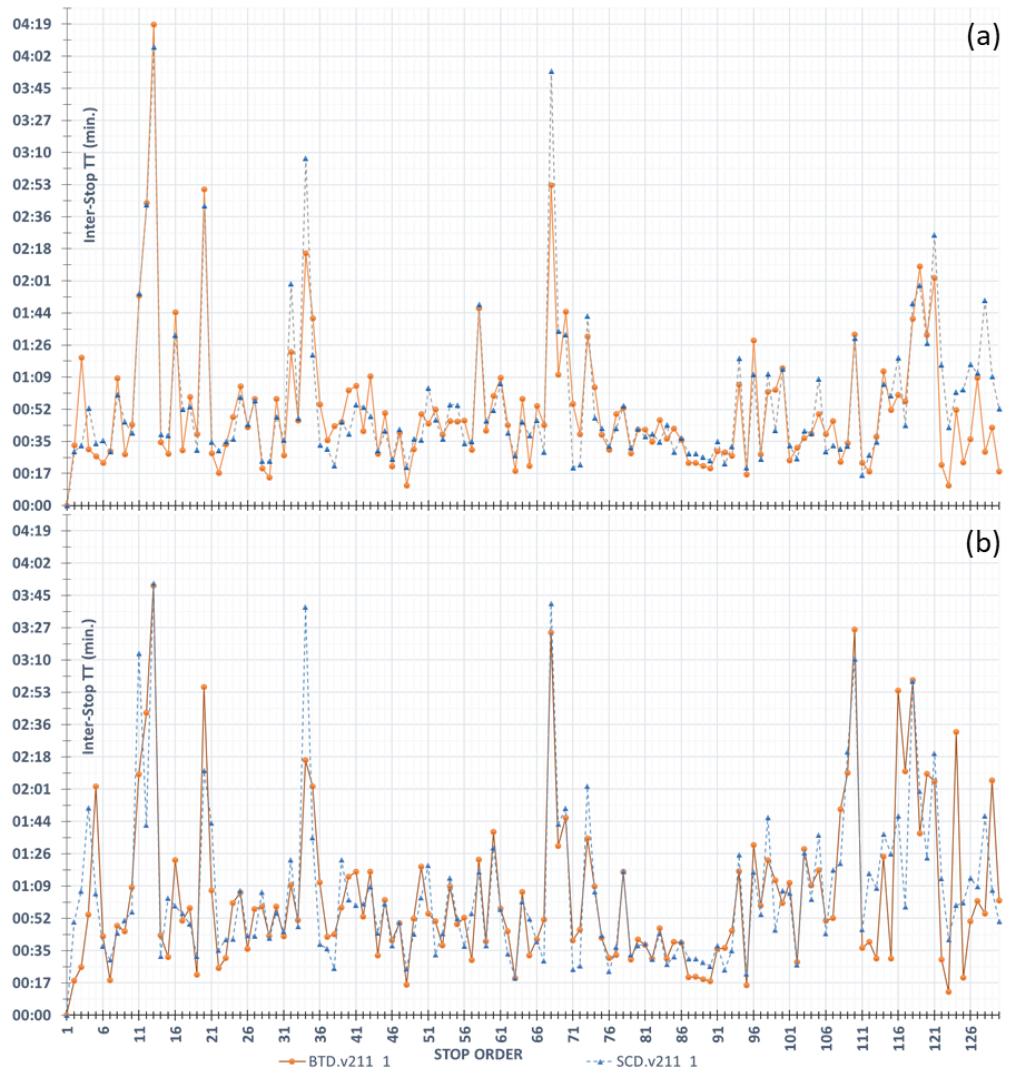
Figure 4.37. BTD – SCD Inter-Stop TT Comparison for the trip (a) v246_1 and (b) v211_1
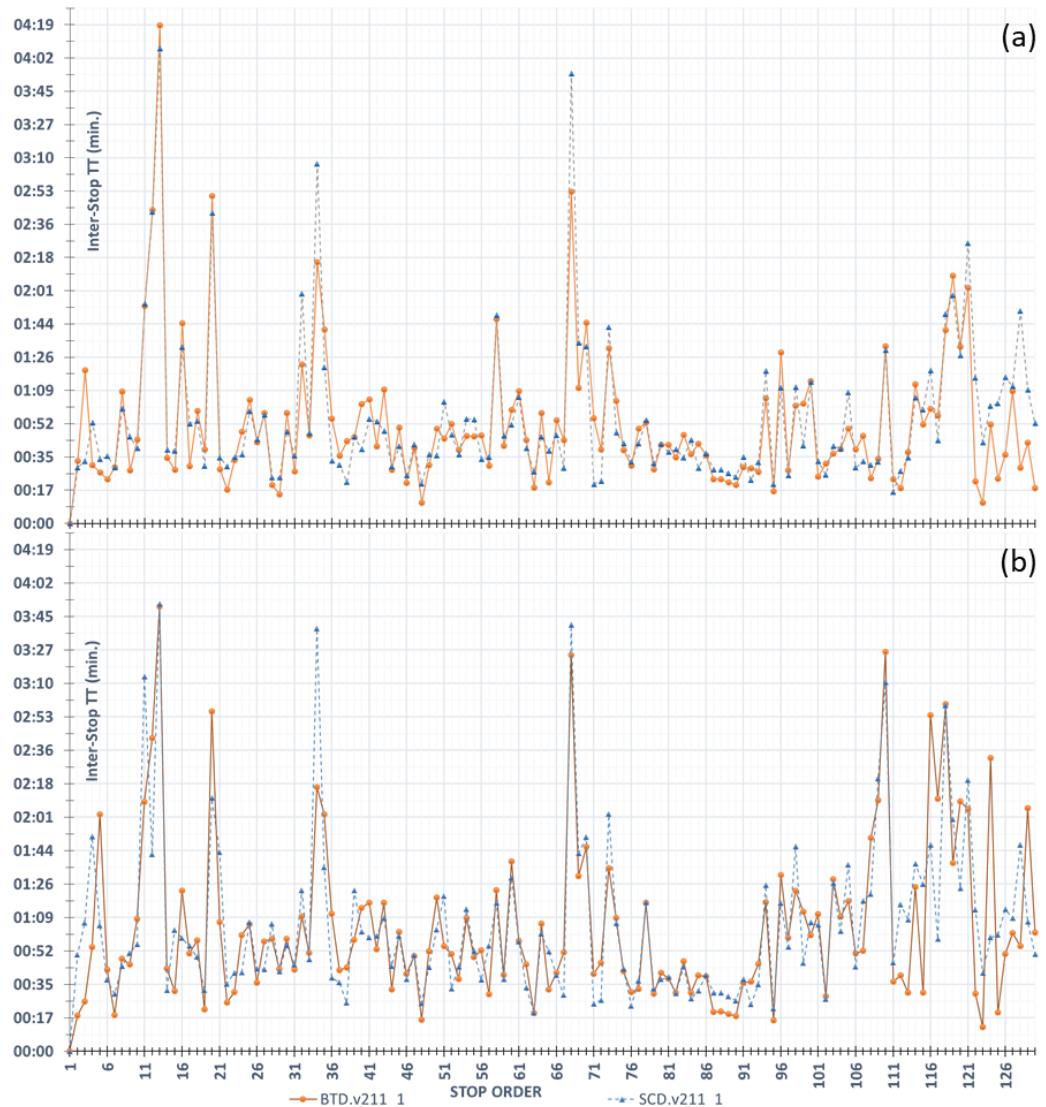
Figure 4.38. BTD – SCD Inter-Stop TT Comparison for the trip (a) v246_3/2 and (b) v632_1

CHAPTER 5

CONCLUSION AND FUTURE RECOMMENDATIONS

5.1   Conclusion

In this thesis, a GIS-based framework has been developed to generate the BTD along all the bus stops using a smart card dataset from the Smart Card Automated Fare Collection Systems (SC-AFC) of Konya Public Transit Network. This framework would be particularly useful for PT analysts and administrators who do not currently have a GIS-based network dataset and GTFS-based schedule and Bus Stop data.

A comparison has been made by using the BTD based trajectory and smart card based trajectory to see the advantages and disadvantages of both datasets. Time-space based diagrams are designed to evaluate the reliability measures visually and reliability indicators calculated to estimate the route based TT reliability.

It can be inferred from this study that using SCD instead of bus trajectory data can be more feasible for public transit administrations as long as they establish the GIS infrastructure of their Public Transit Networks. Although using a weekly dataset could give this kind of useful result, more detailed and realistic analyses could be made by using a larger set of data with the help of a well-designed geodatabase and a LRS. Moreover, it is possible to calculate public transit reliability measures by using SCD with a framework as developed in this study.

5.2   Main Contributions

It has been observed in the study that the research questions mentioned in the introduction chapter have been answered along the methodology and study cases. As responses to the original research questions;

i.   Effectiveness of SCD to estimate the PT reliability?

The answer to the first question has been discussed under the findings titles of the previous chapter. It is observed throughout the study that such provided SCD used in this study is especially effective on deriving the total travel time reliability as one of the PT reliability measures since estimated total travel times can give a relatively better idea about the reliability considering inter-stop travel times. SCD with insufficient distribution and number of records can not be useful for inter-stop travel time reliability evaluations.

ii.   The prerequisites in using SCD in PT reliability estimation

The prerequisites in using the SCD in PT reliability have been investigated, explained and examined along the background, methodology and case study chapters. First, a proper smart card dataset with assigned stop order values, coordinates, and timestamps needs to be provided. Only such a dataset can be used in the developed methodology. Then, the number of smart card readings and the distribution of these smart card readings along the bus line and bus stops should be sufficient to estimate inter-stop travel times and total travel times. At least one smart card boarding record in the dataset for each bus stop would be ideal for the estimation of inter-stop and total travel times of a trip. More SCD for each bus stop would be useful for a better estimation process that reflects the ground truth better. It is almost impossible to estimate travel time and derive travel time reliability, especially for the night trips with very few smart card readings. The second research question has also been explained at the beginning of the methodology chapter. How such provided SCD has to be preprocessed and on what kind of PT line data have to be used to conduct LR processes.

iii.   How SCD should be used to estimate PT reliability?

The answer to the third research question has been discussed in detail in the methodology section. Using the developed framework and GIS tools, it has

been revealed how smart card data can be used until the travel time reliability phase. A customized Linear Referencing process and the GIS tool developed for it were at the core of the methodology. In addition, alternative usage areas in terms of identification of real boarding locations and possible alighting or congestion locations were discussed by comparing with bus trajectory data.

Furthermore, the results of this research can also make a contribution to the study of other researchers, administrations and some municipalities with limited sources in the field of public transportation. It has been observed that almost all studies over PT reliability based on bus trajectory or smart card data have been conducted by taking advantage of a GIS-based network dataset and a stop-based schedule network in GTFS format. Some of these studies have been used sensor data that is more accurate and cleaner than GPS based datasets. On the other hand, most of the studies have been carried on for metro, tram or train networks that are relatively easy to implement linear referencing systems. At the same time, the data sources like the SC-AFC systems used in most of these studies contain both tap-on and tap-off data. Working with such rich datasets provides researchers with a broader scale and easier analysis. While some metropolitan municipalities have this kind of rich datasets, many other local or metropolitan municipalities do not have a GIS-based network dataset or a linear referencing system. Some metropolitan municipalities publish their route, trip and schedule data in GTFS format but still, many municipalities do not have this kind of datasets. While creating and establishing of a GIS-based routable network dataset and stop based schedule data in GTFS format for a complex Bus PT network is a costly, long and cumbersome process, the methodology used in this thesis can facilitate the work of researchers and municipalities who do not have such dataset or a linear referencing system. They will be able to derive some of the PT reliability measures using their historical SCD or BTD without a GIS-based network dataset and schedule data in GTFS format. On the other hand, this methodology with the developed GIS tools could be an opportunity and the starting point for creating a sustainable linear referencing system and a GIS-based network dataset as a part of APTS for such municipalities that lack of these kinds of a dataset.

Another contribution of this thesis is about using the archived historical SCD to derive travel time reliability measures. Since most PT administrations cannot store and manage bus trajectory GPS datasets over long periods due to the large amount of disk place requirements, it can be seen in this thesis that the use of PT smart card datasets can be more effective and accessible, as they cover a considerable amount of less disk space and include more detailed data about passengers.

5.3   Future Recommendations

First, an integrated, sustainable and interoperability based Advanced Public Transportation Information System (APTIS) should be established by designing a geo-database for the better and easy management of datasets. A PT network dataset and GTFS based trip schedules should be generated in this information system. This kind of information system allows decision-makers to make improved analytics of the PT network in terms of public transit reliability measures.

The presented reliability indicators in this study can be used to calculate indexes for different weather conditions or holiday sessions. Therefore, further evaluations can be conducted to see the reliability of the public transit system under different circumstances. If these calculations conducted for reliability indexes can be made using the dataset having a larger time-span, the results of the study can be more realistic.

A data elimination technique and an outlier detection methodology for estimated travel times can be used before calculating reliability indicators. The trips having travel times estimated by unsatisfactory smart card datasets with bad GPS data or the data unequally distributed, excluded manually in this study since only two study PT lines were used. An extended type of study with a large number of study PT lines and trips could use a good data elimination and outlier detection methodology for the sake of more realistic results.

Smart card datasets are mostly used for inferring spatio-temporal passenger activities (ridership/flow) in the literature. On the other hand, these smart card datasets are provided by tap-on/tap-off based Automated Fare Collection Systems. It is easy to analyze passenger movements (origin-destination (OD) estimation, travel pattern mining, trip purpose detection, route choice modeling) in the entire transit network by using this kind of datasets. Since the tap on/tap off based smart card systems provide twice as much data according to the passive smart card (only tap on) systems in terms of data density, the bus trajectory can be extracted more precisely.

Since many public transit administrations already having a database management system for another purpose, they implement their new Automated Fare Collection Systems or Automated Vehicle Location System to this current database management system used for another purpose. Actually, this is for the ease of management with the current DBMS since the management of a new DBMS brings new costs. As a result, they are deprived of a database management system with a spatial extension like PostgreSQL. If this type of a DBMS is installed with a new AVLS and a geodatabase designed to store the network dataset and an LRS belong to the PT network of the city, all the analysis for public transit performance measures can be done quickly and efficiently. Moreover, the current study could be extended by using additional data. Data used for one-week is not sufficient to represent the real world and cannot be used to detect monthly and seasonal patterns.

REFERENCES

Abkowitz et al. (1978). *Transit service reliability.* United States. Urban Mass Transportation Administration.

Anderson, J. (2019). *Webinar: Evaluating Urban Arterial Reliability Performance Metrics.* Retrieved from https://pdxscholar.library.pdx.edu/: https://pdxscholar.library.pdx.edu/trec_webinar/42/

Bertini and Geneidy. (2003). Generating transit performance measures with archived data. *Transportation Research Record, 1841(1)*, 109-119.

Bigham and Kang. (2013). Building a highway linear referencing system from preexisting reference marker measurements for transportation data management. *URISA (Vol. 25, No. 1).*

Brown and Racca. ((2012). Study and Calculation of Travel Time Reliability Measures.

Çetinel, .. G. (2017). Evaluation of bus dwelling patterns using bus GPS data. *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, (pp. 867-871). Naples.

Chapleau et al. (2008). The ultimate survey for transit planning: Complete information with smart card data and GIS. *Proceedings of the 8th international conference on survey methods in transport: Harmonisation and data comparability*, (pp. 25-31).

Chapleau et al. (2011). Synthesizing AFC, APC, GPS and GIS data to generate performance and travel demand indicators for public transit. *90th Annual Meeting of the Transportation Research Board (CD-ROM).* Washington, DC.

Chu, K. K. A., & Chapleau, R. (2008). Enriching archived smart card transaction data for transit demand modeling. *Transportation research record, 2063(1)*, 63-72.

Curries and Mesbah. (2011). Exploring transit operations performance at a network level using AVL and new GIS visualization. *90th Annual Meeting of the Transportation Research Board.* Washington, DC.

Curtin et al. (2007). A comprehensive process for linear referencing. *URISA Journal, 19(2)*, 41-50.

Curtin et al. (2007). A comprehensive process for linear referencing. *URISA Journal*, 41-50.

Curtin et al. (2019). *DM-16-Linear Referencing*. Retrieved from gistbok.ucgis.org: https://gistbok.ucgis.org/bok-topics/linear-referencing

Curtin, K. T. (2020). *DM-16-Linear Referencing*. Retrieved from gistbok.ucgis.org: https://gistbok.ucgis.org/bok-topics/2020-quarter-01/linear-referencing

Danaher et al. (1999). *Transit Capacity and Quality of Service Manual, Prepared by Kittelson and Associates, Inc. for Transit Cooperative Research Program.* Transportation Research Board.

Dayan et al. (2014). Methodology to Map Routes from Truck Permit Database, Using a Linear Reference System and Network Analysis. *Transportation Research Record, 2460(1)*, 47-57.

Devillaine et al. (2012). Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record, 2276(1)*, 48-55.

Esri. (2019). *www.esri.com*. Retrieved from Esri.

Faroqi et al. (2018). Applications of transit smart cards beyond a fare collection tool: a literature review. *Advances in Transportation Studies, 45*.

Fourie, P. J. (2014). Transforming bus smart card transaction data into vehicle trajectories. *Arbeitsberichte Verkehrs-und Raumplanung, 986*.

Gökaşar and Çetinel. (2017). Evaluation of Bus Dwelling Patterns Using Bus GPS Data. *In 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, (pp. 867-871).

Hickman, M. (2002). Robust passenger itinerary planning using transit AVL data. In Proceedings. *The IEEE 5th International Conference on Intelligent Transportation Systems*, (pp. 840-845).

Kittelson and Associates. (2003). *A guidebook for developing a transit performance-measurement system (Vol. 88).* Washington, DC: Transportation Research Board.

Liao and Liu. (2010). Development of data-processing framework for transit performance analysis. *Transportation Research Record 2143*, 34-43.

Llanes, K. R. (2017). Bus Network Analysis and Monitoring (Doctoral dissertation, PUC-Rio).

Lu, H. (2007). Network smart card review and analysis. *Computer Networks 51*, 2234–2248.

Lyman and Bertini. (2008). Using travel time reliability measures to improve regional transportation planning and operations. *Transportation Research Record, 2046(1)*, 1-10.

Ma and Wang. (2014). Development of a data-driven platform for transit performance measures using smart card and GPS data. *Journal of Transportation Engineering, 140(12), 04014063*.

OORT, N. V. (2011). *Service Reliability and Urban Public Transport Design.*

Pelletier et al. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies, 19(4)*, 557-568.

Pu, W. (2011). Analytic relationships between travel time reliability measures. *Transportation Research Record, 2254(1)*, 122-130.

Pu, W. (2011). Analytic relationships between travel time reliability measures. *Transportation Research Record, 2254(1),* , 122-130.

Rakha et al. (2010). Trip travel-time reliability: issues and proposed solutions. *Journal of Intelligent Transportation Systems, 14(4)*, 232-250.

Rieser, M. (2010). Adding transit to an agent-based transportation simulation: concepts and implementation.

Robinson et al. (2014). Methods for pre-processing smartcard data to improve data quality. *Transportation Research Part C: Emerging Technologies, 49*, 43-58.

Shelfer and Procaccino. (2002). Smart card evolution. *Communications of the ACM, 45(7)*, 83-88.

Singh and Gupta. (2015). Recent trends in intelligent transportation systems: a review. *Journal of Transport Literature*, 30-34.

Song, Y. (2018). Modeling Spatial-Temporal Patterns of Bus Delays at and between Stops Using AVL and APC Data and Semi-Markov Techniques. *In CICTP 2018: Intelligence, Connectivity, and Mobility (pp. 666-675). Reston, VA: American Society of Civil Engineer*.

Trépanier et al. (2009). Calculation of transit performance measures using smartcard data. *Journal of Public Transportation, 12(1)*, 5.

Trépanier et al. (2015). Calculation of Transit Performance Measures Using Smartcard Data. *Journal of Public Transportation*, 79-96.

United Nations. (2018). *2018 Revision of World Urbanization Prospects.*

USDOT. (2006). *Travel Time Reliability: Making It There On Time, All The Time.* Washington, D.C.: , U.S. Department of Transportation, Federal Highway; Texas A&M Transportation Institute with Cambridge Systematics, Inc.

Uyan et al. (2017). Determination of Transportation Networks Base on the Optimal Public Transportation Policy Using Spatial and Network Analysis Methods: a Case of the Konya, Turkey. International Journal of Engineering and. *International Journal of Engineering and Geosciences 2(1),*, 27-34.

Yenice, M. (2012). Konya kentinin planlama tarihi ve mekânsal gelişimi. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi 28-4*.

Zhang et al. (2018). Space–time visualization analysis of bus passenger big data in Beijing. *Cluster Computing, 21(1)*, 813-825.

APPENDICES

A-Supplement for the Methodology

A.1 Definitions and Abbreviations of GIS Tools and Concepts

Along with the tools that take place under the linear referencing tasks, many other common GIS tools have been used throughout this study. The definitions of these GIS tools are listed in Table A.1.1 and the concepts related to these tools are listed in Table A.1.2.

Table A.1.1 Definition of GIS tools

GIS Tools	Definitions
Add Field	It is used to add a new attribute column.
Dissolve	It is used to combine or separate objects.
Create Route	It is used to produce polyline m objects.
Locate Features Along Route	It is used to generate kilometers.
Make Route Event Layer	It is used for the mapping of the generated kilometers.
Sort	It is used to put attributes in a specific order.
Field Calculator	It is used to calculate field values.
Calculate End Time	It is used to generate from-to timestamps and kilometers
Feature Class to Feature Class	It is used to export data and change the table structure
Select Layer by Attribute	It is used to select target data by attribute values.
Select Layer by Location	It is used to select target data by location specifications.
Add Join	It is used to joins a layer to another layer or table based on a common field.
Join Field	It is used to join one or more fields of a table to another table based on a common attribute field.
Summary Statistics	Calculates summary statistics for field(s) in a table.
Feature Class to Feature Class	Converts a feature class in the form of selected fields to another one.
Make feature Layer	Creates a temporary feature layer from an input feature class.
Add Field	Adds a new field to the feature class.
Remove Join	Removes a join established before.

Table A.1.2. Definition of GIS concepts

GIS Concepts	Definitions	Abbreviations
Feature Class	In ArcGIS, a collection of geographic features with the same geometry type (such as point, line, or polygon), the same attributes, and the same spatial reference.	-
Multipart Feature Class	A Feature Class defined as one feature since it references one set of attributes.	-
Feature Dataset	In ArcGIS, a collection of feature classes stored together that share the same spatial reference.	-
Polyline Feature	A digital map feature that represents a place or thing that has length but not area at a given scale.	P
Polyline M Feature Class	A polyline feature class has the ability to store m-values (measurement values) or distance from a starting point along a given line.	PM
Single Polyline Feature Class	A polyline feature class having one part associated with a single record in the attribute table.	SPL
Multiple Polyline Feature Class	A polyline feature class having more than one part associated with related records in the attribute table.	MPL
Single Polyline M Type Of Feature Class	A polyline feature having one part associated with a single record in the attribute table and additionally can store m-values.	SPLM
Multiple Polyline M Type Of Feature Class	A polyline feature having more than one part associated with related records in the attribute table and additionally can store m-values.	MPLM
Point Feature Class	A map feature that has neither length nor area at a given scale, such as a city on a world map or a building on a city map.	MP
Point M Feature Class	A point feature class can store m-values.	MPM
Event Table	A data source containing location information in tabular format (called events) that is used to create a spatial dataset. For example, an event table might contain x,y coordinates, measures or routes.	-
Event Layer	A layer created from an event table	-
Field	A column in a table or attribute table of a feature class that stores the values for a single attribute.	-
Projected coordinate system	A reference system used to locate x, y, and z positions of point, line, and area features in two or three dimensions. A projected coordinate system is defined by a geographic coordinate system, a map projection, any parameters needed by the map projection, and a linear unit of measure.	-

A.2 Bus Trajectory GPS Data Structure

The given BTD was in a MySql dump file format. BTD basically contains Sequence ID of each record in sequential order, ID of the Vehicle driven in a particular trip, timestamp of each record and location information in terms of latitude and longitude. The Table A.2.1 represents all the attributes and descriptions of raw BTD. Along with raw BTD, a bus direction table is provided. This table contains several attributes but necessary ones are listed in the below Table A.2.2. The resolution of the bus GPS is approximately 5 seconds.

Table A.2.1 Raw bus trajectory GPS table structure

ATTRIBUTE	FIELD NAME	Attribute_Description	Notes
SEQUENCE_ID	SID	Sequence ID	GPS data order
PACKET_ID	PID	Packet ID	GPS data packets collected within specific minutes
VEH_ID	VID	Bus Vehicle ID	The identification number of the vehicle on the trip.
TIMESTAMP	TS	Timestamp	Provided in yyyy-mm-dd hh:mm:ss format
SATELLITE_NUMBER		Satellite Number	Number of satellites received data (at least 3 satellites)
LATITUDE	Y	Latitude	Provided in decimal format
LONGITUDE	X	Longitude	Provided in decimal format
HEADING	H	Heading	Direction of the bus in degrees
SPEED	V	Speed (km/hour)	The bus speed in km / h
LINE_ID	LID	Service Line Idendification Number	Identification Number for a Service Line
DRIVER_NO	DID	Driver No	ID of Driver
INSERT_TIMESTAMP	ITS	Insert time of GPS	Database insertion time of the record.

Table A.2.2 Bus trajectory data direction table

Attribute/Field Name	Description
f1	Unique index value
SEGTID	Inter-Stop Segment ID
F_TS	Timestamp of the first record in the segment
L_TS	Timestamp of the last record in the segment
ShiftID	Shift ID including the vehicle id and the trip time interval
tripID	Trip ID
MV_DIR_F	Moving direction (from bus stop)
MV_DIR_L	Moving direction (to bus stop)

A.3 Smart Card Data Structure

Smart card data consists of BTD that are recorded once for each smart card reading during a boarding. This means that Smart card Automated Fare Collection Systems (SAFCS) of the public transit network is an only tap-in type of system. Alighting (tap-out) information is not provided is the smart card dataset. These records are provided in CSV file format daily for a ULID. Attributes of daily smart card data are listed below in Table A.3.1.

Table A.3.1. Raw smart card data (SCD) attribute list

ATTRIBUTE	FIELD NAME	Attribute_Description	Notes
VEH_ID	VID	Vehicle ID	Vehicle ID shows which vehicle is on trip.
TXN_TYPE	TT	Type of Transaction from Smart Card	Txn_type is seen as ABILET, BILET, GBILET, KBILET, TBILET, GLBILET, GDBILET, BKM in smart card data
CARD_ID	CID	Smart Card Serial Number	Character length = 32 Total unique card id = 551470
CARD_TYPE	CT	Smart Card Ticket Type (There are 40 types)	Ticket types are shown in numerical values as 0, 1, 2, 8, 9,10..
LATITUDE	Y	Latitude	Provided in decimal format
LONGITUDE	X	Longitude	Provided in decimal format
TIMESTAMP	TS	Date and Time of Smart Card Use	Provided in yyyy-mm-dd hh:mm:ss format
LINE_ID	LID	Service Line Idendification Number	Identification Number for a Service
SUB_LINE_ID	SLID	Sub_Line Idendification Number	The number showing the small change in routes under the same LID (i.e: 36-0, 36-1 in a particular place going from an upper street)
TYPES	T	Public Transport types	Types show public transportation modes. Tram=0, bus=1
ShiftID	ShiftID	Shift ID	Shift ID in terms of ULID and VID
STripID	STripID	Trip ID	Trip ID in terms of integer numbers
AsgnStopOrd	AsgnStopOrd	Assigned Stop Order Value	Assigned bus stop order values to the records withing 50 m buffer of a bus stop.
AsgnStat	AsgnStat	Status of the vehicle	Outside of the 50m bus stop buffer:1, Inside of the 50m bus stop buffer:2
AsgnStopID	AsgnStopID	Assigned Stop ID	Assigned bus stop id values to the records withing 50 m buffer of a bus stop.
ULID_corr	ULID	Unique Line Idendification Number	Unique LID created as 10*LID+SLID (i.e 36-0 and 36-1 corresponded to 360 and 361) total unique ULID = 395
INDEX	INDEX	INDEX	Daily record index

A.4 Producing Bus Route Feature Class
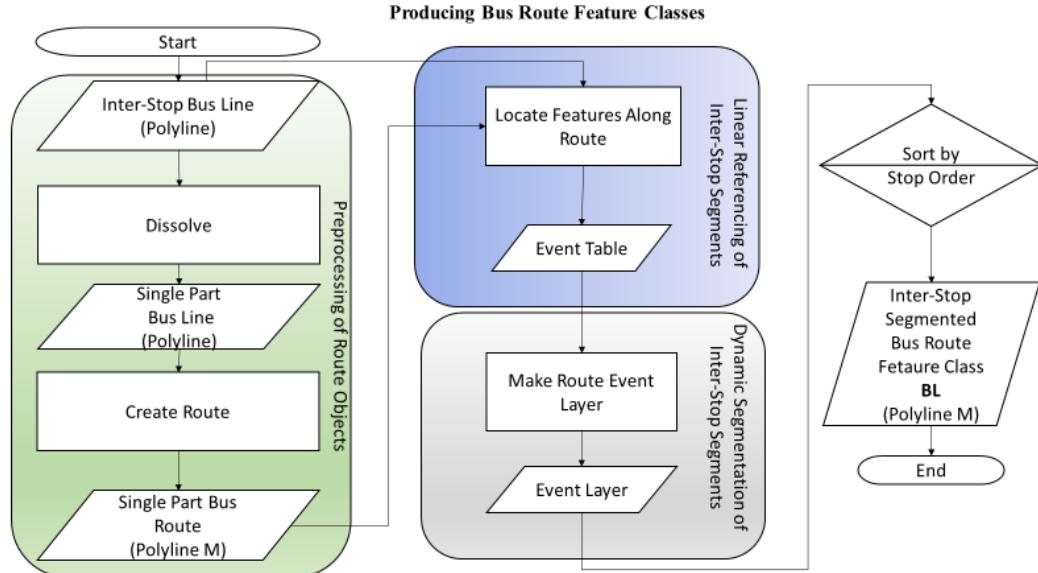
Figure A.4.1 Flowchart of the route production

Table A.4.1 Pseudocodes for the route production

For a given Unique LineID *ULID*, Get the bus line with a Multiple Polyline Feature Class **MPL,**
　　　{ULID, DIR_CHG_C, S_ORD_1, ... }
Get the bus line with a Multiple Polyline Feature Class **MPLM**
　　　{ULID, MEAS, DIR_CHG_C, S_ORD_1, ... }
Get the bus line with a Single Polyline Feature Class **SPL,**
　　　{OBJECTID, ULID}
Get the bus line with a Single Polyline M Feature Class **SPLM**
　　　{OBJECTID, ULID, SHAPE_LENGTH}

FOR ∀ **MPL**
{
　　　CALL Dissolve
　　　　　[**MPL** to **SPL** by Field: *ULID*]
　　　CALL Create_Route
　　　　　[**SPL** to **SPLM** by Field: *ULID*, Measure Source: LENGTH]
　　　CALL Locate_Features_Along_Route
　　　　　[**MPL** to **SPLM** Event Table by Route_Identifier Field: *ULID*, Search_Radius: 0 m, From-Measure
　　　　　Field: *FMEAS*, To-Measure Field: *TMEAS*, Check: Including all fields from input]
　　　CALL Make_Route_Event_Layer
　　　　　[**SPLM Event Table** to **MPLM Event Layer** by Route_Identifier Field: *ULID*, Event Type: Line,
　　　　　From-Measure Field: *FMEAS*, To-Measure Field: *TMEAS*]
　　　CALL Sort
　　　　　[**MPLM** Event Layer, Sort_Field: *S_ORD_1,* Output: **MPLM** Feature Class]
}

A.5 Generation of St Km for Bus Stops
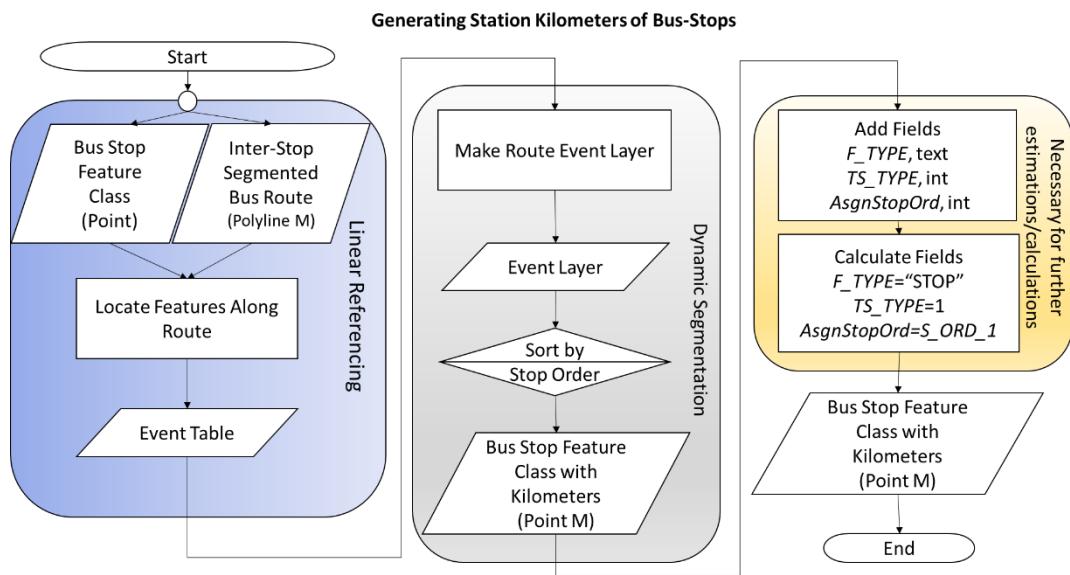
Figure A.5.1. Flowchart for generating St Km of bus-stops

Table A.5.1. Pseudocodes for generating St Km of bus-stops

For a given Unique LineID *ULID*,
Get the bus stops with a Multiple Point Feature Class **MP**
{S_NO_1, S_ORD_1, ULID, DIR_CHG_C,… }
Get the bus stops with a Multiple Point M Feature Class **MPM**
{MEAS, S_NO_1, S_ORD_1, ULID, DIR_CHG_C, AsgnStopOrd,… }
FOR ⊂ **MP**
{
      CALL Locate_Features_Along_Route
            [**MP** to **MPM** Event Table by Route_Identifier Field: *ULID*, Search_Radius: 0 m, Measure
            Field: *MEAS*, Check: Including all fields from input & Keep only the closest route location]
      CALL Make_Route_Event_Layer
            [**MPM Event Table** to **MPM Event Layer** by Route_Identifier Field: *ULID*, Event Type:
            Point, Measure Field: *MEAS*]
      CALL Sort
            [**MPM** Event Layer, Sort_Field: *S_ORD_1,* Output: **MPM** Feature Class]
      CALL Add_Field [*F_TYPE*, Text]
      CALL Add_Field [*TS_TYPE*, short]
      CALL Add_Field [*AsgnStopOrd*, int]
      CALL Field_Calculator
            CALCULATE [*F_TYPE*="STOP"]
      CALL Field_Calculator
            CALCULATE [*TS_TYPE*=1]
      CALL Field_Calculator
      CALCULATE [*AsgnStopOrd =S_ORD_1*]
}

A.6 Data Preprocessing

The provided smart card and bus trajectory datasets must be prepared for further processes. Due to the large volume of these datasets and the difficulties of getting this volume of dataset ready for processing, a DBMS environment has been used in addition to the GIS environment.

A.6.1 Preprocessing for Bus Trajectory Dataset in DBMS

Unlike the smart card dataset, the bus trajectory dataset consists of two separate datasets. The first dataset is the trajectory data obtained from the GPS device of the vehicle. This dataset basically includes the latitude/longitude and timestamps of the bus. Data files were in the MySQL dump file format (*.sql) (Figure A.6.1a). The moving direction of the buses in terms of bus-stop order values was stored in another CSV file. In this second table, basically, timestamp intervals of shift/trip pairs that belong to vehicles are stored (Figure A.6.1b). Therefore these values in the second table regarding moving directions must be inserted into the first table by matching the correct GPS records. The methodology consists of the following steps;

➢ Design a table with suitable field names and data types in PostgreSQL DBMS to import BTD data from MySQL dump files (PostgreSQL)
➢ Import BTD data to the designed table in the DBMS. (PostgreSQL)
➢ Import BTD direction data to the DBMS. (PostgreSQL)
➢ Match and insert directions with shift/trip values to the BTD table from the second table. (PostgreSQL)
➢ Export the resulting tables, separated by shift/trip pairs, as CSV files to import into the ArcGIS Desktop Software.

(a) Initial raw BTD sample

SID	PID	VID	TS	SATNO	Y	X	H	V	LID	DID	ITS
274070433	190	353	2018-10-30 22:45:01	11	37.86598300	32.49218300	84	47	83	1294	2018-10-30 22:44:53
274070434	1057	448	2018-10-30 22:44:51	11	37.82855000	32.46335000	324	19	11	479	2018-10-30 22:44:53
274070435	1166	365	2018-10-30 22:44:52	11	37.86871700	32.54470000	148	12	99	354	2018-10-30 22:44:53
274070436	3915	203	2018-10-30 22:44:51	11	37.79610000	32.47761700	190	62	21	113	2018-10-30 22:44:53
274070437	514	758	2018-10-30 22:44:52	11	37.85088300	32.42106700	197	51	2	568	2018-10-30 22:44:53

(b) BTD direction table sample

SEGTID	F_TS	L_TS	ShiftID	DIRECTION	MV_DIR_F	MV_DIR_L	tripID
ULID443_1288d_2198d_100_101	10/30/2018 18:17	10/30/2018 18:18	443_211_224_2018-10-30 16:18:03_2018-10-30 19:00:37	1	100	101	1
ULID443_1288d_2198d_100_101	10/30/2018 15:57	10/30/2018 15:59	443_246_261_2018-10-30 14:10:06_2018-10-30 19:56:06	1	100	101	1
ULID443_1288d_2198d_100_101	10/30/2018 19:31	10/30/2018 19:32	443_246_261_2018-10-30 14:10:06_2018-10-30 19:56:06	1	100	101	3
ULID443_1288d_2198d_100_101	10/30/2018 09:07	10/30/2018 09:08	443_608_311_2018-10-30 07:09:49_2018-10-30 09:41:37	1	100	101	1
ULID443_1288d_2198d_100_101	10/30/2018 20:21	10/30/2018 20:23	443_632_235_2018-10-30 19:04:20_2018-10-30 20:45:48	1	100	101	1
ULID443_2198d_1287d_101_102	10/30/2018 18:18	10/30/2018 18:19	443_211_224_2018-10-30 16:18:03_2018-10-30 19:00:37	1	101	102	1
ULID443_2198d_1287d_101_102	10/30/2018 15:58	10/30/2018 15:59	443_246_261_2018-10-30 14:10:06_2018-10-30 19:56:06	1	101	102	1
ULID443_2198d_1287d_101_102	10/30/2018 19:32	10/30/2018 19:32	443_246_261_2018-10-30 14:10:06_2018-10-30 19:56:06	1	101	102	3
ULID443_2198d_1287d_101_102	10/30/2018 09:08	10/30/2018 09:09	443_608_311_2018-10-30 07:09:49_2018-10-30 09:41:37	1	101	102	1
ULID443_2198d_1287d_101_102	10/30/2018 20:22	10/30/2018 20:23	443_632_235_2018-10-30 19:04:20_2018-10-30 20:45:48	1	101	102	1
ULID443_1287d_1286d_102_103	10/30/2018 18:18	10/30/2018 18:20	443_211_224_2018-10-30 16:18:03_2018-10-30 19:00:37	1	102	103	1
ULID443_1287d_1286d_102_103	10/30/2018 15:59	10/30/2018 16:00	443_246_261_2018-10-30 14:10:06_2018-10-30 19:56:06	1	102	103	1

Figure A.6.1 Initial samples from BTD dataset (a) and direction table (b)

Fields	Indexes	Foreign Keys	Uniques	Ch

Name	Type	Length	Decimal
SID	int4	32	0
PID	int4	32	0
VID	int2	16	0
TS	timestamp	6	0
SATNO	int2	16	0
Y	numeric	10	8
X	numeric	11	8
H	int2	16	0
V	int2	16	0
LID	int2	16	0
DID	int4	32	0
ITS	timestamp	6	0
DAY_ID	int4	32	0
ULID	int4	32	0
MV_FR	int4	32	0
MV_TO	int4	32	0
SHFTID	varchar	0	0
SHFT_INTRVL	varchar	0	0
TRPID	int4	32	0
SEGTID	varchar	0	0

Figure A.6.2 Resulting BTD fields and data types after preprocessing

SID	PID	VID	TS	SATNO	Y	X	H	V	LID	DID	ITS
270678505	55	608	2018-10-30 07:20:48	13	37.86898300	32.53993300	269	36	44	311	2018-10-30 07:20:46
270679305	57	608	2018-10-30 07:20:58	13	37.86896700	32.53858300	271	44	44	311	2018-10-30 07:20:55
270678887	56	608	2018-10-30 07:20:53	13	37.86898300	32.53930000	272	40	44	311	2018-10-30 07:20:5
270681741	59	608	2018-10-30 07:21:08	13	37.86898300	32.53705000	273	50	44	311	2018-10-30 07:21:23
270679827	58	608	2018-10-30 07:21:03	13	37.86896700	32.53783300	271	47	44	311	2018-10-30 07:21:
270681796	60	608	2018-10-30 07:21:13	13	37.86901700	32.53621700	271	50	44	311	2018-10-30 07:21:24
270681909	62	608	2018-10-30 07:21:23	13	37.86906700	32.53475000	272	33	44	311	2018-10-30 07:21:25
270683585	67	608	2018-10-30 07:21:47	13	37.86913300	32.53330000	276	34	44	311	2018-10-30 07:21:45

DAY_ID	ULID	MV_FR	MV_TO	SHFTID	SHFT_INTRVL	TRPID	SEGTID
480983	443	1	2	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1022d_1021d_1_2
481728	443	1	2	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1022d_1021d_1_2
345	443	1	2	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1022d_1021d_1_2
4 7	443	2	3	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1021d_1020d_2_3
211	443	1	2	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1022d_1021d_1_2
484039	443	2	3	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1021d_1020d_2_3
484146	443	2	3	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1021d_1020d_2_3
485705	443	3	4	443_608_311	2018-10-30 07:09:49_ 2018-10-30 09:41:37	1	ULID443_1020d_1019d_3_4

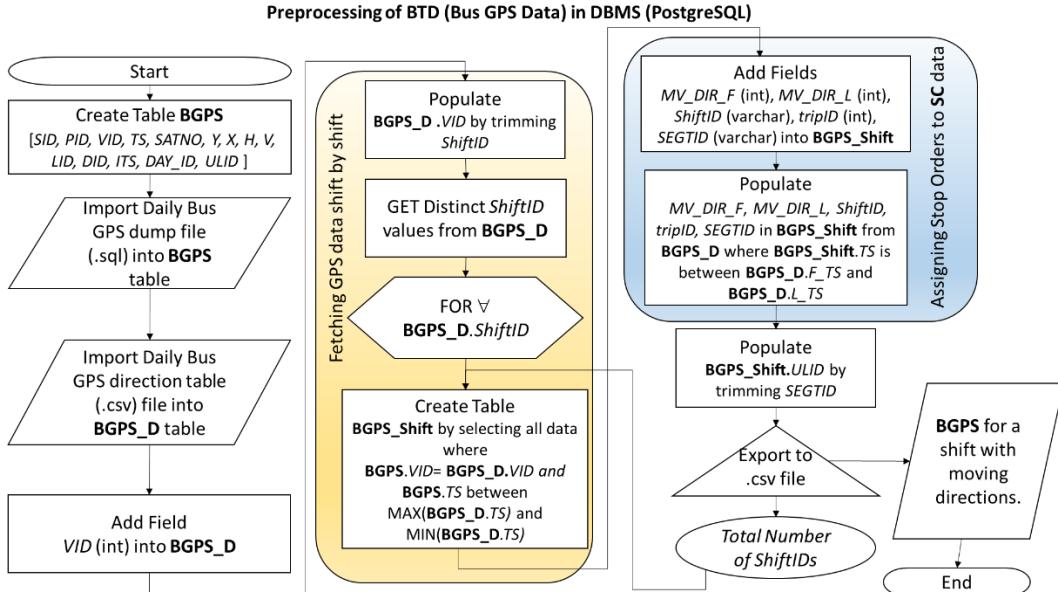Figure A.6.3 Resulting BTD with directions dataset sample

Figure A.6.4 Flowchart for the preprocessing of bus trajectory data in DBMS

Table A.6.1 Pseudo-codes for the Preprocessing of Bus Trajectory Dataset in DBMS

For a given day d, get the BTD dataset **BGPS** { *SEQUENCE_ID, VEHICLE_NO, DATE_TIME, ... }*
For a given day d, get the BTD direction dataset **BGPS_D** { *F_TS, L_TS, ShiftID, tripID, MV_DIR_F, MV_DIR_L, SEGTID, ... }*
FOR ∀ **BGPS**
{
      CREATE Table **BGPS** in PostgreSQL DBMS
          [*SID (int), PID (int), VID (int), TS (timestamp), SATNO (int), Y (numeric), X (numeric) , H (int), V (int) , LID (int) , DID (int) , ITS (timestamp) , DAY_ID (int) , ULID (int)*]
      IMPORT **BGPS** to the **BGPS** Table in PostgreSQL DBMS
}
IMPORT **BGPS_D** to the PostgreSQL DBMS
ADD FIELD *VID* (int)
POPULATE *VID* by trimming *ShiftID*
GET Distinct *ShiftID* values from **BGPS_D**
FOR ∀ **BGPS_D**.*ShiftID*
{
      CREATE Table **BGPS_Shift** by selecting all data where **BGPS**.*TS* between MAX[**BGPS_D**.*TS*] and MIN[**BGPS_D**.*TS*] and **BGPS**.*VID*= **BGPS_D**.*VID*
      ADD FIELDS *MV_DIR_F* (int), *MV_DIR_L* (int), *ShiftID* (varchar), *tripID* (int), *SEGTID* (varchar) into **BGPS_Shift**
      POPULATE *MV_DIR_F*, *MV_DIR_L, ShiftID, tripID, SEGTID* Fields in **BGPS_Shift** from **BGPS_D** where **BGPS_Shift**.*TS* is between **BGPS_D**.*F_TS* and **BGPS_D**.*L_TS*
      POPULATE **BGPS_Shift.***ULID* by trimming *SEGTID*
      EXPORT **BGPS_Shift** to .csv file format
}

A.6.2 Preprocessing of Smart Card Dataset in DBMS

The provided smart card dataset consists of BTD recordings during smart card readings in CSV files separated by days and bus lines. Along with latitude/longitude and timestamps, assigned stop order values are included in the smart card dataset. These files include the columns in Figure A.6.5 The field names and data types are configured in the Postgresql Database Management Software. In addition to these columns, a unique line id (*ULID*) and combined type of shift and trip id (*Ulid_Vid_trip*) fields are added to the dataset to be able to group the data easily in other processes. The design of a smart card table configured in Postgresql is like below in Figure A.6.6.

129

Figure A.6.5 Initial smart card dataset sample

Figure A.6.6 Smart card field names and data types

After these processes, smart card datasets are exported as CSV files day by day in order to be imported into the ArcGIS Desktop Software.

The flowchart for the preprocessing of Smart Card Dataset in DBMS is as shown in Figure A.6.7 and the pseudo-codes for these processes explained in this section are given in Table A.6.2.
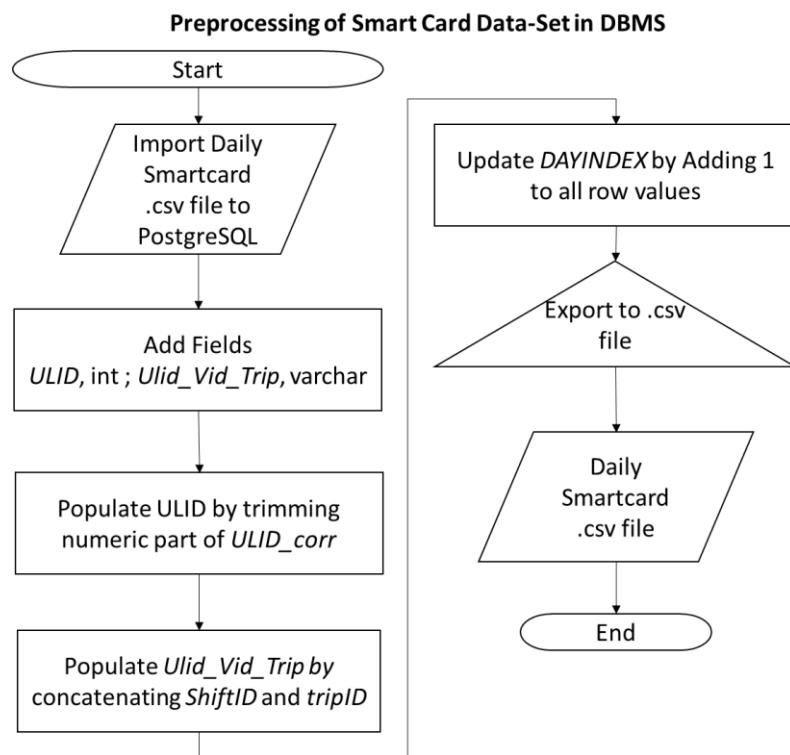
Figure A.6.7 Flowchart for the preprocessing of smart card dataset in DBMS

Table A.6.2 Pseudocodes for Preprocessing of Smart Card Dataset in DBMS

For a given *ULID* and day d , GET smart card dataset **SC** { *VID, CID, ULID_corr, TS, AsgnStopOrd, ShiftID, STripID,... }*
FOR ∀ **SC**
{
       IMPORT **SC** in .csv file format to PostgreSQL DBMS
       ADD FIELDS *ULID (int), Ulid_Vid_Trip (varchar)*
       POPULATE **SC**.*ULID* by trimming numeric part of **SC**.*ULID_corr*
       POPULATE **SC**.*Ulid_Vid_Trip* by concatenating **SC**.*ShiftID* and **SC**.*tripID*
       UPDATE **SC**.*DAYINDEX* by Adding 1 to all row values
       EXPORT to .csv file format
}

A.6.3 GIS Preprocessing of BTD and SCD

The aim of this part is to transfer the necessary "identifier" fields from bus-stop and bus route feature classes to the smart card and bus trajectory feature classes to conduct further Linear Referencing processes. The Linear Referencing processes will be based on these common fields in both smart card or BTD feature class and route feature class.

There are very minor differences between GIS preprocessing of BTD and SCD. While assigning directions to the datasets, Assigned Bus Stop Order (AsgnStopOrd) field is used in the smart card dataset. On the other hand, moving directions (MV_DIR_F) in terms of the first stop order value of the inter-stop segment is used in bus trajectory dataset,

Summary of the GIS preprocessing workflow is as follow:

Step 1: Join the field, including direction values, to the smart card or BTD feature class from the Bus-Stop feature class.

Step2: Conduct Spatial Join in separate directions to transfer bus-stop order values from route feature class to the smart card and BTD feature classes.

Step 3: Merge resulting feature classes of smart card or BTD data generated in different directions

GIS Preprocessing of Smart Card Dataset Page 1

Start

Bus Stop Feature Class with Kilometers **BS** (Point M)

Daily Smartcard Feature Class **SC** (Point)

Assigning directions to **SC** data

Join Field **BS**.*DIR_CHG_CH* into **SC** by matching **SC**.*AsgnStopOrd* and **BS**.*S_ORD_1*

SC with Directions
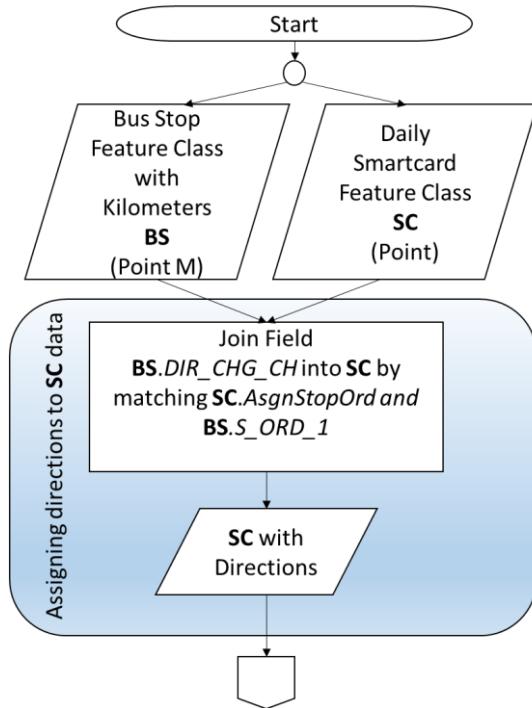
Figure A.6.8 Flowchart of the GIS Preprocessing of SCD Page 1

GIS Preprocessing of Smartcard Dataset Page 2

SC with Directions (Point)

Inter-Stop Segmented Bus Route **BL** (Polyline M)

Assigning Stop Order to **SC** data in Direction 0

Assigning directions to **SC** data in Direction 1

Select Layer by Attribute **SC**.*DIR_CHG_CH*=0

Select Layer by Attribute **BL**.*DIR_CHG_CH*=0

Select Layer by Attribute **SC**.*DIR_CHG_CH*=1

Select Layer by Attribute **BL**.*DIR_CHG_CH*=1

Spatial Join Direction 0

Spatial Join Direction 1

Merge

SC with Stop Orders (Point)

End

Figure A.6.9 Flowchart of the GIS Preprocessing of SCD Page 2

GIS Preprocessing of BTD (Bus GPS Dataset) Page 1

Start

Bus Stop Feature Class with Kilometers **BS** (Point M)

A Shift of Bus GPS Feature Class **BGPS_Shift** (Point)

Assigning directions to **BGPS** data

Join Field **BS**.*DIR_CHG_CH* into **BGPS_Shift** by matching **BGPS_Shift**.*MV_DIR_F* and **BS**.*S_ORD_1*

BGPS_Shift with Directions
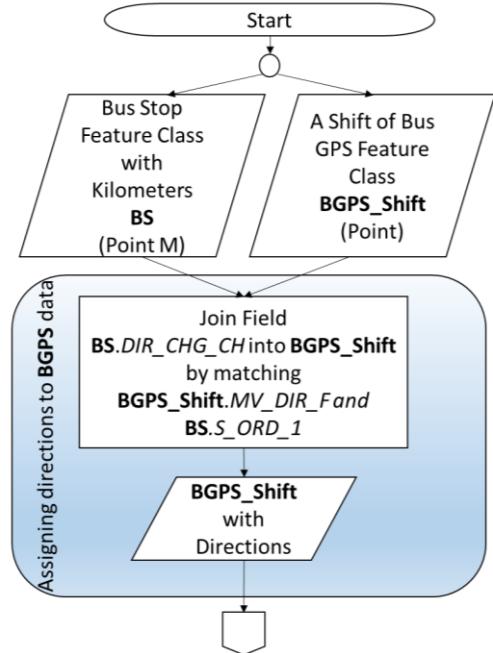
Figure A.6.10 Flowchart of the GIS Preprocessing of Bus Trajectory Data Page 1

GIS Preprocessing of BTD (Bus GPS Dataset) Page 2

BGPS_Shift with Directions (Point)

Inter-Stop Bus Route **BL** (Polyline M)

Select Layer by Attribute **BGPS_Shift**.*DIR_CHG_CH*=0

Select Layer by Attribute **BL**.*DIR_CHG_CH*=0

Select Layer by Attribute **BGPS_Shift**.*DIR_CHG_CH*=1

Select Layer by Attribute **BL**.*DIR_CHG_CH*=1

Assigning Stop Order to **BGPS** data in Direction 0

Spatial Join Direction 0

Spatial Join Direction 1

Assigning Stop Order to **BGPS** data in Direction 1

Merge

BGPS_Shift with Stop Orders (Point)

End

Figure A.6.11 Flowchart of the GIS Preprocessing of Bus Trajectory Data Page 2

Table A.6.3 Pseudocodes for GIS Preprocessing of Smart Card Data

---

For a given *ULID* and day d, GET smart card .csv file **SC** { *VID, CID, ULID, TS, AsgnStopOrd, ShiftID, STripID, Ulid_Vid_Trip, ... }*
For a given *ULID*, GET Bus Stop feature class **BS** { *MEAS, S NO 1, S ORD 1, ULID, DIR CHG C, F TYPE, TS TYPE, AsgnStopOrd,...}*
For a given *ULID*, GET Bus Line **BL** { *ULID, FMEAS, TMEAS, DIR CHG C, S ORD 1, ... }*
CREATE DATASETS for **SC, BS** and **BL** in FILEGEODATABASE
SET Coordinate System [Projected:TUREF, TM:33]
IMPORT **BS** and **BL** into DATASETS
FOR ∀ **SC**
{
    IMPORT **SC** into DATASETS
    CALL JOIN FIELD
        BS.*DIR_CHG_CH* into **SC** by matching **SC**.*AsgnStopOrd and* **BS**.*S_ORD_1*
    CALL SELECT LAYER BY ATTRIBUTE
        BL.*DIR_CHG_CH*=0
    CALL SELECT LAYER BY ATTRIBUTE
        SC.*DIR_CHG_CH*=0
    CALL SPATIAL JOIN sj1
        BL to **SC** [1 to N, Closest in 50m]
    CALL SELECT LAYER BY ATTRIBUTE
        BL.*DIR_CHG_CH*=1
    CALL SELECT LAYER BY ATTRIBUTE
        SC.*DIR_CHG_CH*=1
    CALL SPATIAL JOIN sj2
        BL to **SC** [1 to N, Closest in 50m]
    CALL MERGE [sj1 to sj2, Output: **SC** with direction and stop orders]
}

Table A.6.4 Pseudocodes for GIS Preprocessing of Bus Trajectory Data

---

For a given *ULID*, day and shift, GET BTD .csv file **BGPS_Shift** { *SID, VID, ULID, TS, MV DIR F, MV DIR L, ShiftID, tripID, X, Y, ... }*
For a given *ULID*, GET Bus Stop feature class **BS** { *MEAS, S NO 1, S ORD 1, ULID, DIR CHG C, F TYPE, TS_TYPE,...}*
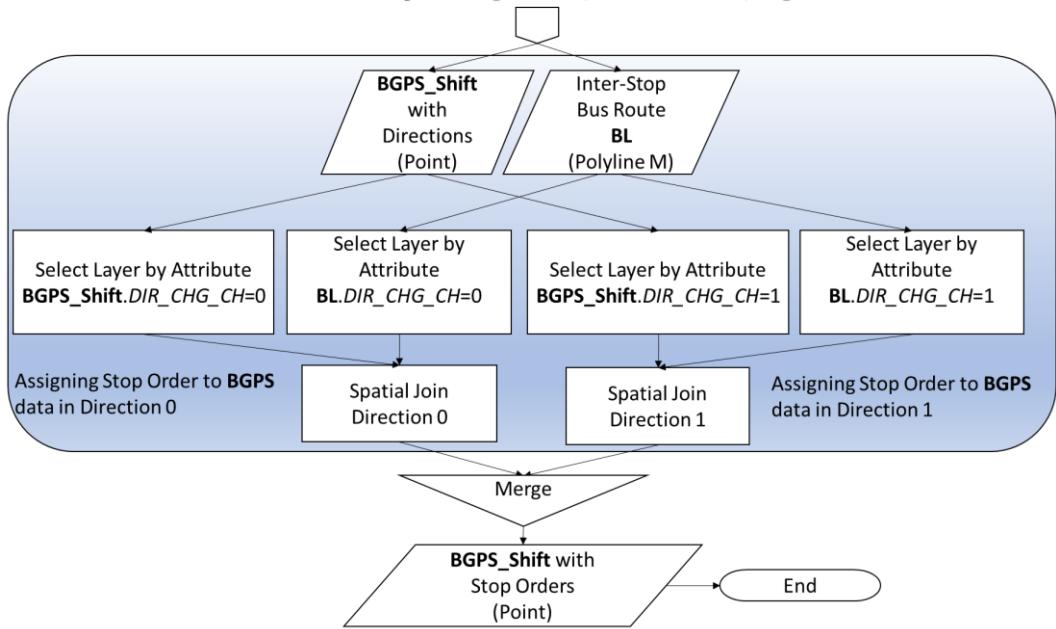For a given *ULID*, GET Bus Line **BL** { *ULID, FMEAS, TMEAS, DIR CHG C, S ORD 1, ... }*
CREATE DATASETS for **BGPS_Shift, BS** and **BL** in FILEGEODATABASE
SET Coordinate System [Projected:TUREF, TM:33]
IMPORT **BS** and **BL** into DATASETS
FOR ∀ **BGPS_Shift**
{
    IMPORT **BGPS_Shift** into DATASETS
    CALL JOIN FIELD
        BS.*DIR_CHG_CH* into **BGPS_Shift** by matching **BGPS_Shift**.*MV_DIR_F and* **BS**.*S_ORD_1*
    CALL SELECT LAYER BY ATTRIBUTE
        BL.*DIR_CHG_CH*=0
    CALL SELECT LAYER BY ATTRIBUTE
        BGPS_Shift.*DIR_CHG_CH*=0
    CALL SPATIAL JOIN sj1
        BL to **BGPS_Shift** [1 to N, Closest in 50m]
    CALL SELECT LAYER BY ATTRIBUTE
        BL.*DIR_CHG_CH*=1
    CALL SELECT LAYER BY ATTRIBUTE
        BGPS_Shift.*DIR_CHG_CH*=1
    CALL SPATIAL JOIN sj2
        BL to **BGPS_Shift** [1 to N, Closest in 50m]
    CALL MERGE [sj1 to sj2, Output: **BGPS_Shift** with direction and stop orders]
}

A.6.4 The Method Used for Preprocessing

A common route identifier is needed while Linear Referencing of SCD to the bus route. This common route identifier was the "Unique Line ID" (*ULID*) field in the linear referencing of bus stops. Since SCD is not snapped to the bus route, the "*ULID*" field is not enough to make linear referencing precisely. The bus routes do not always run straight. Outbound and inbound lines can be very close or even the same line. Route segments can be very close, or intersections of these segments may occur. Linear referencing of smart card data to the bus route should be conducted segment by segment in terms of bus stop intervals. But this is not enough for precise matching of smart card data to its related route segment. A common route segment identifier needs to be used for smart card data and route segment data. Stop order values of the route segments can directly identify a segment of the bus route. On the other hand, the smart card dataset is lack of this stop order field. The most identical field to the stop order field is the "Assigned Stop Order" (AsgnStopOrd) field in the smart card dataset. "Assigned Stop Order" field includes assigned bus-stop order values of the smart card record in the 50-meter buffer area of a bus-stop. Smart card records outside the 50-meter buffer area of bus-stops are assigned to the same stop order value of the preceding smart card record in terms of its timestamp.

Therefore, although it appears that the Assigned Stop Order field can be used for the matching of the smart card data with the right route segment, there is a complication. Assigned Stop Order values do not always match the route segment's Stop Order value. On the other hand, the BTD data includes two fields indicating the vehicle's moving direction in terms of bus stop order values. These fields are similar to that Assigned stop order field in the smart card dataset.

So we need bus-stop order values in the smart card dataset in order to project smart card records to the right route segments during the linear referencing process. Since bus-stop order values are stored in the route feature class, a spatial join process can match and copy these fields to the related smart card feature class. But there are many SCD scattered around complex route segments. On the other hand, smart card

136

records belong to inbound and outbound destinations can be missjoined. It is impossible to make a precise spatial join between smart card data and it's related route segment data. So this spatial join process should be conducted through separate destinations. Initially, direction values have to be transferred to the smart card or BTD feature class from the bus-stop feature class. A non-spatial join for the smart card feature class has to be conducted based on the Assigned Stop Order field and Stop Order field. This non-spatial join process has to be conducted based on the Moving Direction field and Stop Order field during the processing of the BTD data. After that, a spatial join has to be established between route segments and smart card or BTD feature classes in separate directions.

As a result, a combination of spatial and non-spatial join processes conducted through this process in order to obtain stop order fields in the smart card or BTD feature classes. Since there are many smart card feature classes to process, a geoprocessing tool needs to be prepared to facilitate all these works.

A.6.5 Prepared GIS Tool for GIS Preprocessing
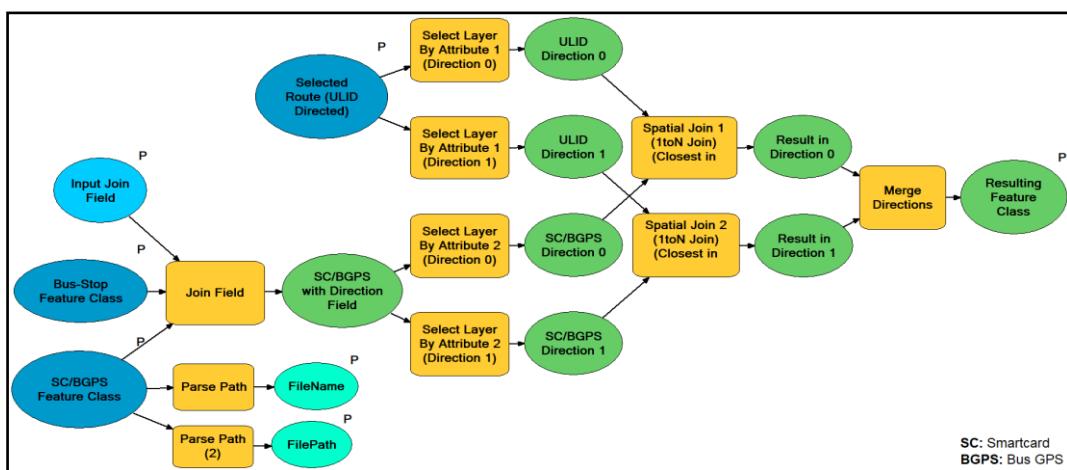
Figure A.6.12 Prepared GIS tool for transferring stop order values to the SCD and BTD in Model Builder

Therefore, these sequences of geoprocessing tools, above-mentioned, string together in a workflow by using Model builder application in ArcGIS desktop. (Figure A.6.12)
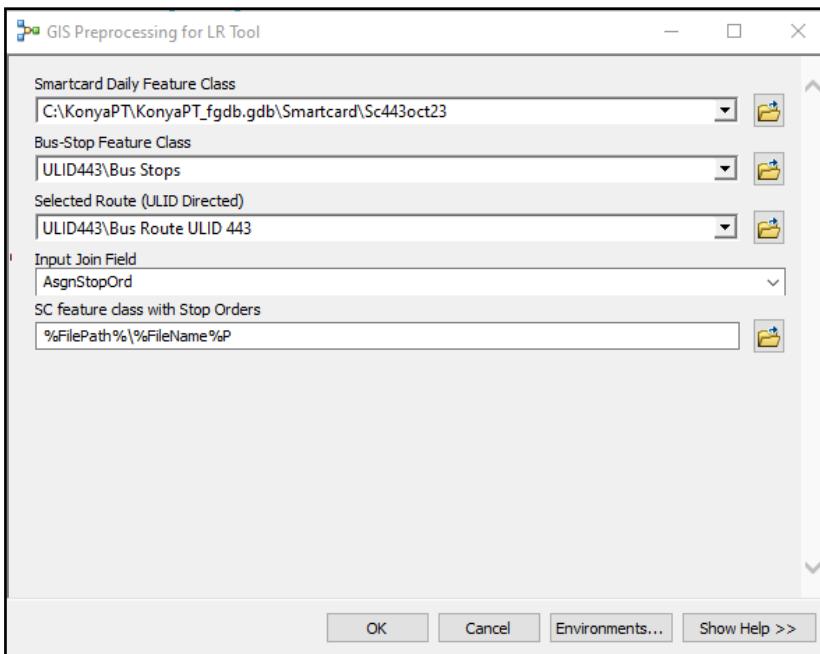
Figure A.6.13 The interface of the tool prepared for GIS preprocessing

After running the tool, an input smart card or bus stop feature class has to be selected along with output location and the field name that the non-spatial join will be based on. Join field have to be selectable since when using this tool for bus trajectory data, another field indicating "moving from-to" field needs to be selected since there is not an assigned stop order id for BTD data. (Figure A.6.13)

The resulting attribute table from a smart card and bus trajectory feature classes of a single trip has shown in Figure A.6.14a and Figure A.6.14b. Eventually, the smartcard and bus trajectory feature classes are ready for the linear referencing process. Some unnecessary fields resulting from previous operations have turned off due to page size. Some of these fields will be deleted to clean up the tables in further processes.

(a) Resulting SCD feature class attribute table with bus stop order values

(b) Resulting BTD (Bus GPS) feature class attribute table with bus stop order values

Figure A.6.14 Resulting smart card (a) and bus trajectory (b) attribute table with bus-stop order values (some fields turned off due to page size)

A.7 LR and DynSeg of BTD/SCD Feature Classes
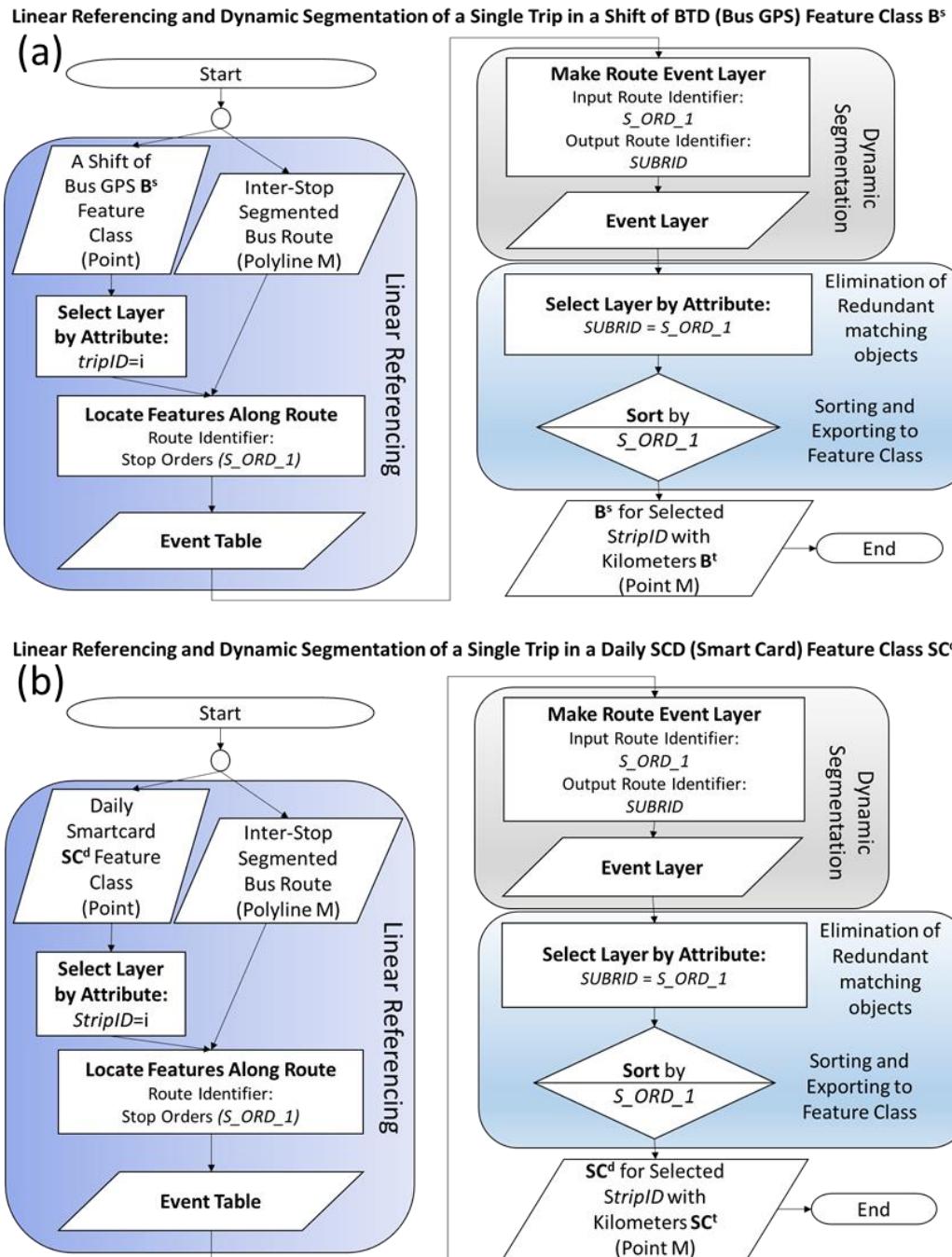
Figure A.7.1 LR & DynSeg of (a) Bus Trajectory Data and (b) Smart Card Data

Table A.7.1 Pseudocodes for LR and DynSeg of Bus Trajectory Data Feature Class

---

For a given *ULID*, d and ShiftID s, get the BTD feature class **B^s** { *SID, VID, ULID, TS, MV_DIR_F, MV_DIR_L, ShiftID, tripID, S_ORD_1,... }*

For a given *ULID, d, s and TripID t, get the* BTD feature class **B^t**

For a given *ULID*, get the Bus Line **BL** { *ULID, FMEAS, TMEAS, DIR_CHG_C, S_ORD_1,... }*

CREATE DATASETS for **B^s** and **BL** in FILEGEODATABASE

SET Coordinate System [Projected: TUREF, TM:33]

IMPORT **B^s** and **BL** into DATASETS

FOR ∀ **B^s**

{

    FOR ∀ **B^t_i**

    {

    CALL Locate_Features_Along_Route [**B^t_i** to **B^t_i _m Event Table** by Route_Identifier Field: *S_ORD_1*, Output_Route_Identifier: *SUBRID*, Search_Radius: 50 m, Measure Field: *MEAS*, Check: Including all fields from input]

    CALL Make_Route_Event_Layer [**B^t_i _m Event Table** to **B^t_i _m Event Layer** by Input_Route_Identifier Field: *S_ORD_1*, Output_Route_Identifier_Field: *SUBRID*, Event Type: Point, Measure Field: *MEAS*]

    CALL Select_Layer_by_Attribute [Input: **B^t_i _m Event Layer,** Expression: *SUBRID=S_ORD_1*]

    CALL Sort [**B^t_i_m Event Layer** by Sort_Field: *TS* Ascending*, Output Feature Class:* **B^t_i_m**]

    }

}

Table A.7.2 Pseudocodes for LR and DynSeg of Smart Card Data Feature Class

---

For a given *ULID* and day d, get the smartcard feature class **SC^d** { *VID, CID, ULID, TS, AsgnStopOrd, ShiftID, STripID, Ulid_Vid_Trip, S_ORD_1,... }*

For a given *ULID, d, ShiftID s, and STripID t, get the* smartcard feature class **SC^t**

For a given *ULID*, get the Bus Line **BL** { *ULID, FMEAS, TMEAS, DIR_CHG_C, S_ORD_1,... }*

CREATE DATASETS for **SC^d** and **BL** in FILEGEODATABASE

SET Coordinate System [Projected: TUREF, TM:33]

IMPORT **SC^d** and **BL** into DATASETS

FOR ∀ **SC^d**

{

    FOR ∀ **SC^t_i**

    {

    CALL Locate_Features_Along_Route [**SC^t_i** to **SC^t_i_m Event Table** by Route_Identifier Field: *S_ORD_1*, Output_Route_Identifier: *SUBRID*, Search_Radius: 50 m, Measure Field: *MEAS*, Check: Including all fields from input]

    CALL Make_Route_Event_Layer [**SC^t_i_m Event Table** to **SC^t_i_m Event Layer** by Input_Route_Identifier Field: *S_ORD_1*, Output_Route_Identifier_Field: *SUBRID*, Event Type: Point, Measure Field: *MEAS*]

    CALL Select_Layer_by_Attribute [Input: **SC^t_i_m Event Layer,** Expression: *SUBRID=S_ORD_1*]

    CALL Sort [**SC^t_i_m Event Layer** by Sort_Field: *TS* Ascending*, Output Feature Class:* **SC^t_i_m**]

    }

}

A.7.1 Prepared LR & DynSeg Tool in Model Builder

Since there are many smart card datasets and much more shift/trips in these datasets to process, an ArcGIS tool has prepared to facilitate by stringing the sequences of geoprocessing tools mentioned above by using Model Builder. (Figure A.7.2)
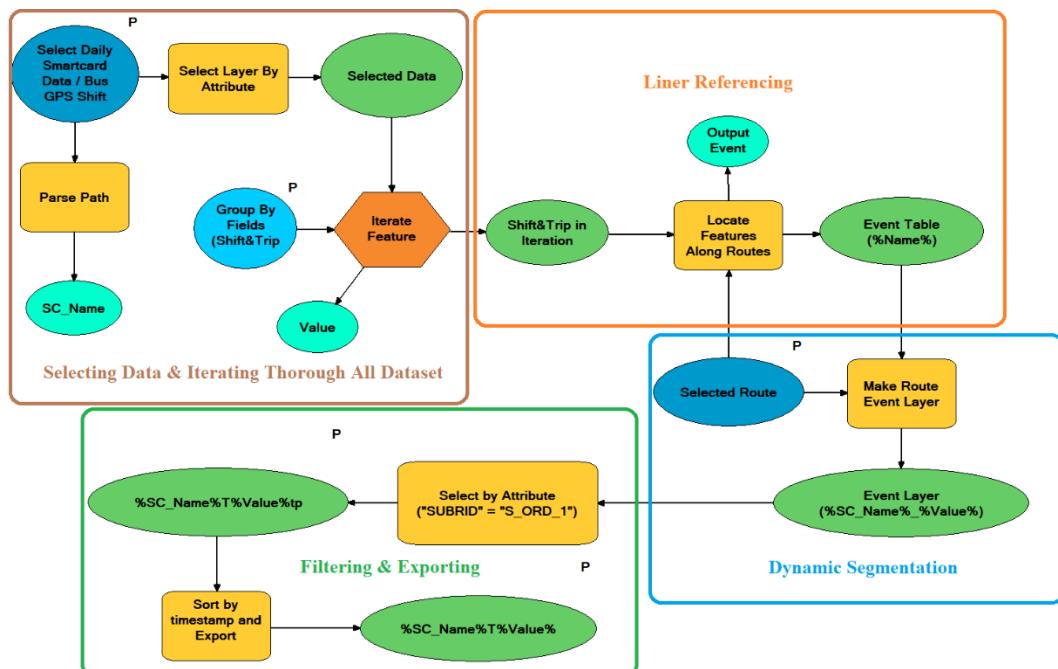
Figure A.7.2 Prepared LR and DynSeg Tool in Model Builder

This tool can be run for a given daily smartcard feature class or a given trip of BTD feature class. As we can see in Figure A.7.2, after LR and DynSeg tasks, a Filtering and Exporting process has been conducted in the generated tool. The filtering part is one of the critical subjects mentioned in the Key Points to Consider title under the Framework section of this chapter. By filtering the input (*S_ORD_1*) and output (*SUBRID*) route identifiers where they are equal to each other, LR and DynSeg processes have been completed without errors. There would be redundant and miss referenced points without this part of the process.
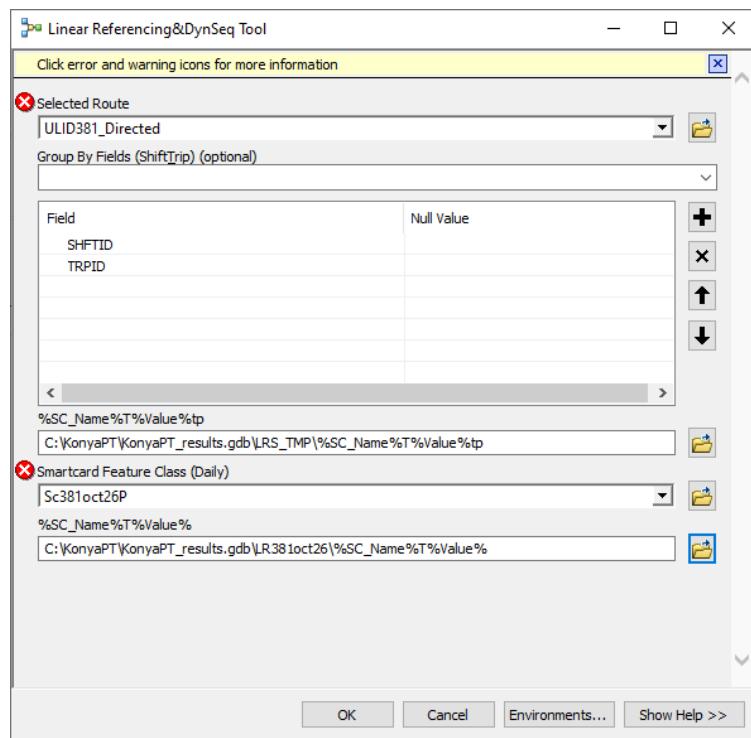
Figure A.7.3 The interface of LR and DynSeg Tool

After running this tool, along with the input route and smartcard feature class, output file locations and trip grouping parameters must be selected. This tool groups daily smartcard data for each shift/trip pairs and conduct the sequence of the process mentioned above. BTD data does not need to be grouped since input data is already a trip dataset. (Figure A.7.3)

At the end of the process, the resulting event layer is sorted according to the timestamp field and stored in the geodatabase by separate file names automatically.

A.8. Postprocessing of Data Fields

Since resulting attribute tables from the linear referencing process include some excess columns and additional fields that need to be inserted, a model has built and run for the cleaning, rearranging and inserting the fields to the dataset of all trips in a day. The same procedure is conducted for both smart card and bus trajectory feature classes. After postprocessing of data fields (cleaning and rearranging fields), some additional fields listed in Table A.8.1 have to be added to the attribute table for further processes in the arrival time estimation part of the study.

Table A.8.1 Additional fields in the postprocessing phase

Field Name	Data Type	Description
STOP_ATE	date	Arrival Time Estimations of Bus Stops
Max_Trip_TS	date	Trip Ending Timestamp
Min_Trip_TS	date	Trip Beginning Timestamp
TS_NRM	date	Timestamps normalized to zero

The processes conducted through this phase are presented in the following flowchart in Figure A.8.2. The pseudo-codes for the processes explained in this section are given in Table A.8.2

Below GIS tool has been generated in the model builder application of ArcGIS to process all daily trips of the smart card or bus trajectory data feature classes that take place in the dataset by combining the processes mentioned above. (Figure A.8.1)
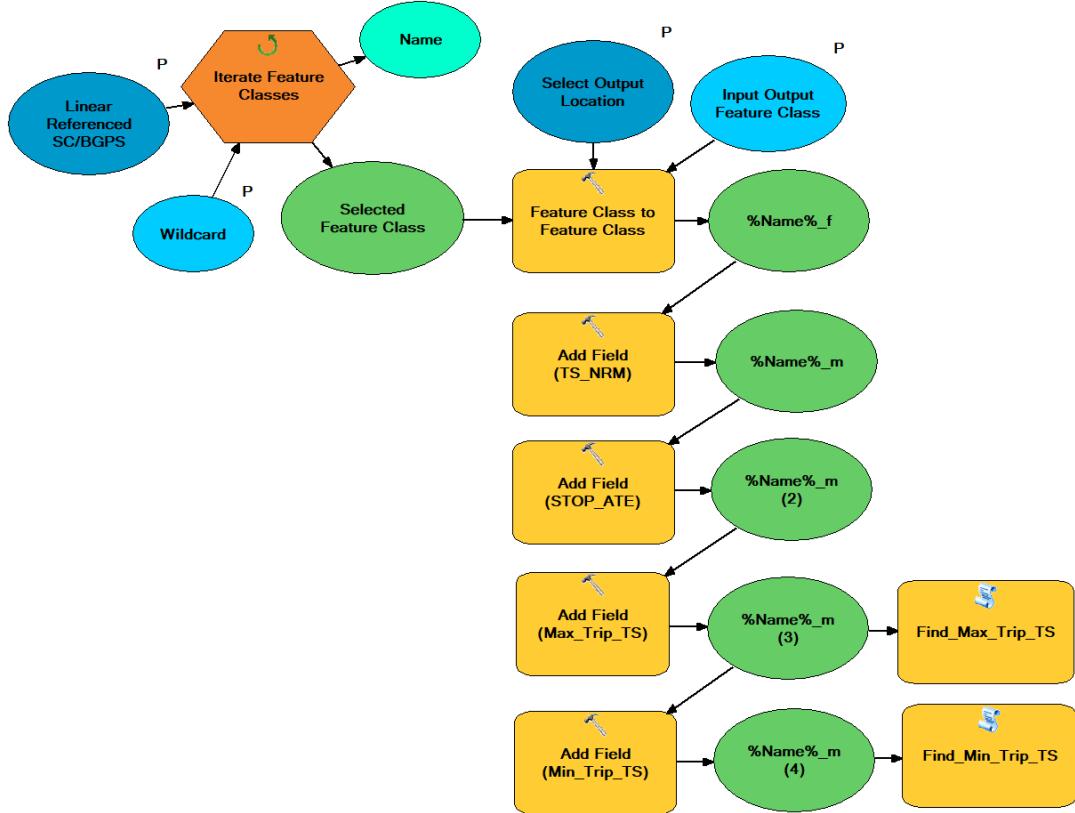
Figure A.8.1 Postprocessing Tool for SCD and BTD
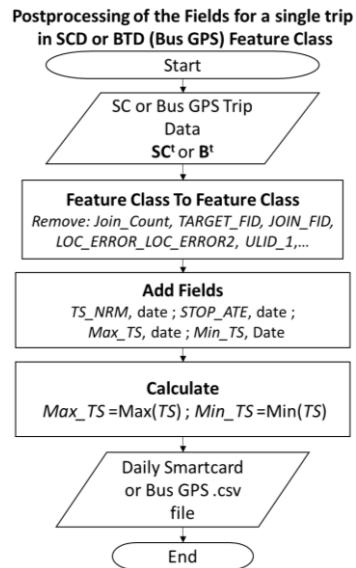
Figure A.8.2 Flowchart for the Postprocessing of fields

Table A.8.2 Pseudocodes for the Postprocessing of the Fields

FOR ∀ **SC^t_i_m** and **B^t_i_m**
{
      CALL Feature Class to Feature Class [Field_Map: Remove; *Join_Count,*
      TARGET_FID, JOIN_FID, LOC_ERROR_LOC_ERROR2, ULID_1,…]
      CALL Add_Field [*TS_NRM*, Date]
      CALL Add_Field [*STOP_ATE*, Date]
      CALL Add_Field [*Max_TS*, Date]
      CALL Field_Calculator
      CALCULATE [*Max_TS* =Max(*TS*)]
      CALL Add_Field [*Min_TS*, Date]
      CALL Field_Calculator
      CALCULATE [*Min_TS* =Min(*TS*)]
      }

A.9 Time-Space and Inter-Stop Travel Time Diagrams

Time-Space and Inter-Stop Travel Time diagrams show the change in the location of a bus along its route in the course of the travel time. In order to generate these diagrams of a bus, kilometers and estimated arrival times of bus stops must be calculated. The station kilometers of bus-stops have already been calculated by Linear Referencing and Dynamic Segmentation in the previous chapter. Therefore, the unknown elements for these diagrams are only the estimated arrival times of bus-stops. Before the estimation process, the necessary data tables need to be prepared by merging the feature classes of smart card or bus trajectory data with the feature class of bus-stops with station kilometers. The flowchart of the preparation phase is in the below Figure A.9.1. The pseudocodes are presented in Table A.9.1 and Table A.9.2.
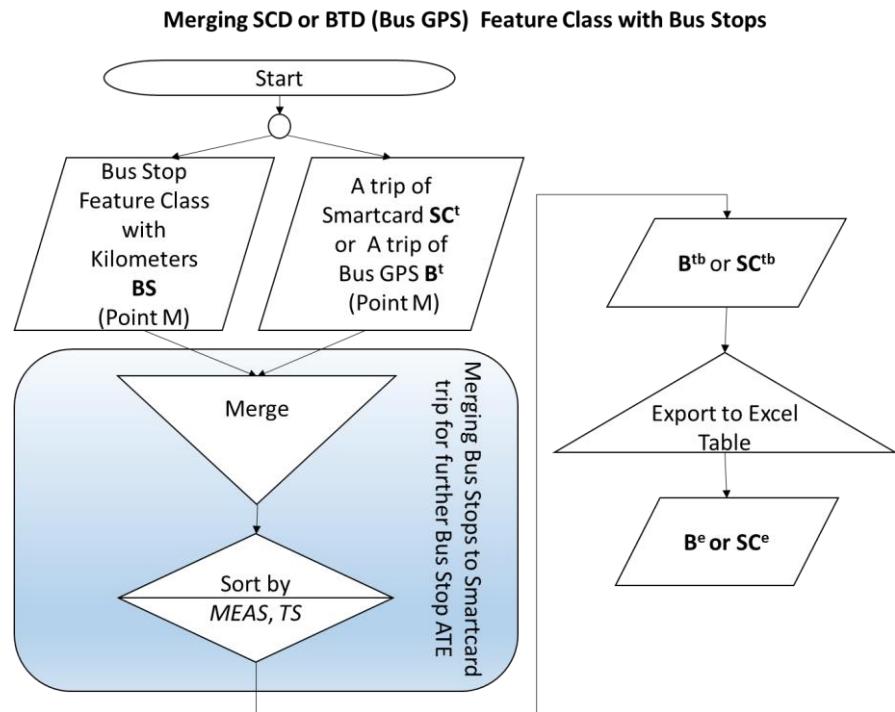
Figure A.9.1 Merging SCD/BTD Feature Class with Bus Stops

Table A.9.1 Pseudocodes for Merging BTD Feature Class with Bus Stops For a given ULID

Get Bus Stop feature class **BS** { *MEAS, S_NO_1, S_ORD_1, ULID, DIR_CHG_C, F_TYPE,*
TS TYPE,…}
Get a trip of BTD Feature Class **B^t**
FOR $\forall$ **B^t_i**
{
        CALL Merge [Input: **BS** and **B^t_i**, Output: **B$^{tb}_i$**]
        CALL Sort [**B$^{tb}_i$**, MEAS, TS, Ascending]
        CALL Table To Excel [**B$^{tb}_i$**]
}

Table A.9.2 Pseudocodes for Merging Smart Card Feature Class with Bus Stops For a given ULID

For a given *ULID*,
Get Bus Stop feature class **BS** { *MEAS, S_NO_1, S_ORD_1, ULID, DIR_CHG_C, F_TYPE,*
TS TYPE,…}
Get a trip of Smartcard Feature Class **SC^t**
FOR ∀ **SC^t_i**
{
        CALL Merge [Input: **BS** and **SC^t_i**, Output: **SC$^{tb}_i$**]
        CALL Sort [MEAS, TS, Ascending]
        CALL Table To Excel [**SC$^{tb}_i$**]
}

A.9.1 Generated Tool for Merging BTD/SCD Feature Classes with Bus Stops

Another tool has been built in the model builder and ran for the preparation of time-space diagram tables (Figure A.9.3 and Figure A.9.4). This model, respectively;

1. Merges smartcard and bus-stops feature classes,
2. Sorts the resulting attribute tables according to the kilometers and timestamps,
3. Adds the normalized timestamp field for the calculation of normalized timestamps,
4. Finally, exports attribute tables to the spreadsheet files in Excel format.
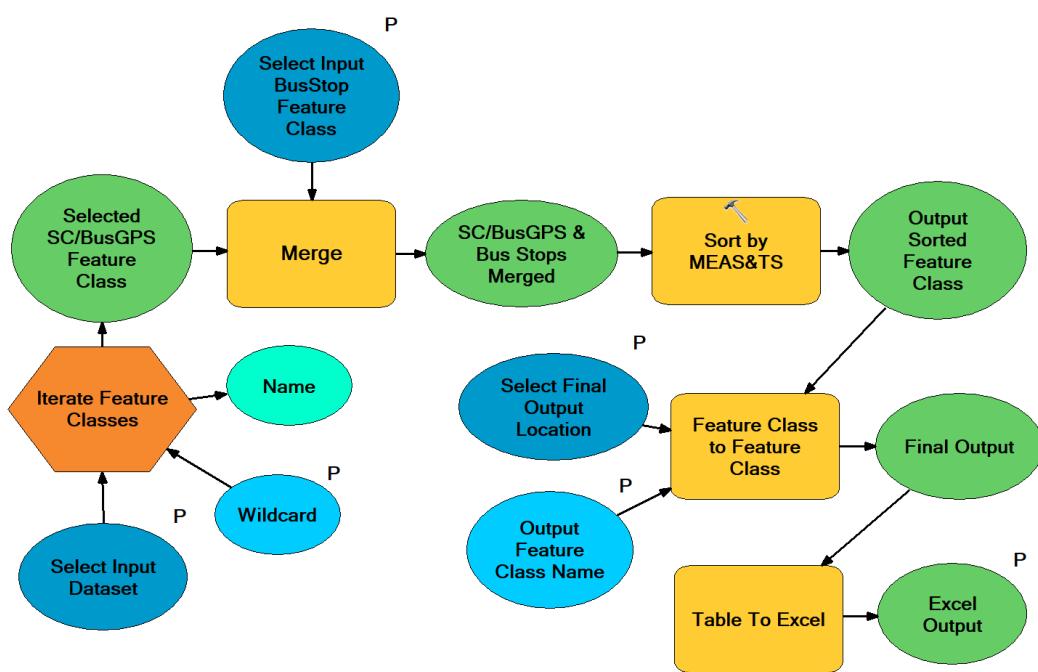
Figure A.9.3 Produced tool for merging BTD or SCD feature class with bus stops (Time-Space Table Generation) in Model Builder
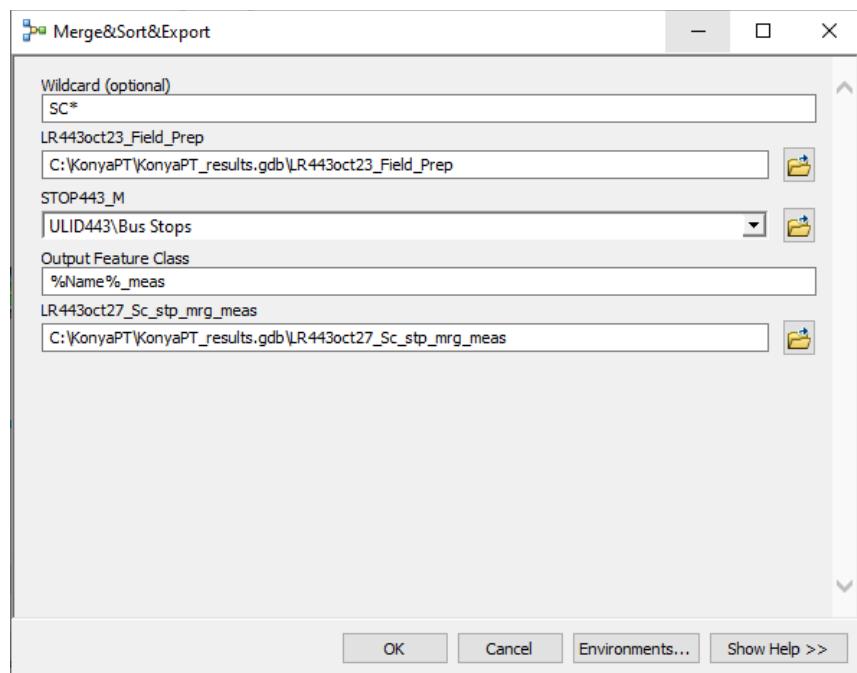
Figure A.9.4 Merge, Sort and Export tool interface

This tool, with the interface was shown in Figure A.9.4, accomplishes all the tasks for the feature classes that include the selected wildcard characters in the selected feature dataset folder and save the results to the selected geodatabase. Since bus-stops are lack of timestamp field, resulting attribute table after the merge process (see Figure A.9.5) has to be sorted according to the kilometer field. A sorted object id field is useful to see which bus-stop is between which smart card data in the later interpolation and extrapolation process. Null values in the timestamp fields belong to the bus-stop points. These Null values have to be estimated in order to generate time-space and inter-stop travel time diagrams.

(a)

Sc443oct23PT443_211_1_f_meas

OBJ	Shape *	TS	TS_NRM	MEAS	F_TYPE	TS_TYPE	DIR_CHG_CH	S_ORD_1	S_ORD_2	AsgnStopOrd	AsgnStat	Ulid_Vid_Trip	FMEAS	TMEAS	ULID	DESTINATION
1	Point M	<Null>	<Null>	0.0000	STOP	1		0	1	2	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
2	Point M	<Null>	<Null>	0.2674	STOP	1		0	2	3	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
3	Point M	10/23/2018 19:32:29	<Null>	0.6218	<Null>	<Null>		0	2	3	3	1	443_211_1	0.2674	0.6242	443	BOSNA SANAY
4	Point M	<Null>	<Null>	0.6242	STOP	1		0	3	4	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
5	Point M	<Null>	<Null>	1.2013	STOP	1		0	4	5	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SA
6	Point M	<Null>	<Null>	1.5685	STOP	1		0	5	6	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA S
7	Point M	<Null>	<Null>	1.9539	STOP	1		0	6	7	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA
8	Point M	<Null>	<Null>	2.2749	STOP	1		0	7	8	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
9	Point M	<Null>	<Null>	2.7496	STOP	1		0	8	9	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
10	Point M	<Null>	<Null>	3.1077	STOP	1		0	9	10	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
11	Point M	<Null>	<Null>	3.4179	STOP	1		0	10	11	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
12	Point M	<Null>	<Null>	4.0231	STOP	1		0	11	12	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA SANAY
13	Point M	10/23/2018 19:40:00	<Null>	4.0392	<Null>	<Null>		0	11	12	11	1	443_211_1	4.0231	4.7114	443	BOSNA SANAY

DESTINATION	S_NO_1	S_NO_2	S_NAME_1	VID	TT	CID	CT	ShiftID	INDEX	STripID	ATUS	USID
BOSNA SANAYI	1022d	<Null>	Akraba Sokak	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	1021d	<Null>	Nefer Sokak	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	1021d	1020d	<Null>	211	BILET	0471C5B1FE982C6EA143779C646729		2	443_211	2	1	44D	ULID443_2_3_1021d_1020d
NA SANAYI	1020d	<Null>	Büyükşehir Spor Ve Kongre Merkez	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
SANAYI	1019d	<Null>	Zeliha Seymen İlkokulu	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
SNA SANAYI	1018d	<Null>	Hacı Kalfa Sokak	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
SNA SANAYI	1017d	<Null>	Mevlana Kültür Merkezi	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	1016d	<Null>	Cemre Sokak	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	1015d	<Null>	Namık Paşa Camii	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	1014d	<Null>	Gaziler Derneği	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	1130a	<Null>	Rumi Sokak	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	2407a	<Null>	Şerafettin - Samanpazarı1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
BOSNA SANAYI	3d	2407a	<Null>	211	BILET	1E65809352755A52807C18E7DCF603		0	443_211	7	1	44D	ULID443_12_11_3d_2407a

(b)

bgpsu443o30v211d224_1618_1900_R_T443_211_224_1_f_meas

OBJ	Shape *	TS	TS_NRM	MEAS	F_TYPE	TS_TYPE	DIR_CHG_CH	S_ORD_1	S_ORD_2	ShiftID	tripID	FMEA	TMEAS	shft_intrvl	ULID	DEST	VID
1	Point M	10/30/2018 16:40:28	<Null>	0.0000	<Null>	<Null>	0	1	2	443_211_22	1	0.0000	0.2674	2018-10-30 16:18:	443	BOSNA S	211
2	Point M	<Null>	<Null>	0.0000	STOP	1	0	1	2	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA S	211
3	Point M	10/30/2018 16:40:33	<Null>	0.0558	<Null>	<Null>	0	1	2	443_211_22	1	0.0000	0.2674	2018-10-30 16:18:	443	BOSNA S	211
4	Point M	10/30/2018 16:40:38	<Null>	0.1335	<Null>	<Null>	0	1	2	443_211_22	1	0.0000	0.2674	2018-10-30 16:18:	443	BOSNA S	211
5	Point M	10/30/2018 16:40:43	<Null>	0.2142	<Null>	<Null>	0	1	2	443_211_22	1	0.0000	0.2674	2018-10-30 16:18:	443	BOSNA S	211
6	Point M	<Null>	<Null>	0.2674	STOP	1	0	2	3	<Null>	<Null>	<Null>	<Null>	<Null>	443	BOSNA S	211
7	Point M	10/30/2018 16:40:48	<Null>	0.2890	<Null>	<Null>	0	2	3	443_211_22	1	0.2674	0.6242	2018-10-30 16:18:	443	BOSNA S	211

VID	SID	PID	SATN	H	V	LID	DID	MV_DIR_F	MV_DIR_T	SEGTID	F_TS	L_TS	RID	ATUS	USID	S_NO_1	S_NO_2	S_NAME_1
211	27289231	59	13	266	55	44	224	1	2	ULID443_1022d_1	10/30/2018 16:4	10/30/2018 16:4	443	44D	ULID443_1_2_1022d_1021d	1022d	1021d	<Null>
<Null>	<Null>	<Nul>	<Null>	<Nul	<Nul	<Nul	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	443	<Null>	<Null>	1022d	<Null>	Akraba Sok
2	27289234	60	13	268	52	44	224	1	2	ULID443_1022d_1	10/30/2018 16:4	10/30/2018 16:4	443	44D	ULID443_1_2_1022d_1021d	1022d	1021d	<Null>
2	289253	61	13	270	58	44	224	1	2	ULID443_1022d_1	10/30/2018 16:4	10/30/2018 16:4	443	44D	ULID443_1_2_1022d_1021d	1022d	1021d	<Null>
2	27289292	62	13	272	56	44	224	1	2	ULID443_1022d_1	10/30/2018 16:4	10/30/2018 16:4	443	44D	ULID443_1_2_1022d_1021d	1022d	1021d	<Null>
<Null>	<Null>	<Nul>	<Null>	<Nul	<Nul	<Nul	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	443	<Null>	<Null>	1021d	<Null>	Nefer Sokak
211	27289326	63	13	271	51	44	224	1	2	ULID443_1022d_1	10/30/2018 16:4	10/30/2018 16:4	443	44D	ULID443_2_3_1021d_1020d	1021d	1020d	<Null>

Figure A.9.5 The attribute tables of the resulting feature class for (a) SCD and (b) BTD after using merge, sort and export tool
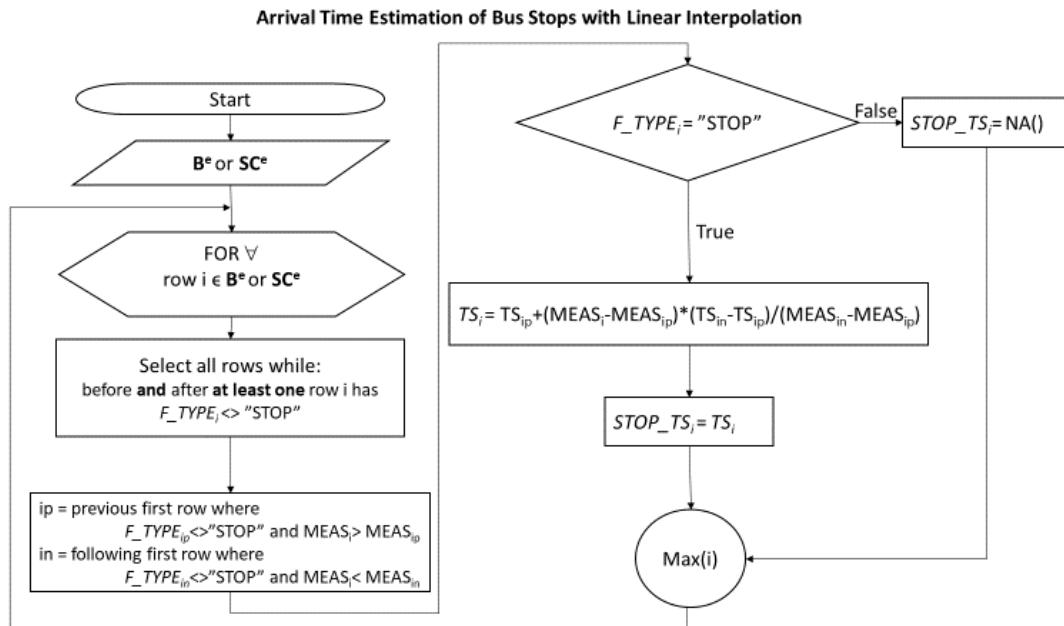
A.10 Arrival Time Estimation (ATE)

Figure A.10.1 Arrival Time Estimation of Bus Stops with Linear Interpolation

Table A.10.1 Arrival Time Estimation of Bus Stops with Linear Interpolation

For a given *ULID,* d, s and t, get the BTD Excel table $\mathbf{B^e}$ { *MEAS, SID, VID, ULID, TS, TS NRM, STOP_TS, Min_TS, Max_TS, ShiftID, tripID, S_ORD_1,… }*
or get the smartcard Excel table $\mathbf{SC^e}$ { *MEAS, CID, VID, ULID, TS, TS NRM, STOP TS, Min TS, Max TS, ShiftID, StripID, S_ORD_1, F_TYPE,… }*
FOR $\forall$ $\mathbf{B^e}$ or $\mathbf{SC^e}$
{
       FOR $\forall$ row i $\in$ $\mathbf{B^e}$ or $\mathbf{SC^e}$
              ip = previous first row where F_TYPE_{ip}<>"STOP" and $MEAS_i$> $MEAS_{ip}$
              in = following first row where F_TYPE_{in}<>"STOP" and $MEAS_i$< $MEAS_{in}$
              Select all rows while;before **and** after **at least** one row i has F_TYPE_i<> "STOP"
              {
              IF F_TYPE_i = "STOP":
                     TS_i = TS_{ip}+($MEAS_i$-$MEAS_{ip}$)*(TS_{in}-TS_{ip})/($MEAS_{in}$-$MEAS_{ip}$)
                     $STOP\_TS_i$ = TS_i
              ELSE:
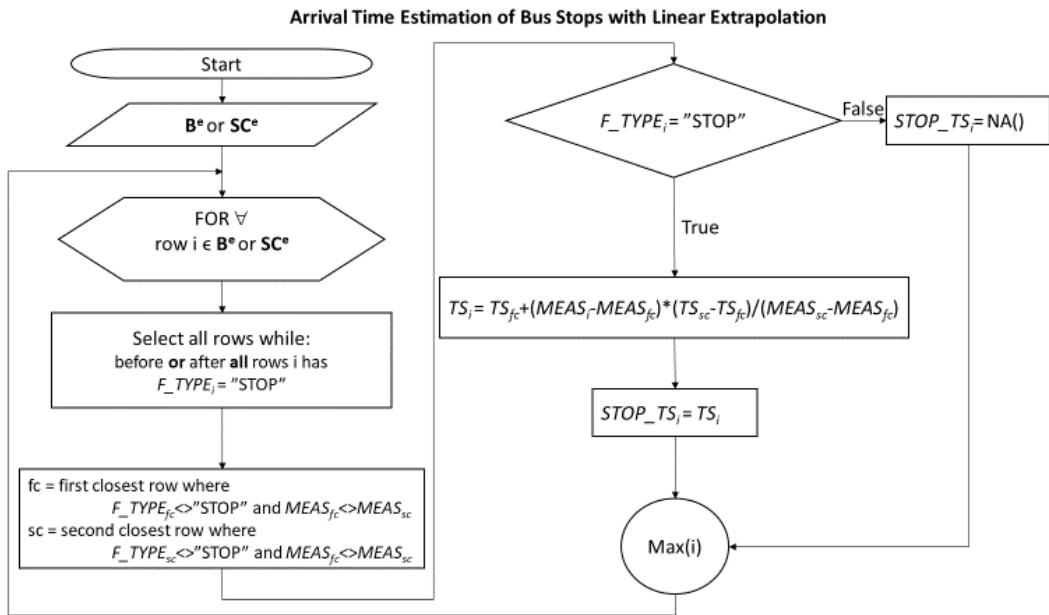 $STOP\ TS_i$= NA()
              }
}

Figure A.10.2 Arrival Time Estimation of Bus Stops with Linear Extrapolation

Table A.10.2 Arrival Time Estimation of Bus Stops with Linear Extrapolation

For a given *ULID*, d, s and t, get the BTD Excel table $\mathbf{B^e}$ { *MEAS, SID, VID, ULID, TS, TS_NRM, STOP_TS, Min_TS, Max_TS, ShiftID, tripID, S_ORD_1, F_TYPE,… }*
or get the smartcard Excel table $\mathbf{SC^e}$ { *MEAS, CID, VID, ULID, TS, TS_NRM, STOP_TS, Min_TS, Max_TS, ShiftID, StripID, S_ORD_1, F_TYPE,… }*
FOR ∀ $\mathbf{B^e}$ or $\mathbf{SC^e}$
{
      FOR ∀ row i ∈ $\mathbf{B^e}$ or $\mathbf{SC^e}$

           fc = first closest row where F_TYPE_{fc}<>"STOP" and $MEAS_{fc}$<>$MEAS_{sc}$
           sc = second closest row where F_TYPE_{sc}<>"STOP" and $MEAS_{fc}$<>$MEAS_{sc}$
           Select all rows while;before **or** after **all** rows i has F_TYPE_i="STOP"
           {
      IF F_TYPE_i = "STOP":
           $TS_i = TS_{fc}+(MEAS_i-MEAS_{fc})*(TS_{sc}-TS_{fc})/(MEAS_{sc}-MEAS_{fc})$
           $STOP\_TS_i = TS_i$
      ELSE:
           $STOP\_TS_i$= NA()
      }
}