Title
IImproving Public Transit Resilience by Leveraging Smartcard Data for Rapid Decision-Making under Highly Dynamic Conditions

Author
Caicedo Castro, Juan David

Publication Date
2023

Improving Public Transit Resilience by Leveraging Smartcard Data for Rapid Decision-Making under Highly Dynamic Conditions

by

Juan David Caicedo Castro

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

and the Designated Emphasis

in

Global Metropolitan Studies

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Joan Walker, Chair
Professor Marta González
Professor Daniel Chatman

Spring 2023

Improving Public Transit Resilience by Leveraging Smartcard Data for Rapid Decision-Making under Highly Dynamic Conditions

by

Juan David Caicedo Castro

Abstract

Improving Public Transit Resilience by Leveraging Smartcard Data for Rapid Decision-Making under Highly Dynamic Conditions

by

Juan David Caicedo Castro

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

and the Designated Emphasis in

Global Metropolitan Studies

University of California, Berkeley

Professor Joan Walker, Chair

The COVID-19 pandemic presented an unprecedented challenge to public transit systems worldwide, testing the ability of transit agencies to make swift and difficult decisions. Balancing the need to provide essential workers with reliable transportation options while dealing with suddenly reduced revenue and increased operational costs forced many agencies to reduce transit services, making life difficult for countless commuters relying on transit. While the pandemic spotlighted the importance of decisive action during times of crisis, disruptions to public transit systems are a regular occurrence. Disruptions vary significantly in time, duration, location, scope, frequency, and impact. Terrorist attacks, the aftermath of violent protests, and natural disasters are a few examples that can drastically change the transit system and our travel needs. Responding quickly and effectively to these disruptions by making informed decisions is crucial not only for minimizing the impact of the disruption, but also for maintaining the reliability of public transit systems for those who depend on them, ensuring they remain available and accessible when needed most.

In recent years, transit agencies have increasingly begun to adopt automated fare collections (AFC) systems, also known as smartcards, to collect fares, making valuable data readily available for analysis. In response, the research community has proposed a number of inferential methods and predictive models to analyze transit patterns. More than 300 research papers have been published on short-term ridership prediction models in the last decade. Despite the abundance of literature, we, the research community, have yet to generate practical knowledge for decision-making. Lack of reproducibility due to inadequate documentation, inaccessible tools and resources, and code and data-sharing issues are prevalent; hindering

our ability to create collective wisdom. The problem is further compounded by the lack of rigorous statistical analysis in many studies, making it difficult to identify relevant research, particularly when metrics across studies are incomparable.

Adding to this chaos, there is also a significant gap in inference methods and predictive models to account for disruptions in general. Current inference methods, which are mostly used to enrich transaction data with other data sources, are designed for stable periods and do not capture the changing transit behaviors of transit users during disruptions, such as those caused by the COVID-19 pandemic. Beyond the pandemic, station closures, a common cause of disruptions, are not accounted for in any predictive model. This further hinders the ability of transit agencies to utilize such models for decision-making during disruptions.

This dissertation demonstrates how to use readily available smartcard data to quickly support decision-making in highly dynamic conditions. The objectives of the dissertation are to:

1. Enhance the value of smartcard data by integrating it with other data sources to provide more comprehensive, near-real-time insights during disruptions.
2. Create an open-source repository for short-term ridership prediction models to facilitate accurate and reliable comparisons, and to accelerate advancements of the models.
3. Compare and assess the performance of state-of-the-art methods for short-term ridership prediction in a highly dynamic condition.
4. Improve the accuracy and reliability of short-term ridership prediction models during disruptions by integrating station closure information.

All work is done using the transaction data of 147 stations in the Bus Rapid Transit (BRT) system in Bogotá, Colombia, from August 2015 to May 2021. This dataset includes two highly dynamic conditions: a month-long protest that occurred in November-December 2019 and the COVID-19 pandemic that began in March 2020. Both conditions drastically changed the travel needs of the users and the transit system. Additionally, the dataset covers several station closures, which are not necessarily associated with these two primary events.

In the study described in Chapter 2, I examined the changes in demand for the BRT system in Bogotá during the COVID-19 pandemic. Despite the disruptions caused by the pandemic, transit agencies had to continue providing services as it was a crucial transportation option for many essential workers. However, a comprehensive understanding of the remaining transit users, as well as how the demand adjusted during the recovery period across different population segments, was lacking, despite being critical information to inform the prioritization of service changes, station cleanings, and the effective implementation of safety measures. I enriched smartcard data transactions with block-level demographic data from the 2018 census to infer the socio-economic strata of BRT users, a variable that can be used as a proxy for income. I present a methodology that combines behavioral modeling techniques with well-established inference methods to capture the dynamic heterogeneity of transit use by strata at different points in time, and use this information to estimate a probability vector that assigns frequent transit users to a stratum. At the beginning of the pandemic, the reduction in transit use for all strata was similar, at around 85 to 90%, but members of lower strata returned to transit at a pace five times faster than members of higher strata. The proposed method had the advantage of providing near real-time insights into the evolution of the COVID-19 pandemic using existing and available data.

Chapter 3 describes a study in which I explored predictive models to anticipate ridership and enable proactive resource allocation ahead of time. It was difficult to reliably compare and test models, even under stable conditions, let alone during highly dynamic conditions. Therefore, I built an open-source infrastructure with running code for five major methodologies commonly used in the literature, including econometric and deep learning approaches. The open-source code also provides the Bogotá BRT smartcard dataset for the five-year period for other researches to reproduce and replicate results.

While the literature on short-term ridership prediction focuses on relatively stable condition, I systematically compared their performance with two highly dynamic conditions, a one-month-long labour protest, and the COVID-19 pandemic. In stable conditions, most tested models performed relatively similarly, with forecasting errors varying from 8.5 to 12%. However, all models performed significantly worse in both highly dynamic conditions relative to stable conditions. In the protest condition, increases in prediction error ranged from 14 to 24%. During the COVID-19 pandemic, increases in prediction error ranged from 12 to 82%. Notably, in the COVID-19 pandemic scenario, a Recurrent Neural Network (RNN) model, with a long-short term memory (LSTM) cell stood out by outperforming other models and adapting faster to disruptions. The prediction error stabilized within about 1.5 months, while other models had higher error rates even a year after the start of the pandemic.

In Chapter 4, I noted that some disruptions resulted in temporary station closures, causing an increase in demand at nearby transit stations. But station closure information was not incorporated into any short-term predictive models. To improve predictions for stations affected by closures, I proposed a new model that uses graph theory and the attention mechanism. The model aimed to capture spatiotemporal correlations, so that the prediction at one station is sensitive to what happens in another station in a previous time step. By incorporating station closure information into the proposed model, I reduced the prediction errors by 3.3 to 23.5% for station closures compared to existing models in the open-source codebase infrastructure developed in Chapter 3. These improvements were observed across multiple metrics and modeling strategies.

In summary, the contributions of this dissertation are to:

1. Develop a method to enrich smartcard data with socio-economic information for highly

dynamic conditions.

2. Expand the understanding of the change in transit use by different socio-demographic groups during the COVID-19 pandemic in Bogotá.

3. Create an open-source codebase infrastructure to consolidate short-term ridership prediction research to perform systematic, reliable and statistically rigorous benchmarks tests to accelerate the advancements of the models.

4. Systematically compare the performance of state-of-the-art methods for short-term ridership prediction during both stable and highly dynamic conditions.

5. Demonstrate the importance of including station closures to improve the prediction accuracy for nearby stations.

6. Propose a novel modeling framework that captures spatiotemporal correlations of the transit network, improving overall performance and the prediction accuracy for stations impacted by other station closures.

7. Integrate the new modeling framework in the open-source codebase infrastructure for other researchers to replicate, reproduce, compare, improve, and build upon our model.

Smartcard data provides timely information about transit demand that can inform operational and tactical decisions of transit agencies. While not as rich in demographic information as surveys, these data provide timely and relevant information that is critical during disruptions. This information can help transit agencies understand how users adjust travel during disruptions and prioritize populations that may be disproportionately affected. Accurate forecasting can also give transit agencies additional time to plan resource reallocation and take appropriate actions. The creation of an open-source codebase infrastructure lowers the barrier to reproducibility. It allows for reliable and systematic comparisons of model performance, identifying relevant research and ultimately generating knowledge to improve decision-making.

Thorough this dissertation, I empirically demonstrate the contributions by leveraging publicly available smartcard data from the BRT system in Bogotá spanning more than five years, including multiple highly dynamic periods of time affecting transit ridership. I showcased the power of smartcard data in understanding the evolution of the impact of the COVID-19 pandemic on frequent transit users from different population segments. In the spirit of collaboration, I built an open-source codebase to foster more innovative solutions, enabling researchers, practitioners, and transit agencies to generate collective wisdom that advances scientific knowledge. I proposed a new modeling framework that leverages station closure information to improve overall short-term ridership forecast. In conclusion, my work demonstrates the potential of utilizing smartcard data to support time-sensitive decision-making and policies, particularly in the context of disruptions. Moreover, I developed an open-source infrastructure to expedite innovation in this space.

To my parents, and Mauri

Contents

List of Figures

List of Tables

Acknowledgments

As I reflect on my academic journey, I can only be grateful for the many individuals who have contributed to the successful completion of this dissertation. This journey has been filled with moments of joy, growth, and accomplishments but also self-doubt, exhaustion, and frustration. Their unwavering belief in me, encouragement, and kindness have given me the strength and courage to persevere.

First and foremost, I would like to express my deepest appreciation to my husband, Mauri, who has been my rock throughout this journey. His unconditional love, and support have been invaluable to me. He has been there when I most needed it most, and a feel extremely fortunate to have him by my side.

I am immensely grateful to my advisor, Joan Walker, for her guidance, knowledge, and expertise in shaping my research. Her passion for the field, love for the research, and commitment to excellence have been an inspiration to me. She not only taught me how to conduct research but also challenged me to think deeper and ask why instead of hows.

I would also like to thank my mentors, Paul Waddell, Marta Gonzáles, and Daniel Chatman for their support and encouragement throughout my academic journey. Their passion for building a better world and looking beyond the data has inspired most of my work, and I am grateful for their guidance and mentorship.

To my parents, who have always wanted the best for me, and all the sacrifices they made to make it that happen, I express my deepest gratitude. Their unconditional love and encouragement have been a constant source of motivation, and I am grateful for their unwavering faith in me.

I also want to acknowledge my siblings, Maria Isabel and Carlos. They have been an invaluable inspiration throughout my life. I am grateful to them for bringing my nephews and niece, Benjy, Martin, and Amelia, into the world. Their innocent joy and laughter have been a constant source of motivation, helping me to persevere through the challenges. I am fortunate to have such a loving family, and I am grateful for the role they have played in my life.

Finally, I would also like to thank my friends, Jeremy, Jessica, Jean, Sol, Soomin, and Esther, who have been part of this journey since my first year in Berkeley. Their support and encouragement have been invaluable, and I am grateful for the friendships we have formed, the experiences we have shared, and the love we have for the field and the research. And all my other friends, who have supported me, distracted me when I needed a break, and expressed their love in multiple ways, this is also for you.

Chapter 1

Introduction

Public transit has always been an integral part of my life. In fact, I only learned to drive a car at the age of 27 when I moved to the United States. Before that, public transit enabled me to move around as needed. Initially, I relied on transit because my parents' schedules did not align with mine, and it was the only viable alternative to walking. Growing up in Colombia, my story is far from unique. Hundred of millions of people around the world rely on public transit every day for commuting to work and school, socializing with friends, attending events, and accessing healthcare services.

Throughout my nearly 20-year experience utilizing public transit, I have faced many challenges, such as overcrowded buses, lengthy commutes, extended waiting periods, aging vehicles, and safety issues. These frustrations have often led to wasted time that could have been spent with family and friends. The COVID-19 pandemic only emphasized the importance of developing a resilient transit system that can provide an effective service, even in times of uncertainty. I couldn't help but wonder who is in charge of making transit better and if, in some way, I could contribute to it in a meaningful way.

While working at the Secretaria Distrital de Movilidad, the agency responsible for managing Bogotá's transportation infrastructure, there was much talk about data – especially the "big data" revolution that many believed could solve numerous transportation issues, including public transit. However, there was a clear gap between collecting new data and being able to use it for actionable policies, as the agency lacked the resources, skills, and knowledge to do so.

On the other hand, my experience in research showed a completely different picture. Over the last decade, there has been a significant increase in publications covering a wide range of topics on transit, using various datasets, and employing numerous methods. This highlighted a noticeable divide between the valuable insights generated by academic research and the ability of practitioners to apply that knowledge effectively in real-world situations, especially in the global south context. This disconnect, combined with my passion for public transit, inspired the contributions made throughout this dissertation.

1.1 Motivation

Public transit is critical in providing access to opportunities, generating economic benefits, promoting sustainable transportation, and improving public health, particularly for low-income populations. In numerous cities across the Global South, employment prospects and opportunities are predominantly concentrated in downtown areas, while low-income populations tend to reside in the city's periphery [40, 17]. This spatial mismatch between housing and employment opportunities results in greater travel distances, rendering walking and cycling trips less viable options. Private motorized transportation is often beyond the reach of low-income individuals and even more in the global south context [88], leaving many people reliant on public transit as their sole affordable means of accessing various destinations, including jobs and opportunities. Public transit is also a more environmentally sustainable option than private modes of transportation, with buses emitting a maximum of 33 CO_2-equivalent grams per passenger per kilometer compared to 263 CO_2-equivalent grams for private vehicles [95]. Additionally, public transit can improve public health outcomes by encouraging physical activity, such as increased walking and cycling to access and egress public transit services [12, 86, 71].

However, public transit systems are vulnerable to various anticipated and unforeseen disruptions, including maintenance operations, construction activities, accidents, natural disasters, technical malfunctions, security threats, protests, and public health emergencies [36]. The size and scope of these disruptions may fluctuate over time and space, occur with different frequencies, and have varying degrees of impact, but all can alter the travel needs and circumstances of commuters. Disruptions to public transit cause prolonged waiting times, service suspensions, and station closures, leading to trip cancellations or mode changes[111], consequently causing user dissatisfaction [21]. This shift can further increase the dependence on private vehicles, leading to a decline in the overall market share of public transit [111, 75]. Such consequences could undermine the advantages of public transit during disruptions, particularly for individuals who cannot drive due to physical limitations or those with low incomes who cannot afford a car.

To mitigate and reduce the impact of the disruption, transit agencies need to make informed decisions fast. Figure 1.1, based on studies by [36] and [114], depicts a resilient system's time-variant performance. The system operates stably during the pre-disturbance period, with its resilience relying on its reliability. Once there is a disturbance in the system, the performance level decrease until a reaction/response decision is made to start the recovery period. The sooner an action plan is implemented, the faster the system will return to planned operations.

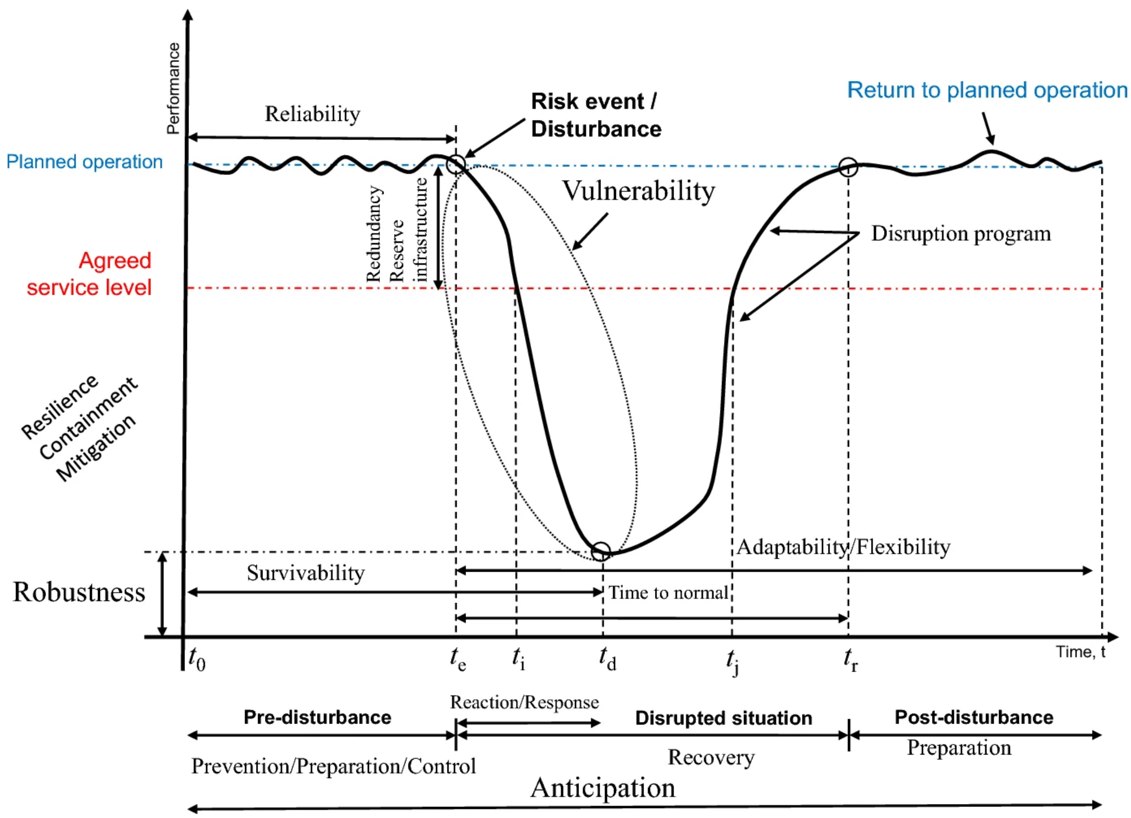

Figure 1.1: A diagram illustrating the system performance during a risk event. Taken from[36, 114]

During these disturbances, public transit agencies can make several decisions, including modification of vehicle scheduling, fleet size per type of vehicle, driver schedules, and driver roster [50]. For instance, there exist methods to implement new vehicle scheduling that can be put in place to minimizing waiting times [119], travel time [16], as well as operational and user costs [123], and even incorporating fueling constraints [61]. However, one critical input for these methods is transit demand. Therefore, an accurate forecast can inform transit operations in advance, providing them with additional time to allocate their resources efficiently where and when needed.

In recent years, transit agencies have increasingly adopted smart cards to collect fares, making valuable data readily available for analysis, and can potentially be used to inform decisions during disruptions. In response, the research community has proposed many inference methods and predictive models to unravel transit patterns and behaviors.For instance, over the last decade, there has been a steady growth in the number of publications on short-term ridership prediction models, with over 300 methodological research papers published on this topic during this time period, as illustrated in Figure 1.2. Despite the abundance of literature, we, the research community, have yet to generate practical knowledge for decision-making. Lack of reproducibility due to inadequate documentation, inaccessible tools and resources, and code and data-sharing issues are prevalent, hindering our ability to create collective wisdom. The difficulty in identifying relevant research, particularly when metrics across studies are incomparable, further compounds the problem. Adding to this chaos, there is a significant gap in inference methods and predictive models to account for highly dynamic conditions. One example is station closures, a common consequence of disruptions, which are not accounted for in any predictive model. This further hinders the ability of transit agencies to utilize such models for decision-making during disruptions.

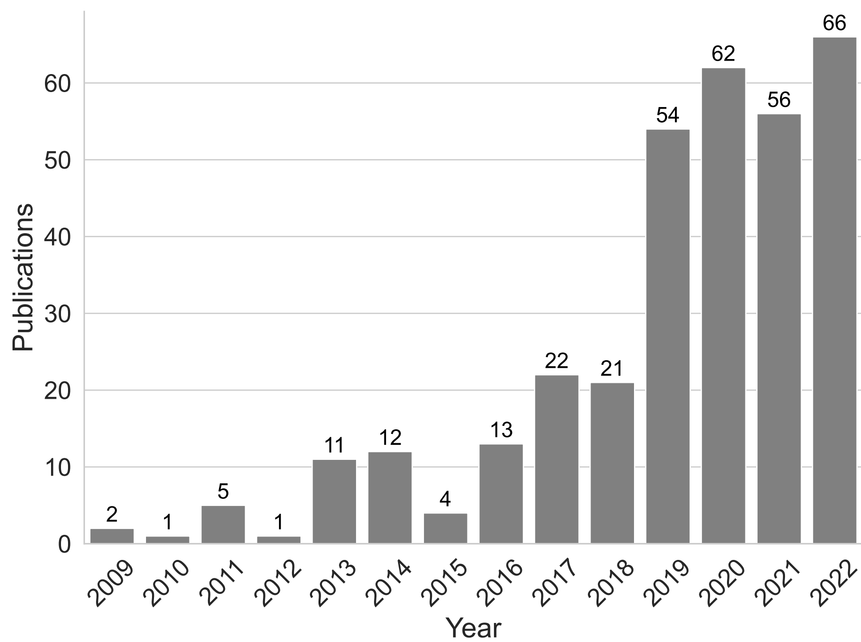

Figure 1.2: Number of Publications on Short-Term Ridership Prediction by Year. Query performed on 11/18/22 with the search term ‘short-term passenger flow prediction’. Only methodological papers related to short-term ridership prediction are included in this plot

1.2 Research Objectives

Therefore, the objective of this dissertation is to use already available data to rapidly inform the decision-making process in the face of disruptions in the public transit system. More specifically, the research objectives are:

1. Enhance the value of smartcard data by integrating it with other data sources to provide a more comprehensive, near-real-time insights during disruptions.
2. Create an open-source repository for short-term ridership prediction models to facilitate accurate and reliable comparisons and to accelerate advancements of the models.
3. Compare and assess the performance of state-of-the-art methods for short-term ridership prediction in a highly dynamic condition.
4. Improve the accuracy and reliability of short-term ridership prediction models during disruptions by integrating station closure information.

For this dissertation, we use the smartcard data for 147 stations of the Bus Rapid Transit (BRT) system in Bogotá, Colombia, spanning from August 2015 to May 2021. The dataset is publicly available at the transit agency's open data portal, and daily transactions are updated daily. Additionally, the time frame contains two highly dynamic conditions. First, a month-long protest in November-December 2019, and the COVID-19 pandemic started in March 2020.

1.3 Inference Methods During Disruptions

The COVID-19 pandemic-induced lockdown resulted in a drastic change in the travel needs of the population, leading to a significant reduction in transit demand in many cities worldwide, with some cities experiencing a decline of up to 90% [51, 105, 52]. This situation presented new challenges for transit agencies, as they had to maintain a reliable and efficient transportation option while dealing with decreased revenue and increased operational costs. In response, many transit agencies reduced their service offerings, impacting those still relying on transit. However, understanding who continued to use transit and for what reasons was unclear yet important for informing strategies to adapt service changes, prioritize safety measures, and reduce operational costs.

Particularly in Bogotá, measures to contain the spread of COVID-19 led to significant changes in public transit operations. To comply with social distancing requirements, public transit services were reduced to 35% capacity. In addition, peak-and-plate measures, a strategy that imposes restricted use of private vehicles, were temporarily lifted to disincentivize the use of public transit. A few months after the start of the pandemic, construction and manufacturing jobs were allowed to reopen with differentiated shifts to avoid crowds, further changing the dynamics and needs of the transit demand. The significant changes in public transit demand brought about by the ongoing impacts of the pandemic underscore the critical need to acquire knowledge of the changing patterns of transit demand under highly dynamic conditions, particularly for those who are disproportionately affected.

One approach to gaining a better understanding of these changing patterns of transit demand is through the collection of survey data [5, 106]. Surveys can provide a wide range of socio-demographic and economic data, but they are costly and time-consuming to collect and analyze. In addition, surveys only provide cross-sectional data and do not track the evolution of travel patterns unless multiple and frequent surveys are collected. During the COVID-19 pandemic, prolonged contact with people outside the household was limited, adding an extra layer of complexity to collecting survey data.

Smartcards, on the other hand, provide rich panel data on transit use for almost the entire system but lack social, demographic, and economic variables. While some research combines transit data with other sources to infer population variables, such as socio-demographic and economic characteristics [27, 31, 4, 45, 34, 18, 65, 18, 65, 76, 42], these methods typically rely on the frequency and regularity of transit use, making them unsuitable for highly dynamic conditions. As a result, current inferential methods cannot be used to accurately infer population variables for highly dynamic conditions, limiting their potential usefulness in such circumstances.

These limitations highlight the need for innovative methods to capture the new dynamics of transit usage during highly dynamic conditions and infer social, demographic, and economic variables that can inform the development of effective strategies for maintaining reliable and efficient transportation options.

In Chapter 2, I present a methodology that integrates behavioral modeling techniques with advanced inference methods to capture the dynamic heterogeneity of transit use by different population segments at different points in time. Using this approach, I leverage the time-variant heterogeneity in transit usage to estimate a probability vector that assigns frequent transit users to the different population segments. I showcase my methodology with the strata variable, a socio-economic classification that can serve as a proxy for income, as an example. I show that lower-strata individuals return to transit at a pace five times greater than middle- and high-strata individuals. This finding highlights the importance of enriching smartcard data with other data sources to gain nearly real-time insights into how different population segments adjust their use of public transit in response to highly dynamic conditions, ultimately improving the ability of transit agencies to make quick and informed decisions.

1.4 Short-term ridership prediction benchmarking under highly dynamic conditions

Highly dynamic conditions, defined as both unexpected and uncertain, can drastically change our travel needs and, in general, the transportation system. For public transit agencies, responding effectively and fast to these disruptions is crucial to maintaining the reliability and accessibility of essential activities, especially for those who rely exclusively on public transit. One strategy to respond to changes in demand during highly dynamic conditions is to adapt vehicle schedules, fleet size, driver schedules, and rosters accordingly. However, to effectively make these adaptations, it is necessary to understand how demand adjusts to these conditions. Short-term ridership prediction models have the potential to inform these decisions in advance, thus allowing for better planning and more effective resource allocation. However, the implementation of these models has been limited to short, stable time frames and a small number of stations. How these models perform in highly dynamic environments has not been studied, nor has their performance been systematically compared.

Despite the development of multiple models for predicting short-term ridership demand using complex methods, a common platform for testing the performance of state-of-the-art methods does not exist. Consequently, a reliable benchmark process for systematically comparing and identifying relevant literature, even under stable conditions, is lacking. In many cases, code sharing and data are unavailable, which hinders the ability of other researchers to replicate results, reproduce methods for other regions, and compare model performances reliably. Moreover, the metrics used for analysis are often not comparable across studies, making it challenging to identify relevant research. As a result, current benchmark practices only use a subset of models that researchers have the capacity to implement, resulting in increased research time and unreliable comparisons, which limits the ability of the research community to build collective wisdom and advance science.

To reduce barriers to implementing state-of-the-art methods, I make available an open-source codebase infrastructure, creating a collaborative space that enables collective efforts and accelerates advancements in the models. The infrastructure includes five commonly used methods in the literature, ranging from traditional ARIMA and SARIMA models to deep learning approaches such as Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) with a Long-Short Term Memory (LSTM) cell for the latter. By developing a comprehensive open-source codebase infrastructure, I eliminate some of the challenges of limited code and data availability for replicating and reproducing results, facilitating more reliable comparisons and facilitate innovation.

In Chapter 3, I make use of the open-source codebase infrastructure to compare model performance for stable conditions and two highly dynamic conditions, namely, a month-long protest and the COVID-19 pandemic. I show that most tested models perform similarly in stable conditions, with forecasting errors ranging from 8.5 to 12%. However, in both highly dynamic conditions, all models performed significantly worse compared to stable conditions. Specifically, during the protest condition, prediction errors increased for all models, ranging from 14 to 24%. Similarly, during the COVID-19 pandemic, the prediction error increased by an average of 12 to 82%.

Through this comparison, I also explore the advances of adaptive learning systems, in which models are constantly learning new patterns as new information becomes available. I demonstrate that the combination of the LSTM model and adaptive learning not only outperformed other models but also adapted faster during the COVID-19 pandemic. While the other models employing adaptive learning techniques continued to exhibit higher error rates even a year after the start of the pandemic, the prediction error of the LSTM model stabilized within approximately 1.5 months. In conclusion, I highlight the effectiveness of employing adaptive learning systems, specifically the LSTM model, in improving short-term ridership prediction and enabling the models to adapt quickly to highly dynamic conditions such as the COVID-19 pandemic.

1.5 Incorporating Station Closure Information

Public transit systems are vulnerable to planned or unplanned disruptions that can lead to temporary closures of transit stations. Such closures can have a significant impact on travel demand patterns, as travelers may be forced to alter their travel plans or use alternative transportation modes[99, 125, 84]. For instance, in Bogotá alone, over a five-year period, there were more than 10,000 closures lasting more than two hours, highlighting the need to incorporate station closure information in short-term ridership prediction models. Smartcard transaction data from Bogotá also revealed that station closures led to increased demand at nearby stations. However, to effectively optimize resource allocation, transit agencies require precise knowledge of when, where, and to what extent the demand of neighboring stations will be affected by the closures.

Despite the importance of station closures, current model implementations ignore their impact, resulting in non-zero demand predictions when stations are closed. Furthermore, it remains unclear how these models perform under station closures, limiting their ability to inform decisions under these circumstances.

In Chapter 4, I quantify the impact of station closures on the prediction accuracy of nearby stations. I show that the prediction error of neighboring stations increases by a factor of two using the deep learning models available the open-source codebase. This finding supports the idea that station closures have a substantial impact on the accuracy of model predictions. However, existing short-term ridership prediction models do not account for this information.

To close this gap, I present a new modeling framework that draws on recent advancements in graphical neural networks and transformer models for time series analysis. The proposed framework utilizes the attention mechanism of the transformer model to focus on the most important information [6]. Using a graph representation of the transit network, I constrained the attention mechanism to the neighboring stations, allowing the model to capture spatial correlations. The attention mechanism is also applied in the temporal component to capture complex spatiotemporal correlations simultaneously.

Using the open-source codebase, I demonstrate that the proposed framework, combined with station closure information, outperforms other state-of-the-art models. For instance, I show that the prediction error of nearby stations impacted by a station closure improves between 3.0 to 23.47% across multiple metrics and modeling decisions compared to the second-best performing model. Finally, in the spirit of collaboration and sharing, I have made the code implementation for this model publicly available in the open-source codebase. This allows other researchers to replicate, reproduce, compare, improve, and build upon the proposed model to advance the field of short-term ridership prediction.

1.6 Contributions

Overall, this dissertation contributes to the systematic use of existing and available data sources, mostly smartcard data, to understand, analyze, infer, and provide tools to support and improve the decision-making processes under highly dynamic conditions.

Inference methods to enrich smartcard data with demographic information typically relied on stable conditions and frequent and regular use of transit, making them unsuitable for highly dynamic conditions. In this dissertation I develop a method to enrich smartcard data with socio-demographic information that adapts to highly dynamic conditions. The method uses individual smartcard data to build a database of frequent uses from whom I infer the Bus Rapid Transit (BRT) station closest to home (home location) using well-established inference methods. I combine block-level census and land use data to infer a probability distribution of a user belonging to a stratum at any given point in time. Lastly, I showcase the method to track the change of transit patterns of frequent transit users by strata in Bogotá, during the first seven first months the COVID-19 pandemic.

Despite the vast number of existing literature on short-term ridership prediction models, there is a lack of a collaborative effort to compare their performances. To address this issue, I develop an open-source codebase infrastructure with five models, including parametric and deep learning approaches, to consolidate short-term ridership prediction implementations and improve benchmark practices. The main goal of this infrastructure is to enable systematic, reliable and statistically rigorous comparisons of models, while also promoting collaboration to accelerate innovation and the advancement of the field.

Moreover, the literature on short-term ridership prediction methods has been limited to short, stable time frames and a small number of stations. However, how these models perform under highly dynamic conditions had not been systematically compared. I use the open-sourced codebase to compare the performance of state-of-the-art methods for short-term ridership prediction during stable and two highly dynamic conditions, including the COVID-19 pandemic and a month-long protest in Bogotá, Colombia. The benchmark also evaluates two modeling design strategies: multi-output models and adaptive learning, which aim to improve existing models in highly dynamic conditions.

I demonstrate that station closure impact the demand of neighbouring stations, however existing short-term ridership existing models did not make use of this information increased the prediction error of nearby station. In this dissertation, I demonstrate the importance of station closures to improve the prediction accuracy of nearby stations of a station closure for existing methodologies. Additionally, I proposed a new modeling framework that uses the station closure information and the attention mechanism to capture complex spatiotemporal correlations, which further improves the prediction accuracy of stations impacted by station closures. Finally, I make the code implementation of this new model available, enabling other researchers to reproduce, compare, improve, and build upon the model.

We propose a modeling framework that captures spatiotemporal correlations for the transit network, improving overall performance and the prediction accuracy of stations impacted by other station closures. The proposed model includes a graph representation of the transit network and the attention mechanism to simultaneously capture spatiotemporal correlations, especially during station closures. The model is made available in the open-source codebase to allow other researchers to reproduce, compare, improve, and build upon.

1.7 Dissertation Outline

This dissertation is structured as follows:

Chapter 2 presents the methodology used to enrich smartcard data during highly dynamic conditions, along with empirical evidence of its use to track changes in transit demand in Bogotá, Colombia during the COVID-19 pandemic.

Chapter 3 introduces the five selected methodologies used to implement the open-source codebase platform, along with the results of the meta-analysis conducted for stable conditions, the one-month-long protest, and the COVID-19 pandemic.

Chapter 4 centers around the integration of station closure information in short-term ridership prediction models, introducing a new model that utilizes a transformer-based architecture with graph information to improve prediction accuracy in the event of station closures.

Chapter 5 concludes with this dissertation with the main key finding, contributions, and future research directions.

Chapter 2

Influence of socioeconomic factors on transit demand during the COVID-19 pandemic: A case study of Bogotá's BRT system

ABSTRACT[1]

The COVID-19 pandemic restricted most economic and social activities, impacting travel demand for all transportation modes and especially for transit. We hypothesize that the shifts in travel demand varied by socioeconomic status, and we assess the differential impact of COVID-19 on the Bus Rapid Transit (BRT) patronage across various socioeconomic groups in Bogotá. We built a database of frequent transit users with data collected by smartcards in Bogotá's BRT system between January and October 2020. For each user in the database, we labeled their home and work stations. Transactions at other stations are classified as "other." The stratum (a government socioeconomic classification of residential units in Colombia) of a BRT station's service area was assigned using an estimated probability vector for each user belonging to a specific stratum; this data is validated with aggregate strata distributions in the 2019 household travel survey. Our study found that the reduction in transactions for lower-strata users is significantly less than that of the middle and high strata. The magnitude of this difference varies over time but stabilizes after the end of the lockdown. The growth rate of "other" transactions per thousand people is greater than the growth for home and work locations, especially for the lowest strata. Other studies have shown that the radius of gyration (Rg) (a measure of how far indi-

[1]This Chapter is adapted from [13] and has been reproduced with the permission of my co-authors Joan L. Walker, and Marta C. González.

viduals travel away from home) has decreased by about 50% after the lockdowns. Our study shows that when measuring Rg only for users that continued using BRT, the Rg slightly decreased for lower and medium strata but increased for high strata. The contribution of this study is to build a method to classify BRT transactions of frequent users by strata and to describe their trends to expand our understanding of the COVID-19 lockdowns impacts in the Global South context. These results are a starting point to inform policy and decision-makers to guide the recovery efforts to improve transit accessibility and level of service for captive users such as low-stratum users.

2.1 Introduction

One of the major effects of the COVID-19 outbreak and the lockdowns is the drastic reduction of transit demand. In China, mode share for metro dropped from 26% to 14% [51]. In Europe and North America, transit ridership declined by as much as 90% during the first weeks of the lockdown measures [105]. Likewise, the Inter-America Development Bank shows a ridership reduction of 60% to 90% in major cities across Latin America [52].

Since the beginning of the outbreak, many studies have shown a differential impact of the lockdowns on minorities and vulnerable groups, such as gender [3], poverty [98], and small businesses [33]. However, in transportation, most of the analyses have focused on understanding the general behavioral change in travel patterns and mode shares. For instance, [48] used passive data from a routing app to analyze travel times by modes, type of visited locations, and preference of origins and destinations. An ongoing study in Switzerland, using GPS data of 250 participants, shows that train and tram use remains constant at a 60% reduction compared to the baseline, but daily bike distance increased 100% on average [73]. Similarly, [103] found that the average trip time for the bike-sharing system in New York increased from 13 to 19 minutes in dock stations nearby metro stations. In terms of public transit, [77] found a more significant demand reduction in transit stops near schools, universities, and supermarkets.

There is a growing literature that explores the impact on transportation demand by socioeconomic group. For example, [5], using a convenient sample survey of 4395 individuals in Chile, found that 77% of workers from low-income households do not telecommute. A survey of more than 25,000 transit users in North America showed that pandemic-era riders are primarily low-income essential workers with no private vehicle access [106]. Using smartcard data, researchers have found a correlation between socioeconomics and transit use reduction. These studies infer socioeconomic variables , such as income, by different fare values [11] or by matching census data and with home boarding stations [121, 2]. [30] used smartcard data in Bogotá to explore the relationship between transit and strata, informality, and poverty. The study found that lower strata, higher informality, and higher poverty correlated with more transit reduction use. However, this study linked a transaction to the characteristics of the area rather than the individual. Therefore, a transaction of a low-strata user in a high-income area may have counted as a high-strata transaction.

While the literature suggests that travel shifts during COVID-19 have varied by socioeconomic group, little is known about what changes will persist and how the evolution will vary by group. [115] showed that behavioral inertia caused by the pandemic might result in public transit demand of only 73% compared to pre-pandemic levels in New York, but they do not investigate who they are. The governments' role is to provide a safe and efficient transit system [122]. As transit has proven to provide a transportation option to essential services, transit service should accommodate their operation to serve the population that most depend on it [110]. Therefore, it is critical to assess and understand the evolution of the transit demand disaggregated by different socioeconomic groups. An assessment of this kind is necessary to plan transit operations with a vertical equity objective in mind (i.e., allocating services to where they are most needed) rather than horizontal equity (i.e., allocating services proportionally to everyone) [14]. While surveys provide an opportunity to explore the differences in transit demand with rich socioeconomic variables, they are costly and time-consuming. Actions and decisions need to be taken nearly in real-time.

This study assesses and describes the change and evolution of the transit demand by different socioeconomic groups during the COVID-19 pandemic in Bogotá. We leveraged smartcard data from Bogotá's Bus Rapid Transit (BRT) system to assess the differential impact of the COVID-19 pandemic restrictions by socioeconomic groups. Although the data does not contain any socioeconomic variables, we use inference methods [78, 80] to obtain a probability distribution for one socioeconomic variable in the context of Bogotá. Our research objectives are (i) to infer socioeconomic characteristics of BRT frequent users based on smartcard data, and (ii) to use this information to expand our understanding of the COVID-19 lockdowns impacts and trends on transit demand by different strata. While changes in the supply may causes changes in the demand, and changes in demand may also cause changes in supply, this two-way causality is not possible to address with the available data. Therefore, this is not an objective of our study.

The remainder of this paper is organized as follows. Section 2 provides the background information on Bogotá, its transit system, and the COVID-19 timeline. Section 3 summarizes the methodology we used to build a transit users' database of frequent users. Section 4 introduces the three main findings of this research related to transit use change by strata. In Section 5, we conduct an extensive discussion of the results. Section 6 presents the study limitations and future research. Finally, in Section 7, we state the conclusion of this study.

2.2 Background

Bogotá is the largest city in Colombia; according to the 2018 census, it has a population of 7.2 million. The Household Travel Survey (HTS), 2019 estimated that 30% of the trips, or about 4.5 million trips, are made daily in the public transit system. Bogotá's transit system has two main components, the BRT and the Integrated Public Transit System (SITP, for its acronym in Spanish). The BRT system has 152 stations, three cable car stations, and it records 2.5 million transactions a day [108]. Passengers use a contactless card to validate at entry points only; there is no need to validate the exit point as the system uses a flat fare.

Bogotá has a stratum classification, which is an official socioeconomic variable assigned to every residential unit in Colombia. Stratum ranges from one to six depending on physical characteristics of the building (such as construction materials, built squared meters, type and quality of finishes), the environment (such as dominant topography, dominant constructive typification, dominant densities, and public utilities accessibility and coverage), and the urban context (land use, and characteristics and access to the road network) [24]. According to Law 142 of 1994, the purpose of this classification is to provide a differential charge for public utilities, apportion subsidies, and charge extra contributions. Note that the stratum is attached to a psychical living dwelling but not to a specific individual. While strata do not consider the household income, [15] concluded that it is justifiable to understand it as a proxy for income - Low-strata correlates with lower-income and high-strata with higher-income. Many studies in transportation have used strata as a predictor variable in travel behavior [74, 26, 8].

Like many cities and countries, Bogotá and Colombia declared a lockdown in March 2020 aiming to reduce the spread of the COVID-19 virus, which halted most of the region's economic activities. Contrary to the US, Europe, and China, the lockdown measure in Bogotá and Colombia was preventive. The early lockdown allowed time to prepare the health system for a future wave of COVID-19 patients [72]. Bogotá's mayor decreed a drill lockdown on March 20, and then the national government declared a mandatory lockdown on March 24 that lasted until September 1. In April, the government required transit systems to operate at 35% capacity in an effort to stop the spread of the virus. The peak-and-plate measure, a policy that restricts private vehicle use on certain days depending on the last digit of the vehicle's plate, was lifted to avoid public transit use. On April 27, the national government allowed construction and manufacturing jobs to reopen. In Bogotá, a differential shift was imposed to avoid crowds in the public system starting May 11. Construction work was allowed from 10 a.m. to 7 p.m., manufacturing from 10 a.m. to 5 p.m., and retail from noon to midnight. On May 30, Bogotá's "Kennedy" neighborhood had a contagious rate significantly higher than any other, which forced the city to declare a lockdown specific to just that neighborhood. However, the BRT stations within Kennedy were still in operation. Given an increase of COVID-19 cases and intensive-care beds' saturation, the city decided to impose a new lockdown plan. This new plan divided Bogotá's neighborhoods into three groups, which would go on lockdown for 14 days each, starting on July 13, 23, and 31, respectively. It was later announced that the lockdowns would be applied to the fourth group of neighborhoods from August 16 until August 30. On August 25, the city increased transit effective capacity to 50%. On September 1, the national government ended the mandatory lockdown and moved to a new plan of selective isolation for individuals with symptoms or recent exposure. On September 22, the peak-and-plate measure was reinstated as traffic level recovered to a pre-lockdown situation, although these new measures would not apply to vehicles with more than three occupants or health care workers.

2.3 Methodology

The BRT smartcard data is a longitudinal observation of transit users in Bogotá. However, socioeconomic variables and characteristics of the trips are not included. In this section, we describe our framework (Figure 2.1) to infer the trip origin and stratum to complement the data. Our objective was to create a longitudinal data set of frequent transit users (before the lockdown) with inferred stratum and trip origin information.

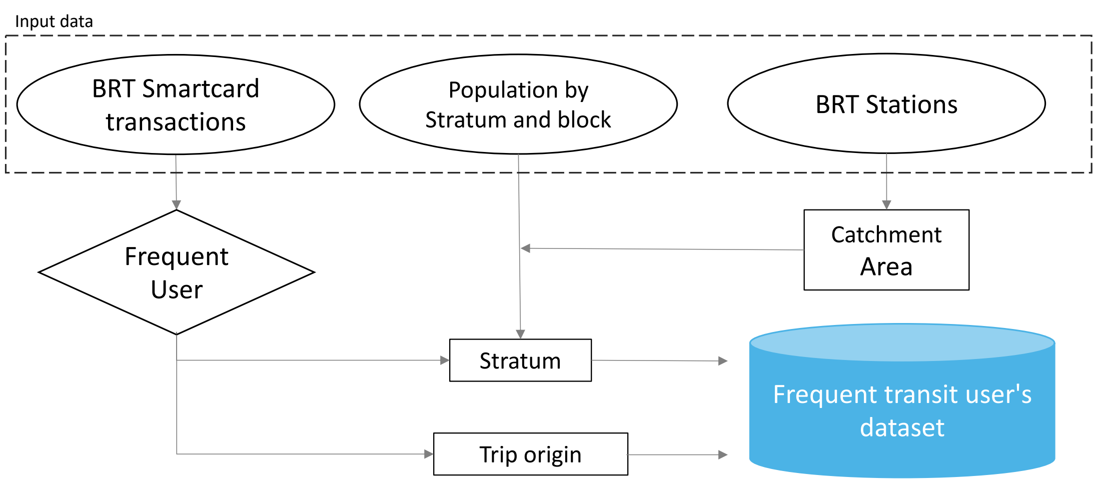

Figure 2.1: Frequent transit users database development framework.

2.3.1 Data input and pre-processing

We used Bogotá's BRT smartcard data of individual transactions from January 1, 2020, to October 12, 2020. A transaction is automatically recorded when a smartcard approaches a reader device at the BRT station entrance. A transaction record contains the date and time, station name, entry point (for stations with multiple entry points), the fare, card type, and a unique identifying number. The unique identifying number links subsequent transactions made by the same card. In total, the dataset has more than 247 million transactions representing about 7.5 million unique smartcards [107]. We also used the official strata classification dataset, a shapefile containing more than 45,000 blocks with the stratum classification [91] and the block-level population extracted from the official census website [23].

We pre-processed smartcard data to obtain an individual's transit travel sequence. We defined a transit travel sequence as all transactions registered by the same card ID. However, a travel sequences may have double tap-ins or internal transfers within the system that need to be removed from the analysis. Consecutive transactions made in the same station by the same card within 30 minutes is assumed to be only one transaction. We also eliminated internal transfers. For example, some feeder stops are not located in the station but a nearby location. In these cases, users need to tap-in when they exit the feeder route stop and tap-in again when they enter the main station; however, this is only one transaction. We also eliminated transactions in "Virtual Stations" as they do not correspond to any physical location. One user may use multiple smartcards, or multiple users may use a unique smartcard, however the data set does not distinguish between the two. Therefore, we assumed that one card represents a unique user. Similar assumptions have been made in similar smartcard data studies [134].

We used the boarding stations and entry points to differentiate transactions of multiple services. Stations with feeder or intercity connections have designated entry points to ensure payment when entering a BRT station. Therefore, a BRT station can have walking, feeder, and intercity entry points, and each transaction is classified accordingly.

2.3.2 Frequent users

The objective of separating frequent and non-frequent users is to distinguish transit travel sequences with enough information to infer their transit travel behavior. Users who do not have enough information may be sporadic users or visitors who use the system for a short amount of time. To compare individuals before and after the lockdowns, the selection of frequent users considered the transactions between January 1, 2020, and March 8, 2020.

To determine frequent users, we characterized each transit travel sequence with three variables: (1) Total number of transactions, (2) Number of active days, and (3) average transactions per active day. A distribution density plot for each variable is shown in Figure 2.2. As expected, the number of transactions decreases exponentially (Figure 2.2A). The distribution of the number of active days suggests that most unique transit users are sporadic (Figure 2.2B). Additionally, the distribution of the number of transactions per active day tails off after five, suggesting that most people have fewer than five transactions a day on average (Figure 2.2C). To better understand the relationship among these variables, we plotted a joint distribution of the number of transactions versus the active days (Figure 2.2D). The plots show a linear relationship between the variables. However, there is a high concentration of sequences with fewer transactions and days of use that are more likely to be non-frequent users. We selected the cutoff point with the following rules: (i) the cutoff point should increase the variance of the joint distribution, and (ii) the average values of the resulting distribution should be greater than the cutoff points. Figure 2.2E shows the joint distribution of the number of transactions and the active days with a cutoff point of 15 transactions and ten active days of use. The density distribution spreads along a line with slope one, suggesting an increased variance. The average values were higher than the cutoff points, suggesting that most infrequent transactions were filtered.

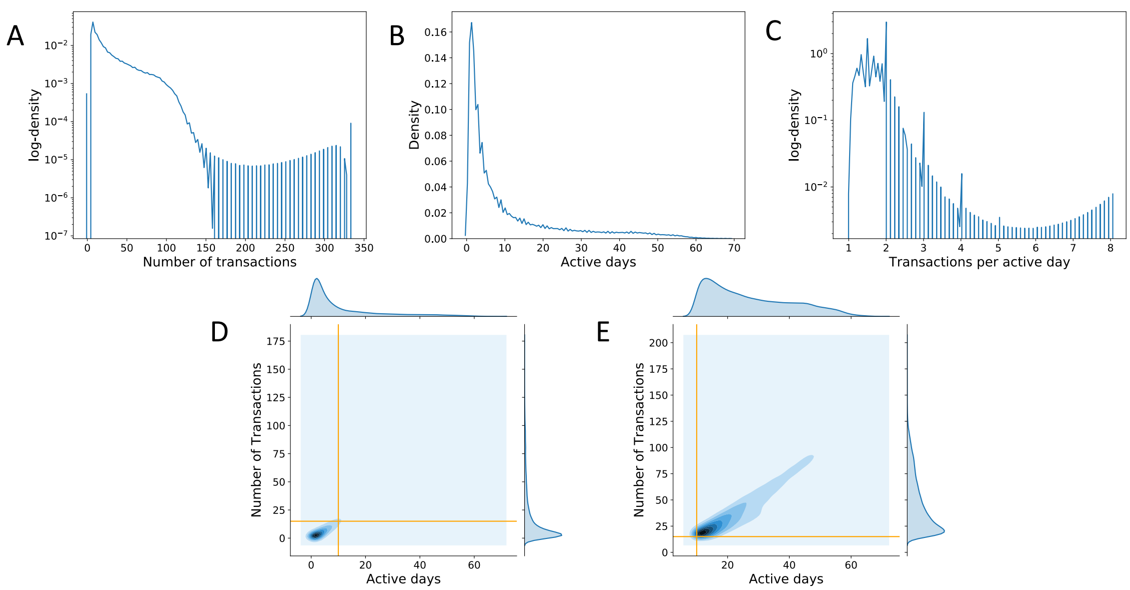

Figure 2.2: Transit travel sequence distributions. **(A)** Distribution of the number of transactions. **(B)**Distribution number of active days. **(C)** Distribution of the number of transactions per day. **(D)** non-filtered joint distribution number of transaction and active days. **(D)** Filter joint distribution number of transaction and active days. Orange line represent the cutoff filter points at 10 active days, and 15 number of transaction.

Frequent transit users were classified as those unique identifier numbers that had transit travel sequences with more than 15 transactions, ten active days of use, fewer than 200 transactions, and average daily transactions fewer than five. After this filtering process, we remained with approximately 2 million users representing 70% of the total transactions. We only used these frequent users in our analysis.

Note that this process may have captured users who may not use transit every day or use transit frequently during a short period. While the inference of these users may be limited (e.g., we may not be able to infer a work location), these users tend to use the system in a certain way. For instance, someone who uses BRT for personal appointments or shopping will have the origin of their transactions most likely be linked to a regular location. Therefore, our study focuses not only on people who use transit frequently for commuting but also who use transit frequently for other activities.

2.3.3 Trip origin - Home, work and other location inference

Transit travel sequences of frequent users reveal important information about each individual's activity locations. The literature suggests a ruled-based approach that leverages the regularity of the transit travel sequence to infer locations such as home and work [27, 31, 4]. For example, the most frequent station is classified as the home station [45, 34]; or, if the time between two consecutive transactions is greater than a regular working session (e.g., 8 hours), then it is a commuting trip [18, 65]. These studies may also add external variables such as land use and household travel surveys to enrich the inference method [18, 65, 76]. [42] also suggests using continuous hidden Markov models to identify clusters in the smartcard data population and use this information to infer home and work locations. However, they tested their methodology with two days of observation.

The purpose of this section is to describe the methods used in this study to infer home and work locations. We also classified transactions at other locations as "other." For simplicity, we used a similar set of rules as [65]. We adjusted some of the rules to reflect that our data set only contains tap-in information. We used the following rules to infer the activity locations:

- Home: Most common first transaction of the day before noon.
- Work: Most common transaction such that difference between consecutive transaction in the same day is greater than 6 hours.
- Other: Transactions in any other station that are not classified as home or work.

Note that some transit sequences may have only one transaction a day or consecutive trips in the same day with less than 6 hours difference. For these users, a work location was not inferred. They could potentially be less frequent riders that use the BRT for discretionary activities, and therefore these transactions were classified as other.

2.3.4 Catchment areas

The catchment area of a BRT station is its area of influence. We assumed that each frequent transit rider in our processed database lives within the catchment area of their inferred home station. From the HTS, we estimated that 89% of the transit users live within the catchment areas and 87% access the BRT by walking. Only 5.9% and 3.8% access the station by other bus services and informal transit, respectively. Additionally, from the smartcard data, only 7.7% of the home location transactions are identified as transfers from other bus services that are not the feeder routes. These percentages validate that our assumption is a reasonable approximation of the reality.

To estimate a catchment area, we followed these principles:

- The area of influence is a buffer with the maximum walking distance a person is willing to walk to access mass public transit.
- Catchment areas of different stations do not overlap.
- Only feeder routes will add to the catchment area outside the maximum walking distance.

The maximum walking distance to a BRT station is estimated from the 2019 HTS. From the survey, we filtered respondents that mention BRT as the transportation mode. We extracted the walking access and egress time from home to BRT station. From the distributions, 90% of the respondents reported access walking time less than 20 minutes. We assumed an average walking speed of 4.5 km/h, representing a maximum walking distance of 1500m. This distance is more likely to correspond to a Manhattan distance. However, we created the buffer with a Euclidean distance. To convert the Manhattan distance to a Euclidean distance, we used the fact that the sum of the squared legs for any right triangle corresponds to the hypotenuse squared. To simplify the conversion, we assumed an isosceles triangle; therefore, the corresponding Euclidean distance is $2c$, where c is the triangle's leg. From the Pythagoras theorem, we also know that for any isosceles triangle, the hypotenuse is $\sqrt{2} * c$. Finally, the conversion coefficient is $\sqrt{2}/2$. Therefore, the distance to create the catchment area buffers is 1000m. A similar radial distance for BRT stations catchment areas is suggested in [54]. If two catchment areas covered the same point, we assigned it to the closest station.

Feeder routes are a free service that amplify the catchment area of the station. Because it is a free service, users are not required to tap-in when they board a feeder bus, and thus, it is not possible to know the exact location of the boarding. However, stations have special entry points for feeder routes that are easy to identify. To account for these services, we defined the catchment area of a feeder route as a buffer of 500 meters along the route, and considered it a different unit of analysis. Feeder route information did not include feeder services at "Portal 20 de Julio"; therefore, transactions from feeder lines at this station were excluded from the analysis. Figure 2.3 shows the catchment areas for both the stations and the feeder routes.

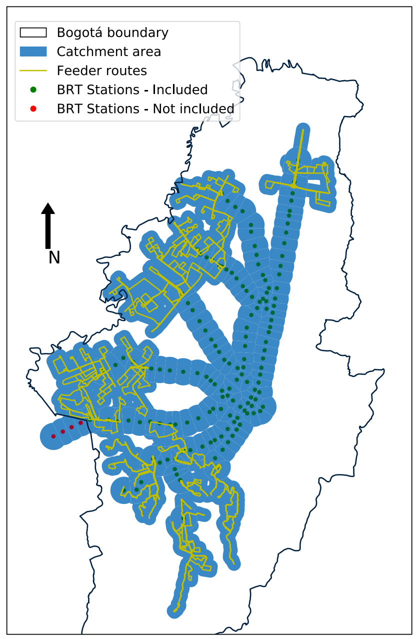

Figure 2.3: BRT stations, feeder routes and catchment areas

2.3.5 Stratum

To infer the stratum of a user, we estimated an s-dimensional probability vector $P_n(s)$ that reflects, for each station n, the probability that a frequent transit rider classified as living in that station's catchment area belongs to a particular stratum s. The probability of station n having stratum s ($P_n(S = s)$) depends on the population of stratum s within the catchment area of station n and the rate at which a strata s uses the BRT ($rate_s$). The importance of the rate is that it capture strata strata heterogeneity in BRT use, which is critical for this study's objective.

$$P_n(S = s) = \frac{Pop_{n,s} * rate_s}{\sum_{j \subseteq S} Pop_{n,j} * rate_j} \tag{2.1}$$

To estimate $rate_s$, we regressed the transactions per station for those users where that station was classified as their home-station with respect to the population by stratum within the catchment area of a station. The form of this regression can be any generalized linear model for continuous or count variables as shown in Eq. 2.2. The rate is the average marginal effect, which can be understood as the change in the number of home-transactions for one increase in a given stratum population (Eq. 2.3).

$$transactions = f(z, \theta) \tag{2.2}$$

$$rate_s = \frac{\partial transactions}{\partial \theta_s} \tag{2.3}$$

To estimate BRT use rates by strata, we used ordinary least square (OLS). Using OLS estimates, we obtained a similar stratum shares distribution as the negative binomial estimates and the 2019 HTS for a week in February. While the negative binomial model performs better in terms of the AIC, it resulted in two outlier stations (large catchment area and a small number of transactions), which disproportionately increased the error measures such as the root mean squared error (RMSE). By the central limit theorem, a normal distribution approximates a negative binomial distribution for large mean values [25], which is the case for this study (the average number of home transactions in the BRT system is 3556 transactions per station/day). One advantage of OLS estimation is that it provides straightforward interpretability of the coefficients (e.g., the average marginal effect of a variable is its estimated coefficient). The implication of using this approximation is that we could potentially get negative prediction or fractional number; however, the objective of performing a regression is not to predict ridership but to capture the heterogeneity in the use of BRT by strata. Since we obtain a similar result with both regressions, we preferred the OLS estimate for its interpretability.

Eq. 2.4 is the regression used to estimate the rates for each stratum. The dependent variable was the number of transactions per station corresponding to users where that station was classified as a home-station. The independent variables were the population by stratum in the catchment area. The estimated coefficients represent the average number of transactions per inhabitant of a given stratum, which we called the stratum transaction rate. Since strata five and six are just 5.3% of the population, we estimated a unique coefficient for this combined population segment.

$$transactions = \theta_0 + rate_1 * pop_{stratum1} + ... + rate_{56} * pop_{stratum56} \tag{2.4}$$

Table 2.1: Sample OLS estimation for strata inference. Samples days are Mondays, Wednesday, Saturday and Sunday

	Home transactions 2020-02-17	Home transactions 2020-02-19	Home transactions 2020-02-22	Home transactions 2020-02-23
stratum1	0.126***	0.113***	0.088***	0.028***
	(0.019)	(0.020)	(0.011)	(0.004)
stratum2	0.069***	0.069***	0.047***	0.016***
	(0.006)	(0.007)	(0.004)	(0.001)
stratum3	0.068***	0.072***	0.031***	0.011***
	(0.012)	(0.012)	(0.006)	(0.002)
stratum4	0.084**	0.085**	0.038*	0.020***
	(0.039)	(0.039)	(0.022)	(0.007)
strata 5 & 6	0.150**	0.154**	0.063	0.024*
	(0.068)	(0.069)	(0.038)	(0.013)
R-squared	0.786	0.777	0.821	0.828
Adj. R-squared	0.793	0.784	0.826	0.833
N	163.0000	163.0000	163.0000	162.0000

* $p < .1$, ** $p < .05$, ***$p < .01$

In stable conditions, the stratum's transaction rate would only vary for weekdays and weekends and during seasonal holidays. However, under highly dynamic conditions, this rate is likely to change daily. Thus, to estimate the impact of COVID-19 lockdowns in BRT use, we estimated the stratum's transaction rate daily. Inter-city trips, including the trips from Soacha, are excluded from the analysis because strata information at the block level is only available in Bogotá. Table 2.1 shows an example of the results for four different days (Monday, Wednesday, Saturday, and Sunday). We noticed that coefficients for higher strata tend to be smaller in magnitude compared to lower strata. These results suggest that individuals from higher strata are less likely to use BRT, especially during the weekends. Note that these results do not include an intercept because it was statistically insignificant for all regressions

2.3.6 Validation

As with most inference methods and passive data, ground truth data is rarely available. Therefore it is not possible to validate these results from a statistical point of view. However, we can compare our results with the aggregate distributions from other data sets such as the household travel surveys and the census. This validation process also ensures that expected transit behavior can be observed, such as the bi-modal distribution of trips during a day, and that most inferred home locations are in residential areas.

To validate the proposed inference methods, we compared the results with expected distributions and distributions from other data sources. Figure 2.4A shows the distribution of the inferred home, work, and other locations for every frequent transit sequence. As expected, the distribution of home locations is mostly concentrated in the residential areas of the city perimeter. Similarly, work locations have a higher concentration in the downtown and the extended downtown of Bogotá, where most jobs and universities are located. Lastly, locations classified as “other” are primarily distributed in both the downtown and residential areas, suggesting that “other” are potentially related to discretionary activities. To verify and strengthen our assumptions, we also plotted the transaction time for each location classification as shown in Figure 2.4B. As expected, the transaction time distribution for home locations peaks in the morning, which corresponds with the beginning of typical working/school hours. Likewise, the transaction time distribution for work locations peaks in the afternoon and coincides with the end of the typical working/school schedule. Transaction times for “other” locations are evenly distributed throughout the day, with some peak in the morning, midday and afternoon. Notice that we only constrained transactions before 12 P.M. to infer home locations; once inferred, every transaction in the home location is counted in the transaction distribution.

To validate the strata inference, we compared the strata distribution of BRT users with the strata distribution of the population [23](Figure 2.4C), and the 2019 HTS [90] (Figure 2.4D). Our results show that the strata share closely matches the strata distribution as estimated from the 2019 HTS. From the HTS, we traced BRT trips with access modes different than walking. We found that they are more likely to be low-strata users, which may explain the underestimation of the share for strata one and overestimation of strata two.

Finally, with the methodology proposed in the study, we built a frequent transit users database with 2,011,067 unique users. Each user has a home and work location (when possible) and a strata probability distribution.

2.4 Results

This section presents the main findings and the statistical analysis to support them. We analysed the trends by strata for the transaction reduction (%), the transactions per 1000 people, and the radius of gyration (Rg). To calculate the transactions reduction on a given day, we compared it to the same day in the base week. For instance, any Tuesday is compared to the transactions of the Tuesday in the base week. We selected the week of February 17, 2020, as the base for these calculations as it is a typical week. We avoided using weeks

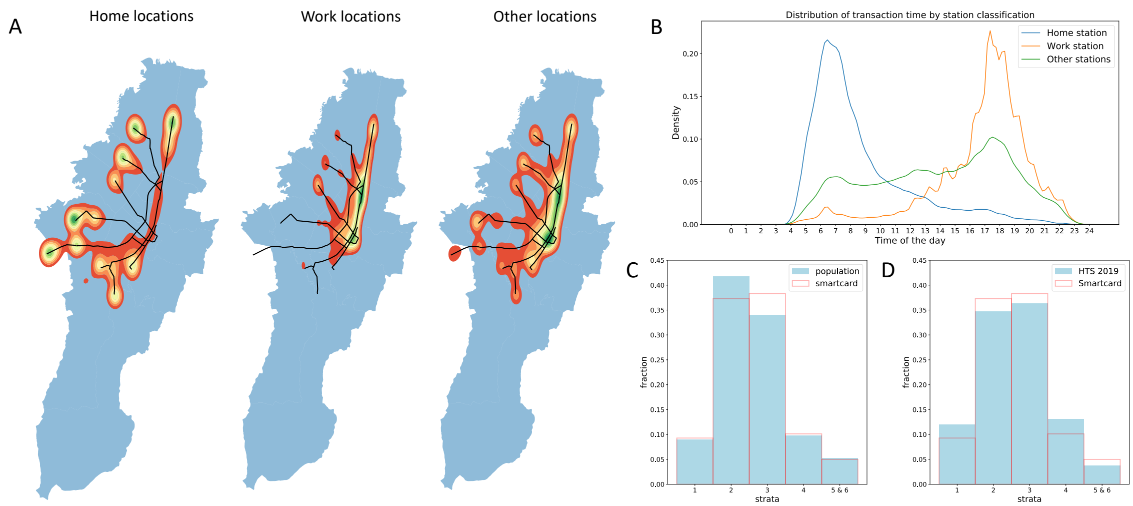

Figure 2.4: Home, work and other location inference validation. **(A)** Density distribution of home, work and other locations. **(B)** Transactions time distribution for home, work and other locations. **(C)** Distribution of population by strata (census, 2018) and BRT user by strata by proposed methodology. **(D)** Distribution of BRT users by strata (HTS, 2019) and proposed methodology.

in January because travel demand is usually atypical. We also avoided the first week of February as, by decree, the first Thursday of February is the day with no car or motorbike. Therefore, we expected transit demand to be higher than usual. Lastly, we defined the radius of gyration (Rg) as the average distances in kilometers reached by a person using BRT during a day, using the home station as an anchor, as shown in Eq. 2.5

$$Rg_n(t) = \sqrt{\frac{1}{S_{n,t}} \sum_{i=1}^{S_{n,t}} d(s_{home,n}, s_i)^2} \tag{2.5}$$

Where $Rg_n(t)$ is the radius of gyration of individual n in the time t, $S_{n,t}$ is the number of stations visited by individual n in time period t, and $d(s_a, s_b)$ is a function that measures the euclidean distance between station s_a and station s_b. Notice that $s_{home,n}$ is the inferred home station of individual n.

For each of these variables (transaction reduction, transactions per 1000 people, and the Rg), we fitted a line using Ordinary Least Squares (OLS), where the dependent variable is a linear function of time, and the slope represents the trend. This analysis's main objective is to have a statistical tool to test hypothesized significant differences in trends by strata. The z test for the difference in two estimated slopes is the one suggested in [79], and we used a 95% confidence level. For each stratum, we fit a line in a defined period after the lockdowns and compared their slopes. If all slopes are statistically similar to each other, we conclude that the strata do not influence BRT demand during the COVID-19 pandemic. For the time variable, we added an indicator to differentiate between weekdays and weekends.

2.4.1 Transactions of lower strata returned to transit at a faster speed

The reduction in the number of transactions by stratum is shown in Figure 2.5. The plot shows the last 7-day transaction reduction to attenuate the weekend effect. We would expect a slight variation during the year under normal circumstances, except for the holiday season in December - January and Easter. At the beginning of the lockdown, there is an initial sharp reduction of transactions of nearly 90% for middle and high strata, while for lower strata, the reduction is 85%. From the beginning of the general lockdown to the start of the sectorized lockdowns, there is a steady growth of transit use for all strata; however, the gap between lower and middle/higher strata increases to 15%, suggesting a more significant transit use growth rate for lower strata. A month after the end of the lockdown restrictions, this difference is approximately 20%. In Figure 2.5 we also see a drastic drop in demand after the sectorized lockdown that affected strata 1 and 2, but not other strata.

To test the difference in the slopes in Figure 2.5, we fit a line to each trend. To capture the difference between weekdays and weekends, we considered the daily reduction instead of the last 7-days average reduction. We modeled two periods, the first between March 20 and July 13 and the second between September 01 and October 12. We did not model the period between the sectorized lockdowns as the trends for strata one and two are not linear. Table 2.2 shows the results for the first period. The slopes for all strata are positive and significantly different from zero. The small standard errors also suggest that these estimates are significantly different from each other. The slope can be interpreted as the daily return to transit rate. For strata one, this rate is 0.24%/day, which is five times higher than the rate for strata five and six. We also show that the reduction of transactions on the weekends is less severe than for weekdays. For instance, for strata five and six, the reductions of transactions on the weekends is 77.5% (100% - 7.568% - 14.939%), but for weekdays, the reduction is 92% (100% - 7.568%).

The results for second period are shown in Table 2.3. The slope estimates are all positive

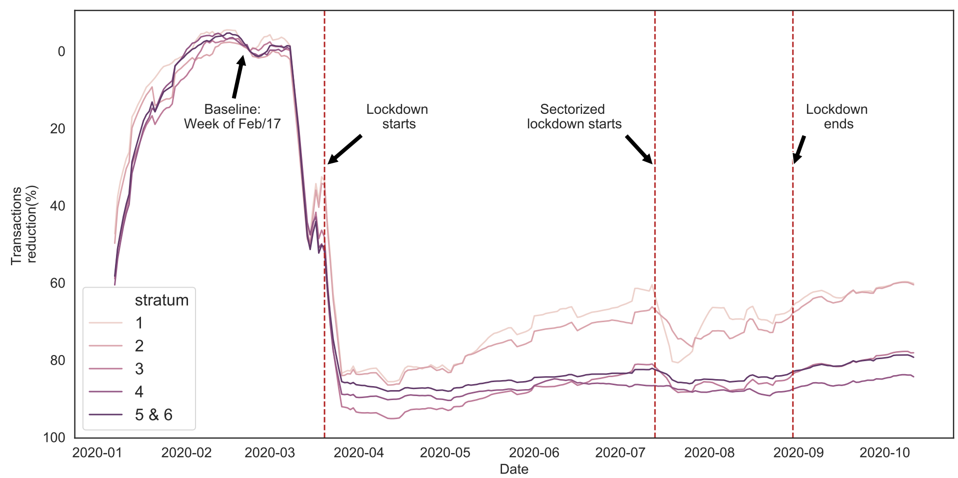

Figure 2.5: 7-days rolling average reduction of transactions (%) by strata. Reference is the week of February 17, 2020. Red dotted lines mark the start of the lockdown (Mar/20), the sectorized lockdown (Jul/13) and the end of the lockdown (Sep/01).

Table 2.2: OLS estimation before sectorized lockdownn. Dependent variable is the transactions reduction (%). Estimation period: Mar/20 to Jul/13 (99 days).

	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Strata 5 & 6
constant	12.110***	12.118***	3.054***	6.691***	7.568***
	(0.654)	(0.479)	(0.419)	(0.454)	(0.674)
slope	0.248***	0.195***	0.135***	0.044***	0.054***
	(0.010)	(0.007)	(0.006)	(0.007)	(0.010)
weekend	3.607***	4.933***	3.845***	11.296***	14.939***
	(0.688)	(0.503)	(0.440)	(0.477)	(0.709)
R-squared	0.877	0.900	0.852	0.862	0.831
Adj. R-squared	0.879	0.902	0.855	0.865	0.834
N	99	99	99	99	99

* $p < .1$, ** $p < .05$, ***$p < .01$

and significant but not significantly different from each other. The constant estimate shows the difference in the transaction reductions between low strata (1-2) versus middle and high strata (4-6) is between 15% and 25%. Since the slopes grow at the same rate, the disparity between low and middle/high strata remains stable in this period.

Table 2.3: OLS estimation after sectorized lockdown. Dependent variable is the transactions reduction(%). Estimation period:Sep/01 to Oct/12 (41 days).

	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Strata 5 & 6
constant	34.091***	32.154***	14.545***	8.740***	12.994***
	(0.998)	(0.792)	(0.708)	(0.935)	(0.880)
slope	0.095**	0.137***	0.118***	0.073*	0.089**
	(0.041)	(0.032)	(0.029)	(0.038)	(0.036)
weekend	7.783***	9.243***	11.247***	15.625***	17.134***
	(1.058)	(0.839)	(0.750)	(0.991)	(0.933)
R-squared	0.608	0.789	0.865	0.867	0.899
Adj. R-squared	0.628	0.799	0.872	0.873	0.904
N	41	41	41	41	41

* $p < .1$, ** $p < .05$, ***$p < .01$

2.4.2 Transactions at "Other" locations are growing faster than home and work locations.

This section focuses on how the COVID-19 lockdowns have affected the transactions by the location classification (home, work, and others). In Figure 2.6 we plotted the number of transactions per 1000 people by transaction location and strata. This variable was selected to compare the behavioral change of transit use during the recovery period. As expected, strata 4 to 6 have low values for transactions x 1000 people, and they are relatively constant until the end of the lockdowns. For strata one to three, the plot shows a drastic reduction at the beginning of the sectorized lockdowns for all transaction type. However, other strata only show a slight decrease. After the end of the lockdown restrictions, the plot suggests that transactions at "other" locations are growing faster than home and work trips for all strata, but it is steeper for strata one to three.

To test this last hypothesis, we fit these trends with a line to compare the value of the slopes. For this test, we only selected the data from September 1, until October 12. The results are shown in Table 2.4. The slopes are all positive and significant at the 95% level, except for work in strata one. For all strata, the slope for "other" transactions is greater than home and work. However, for strata one to three, it is significantly steeper than any other strata. Notice that for strata one, the slope for "other" is more than double the home transactions slope.

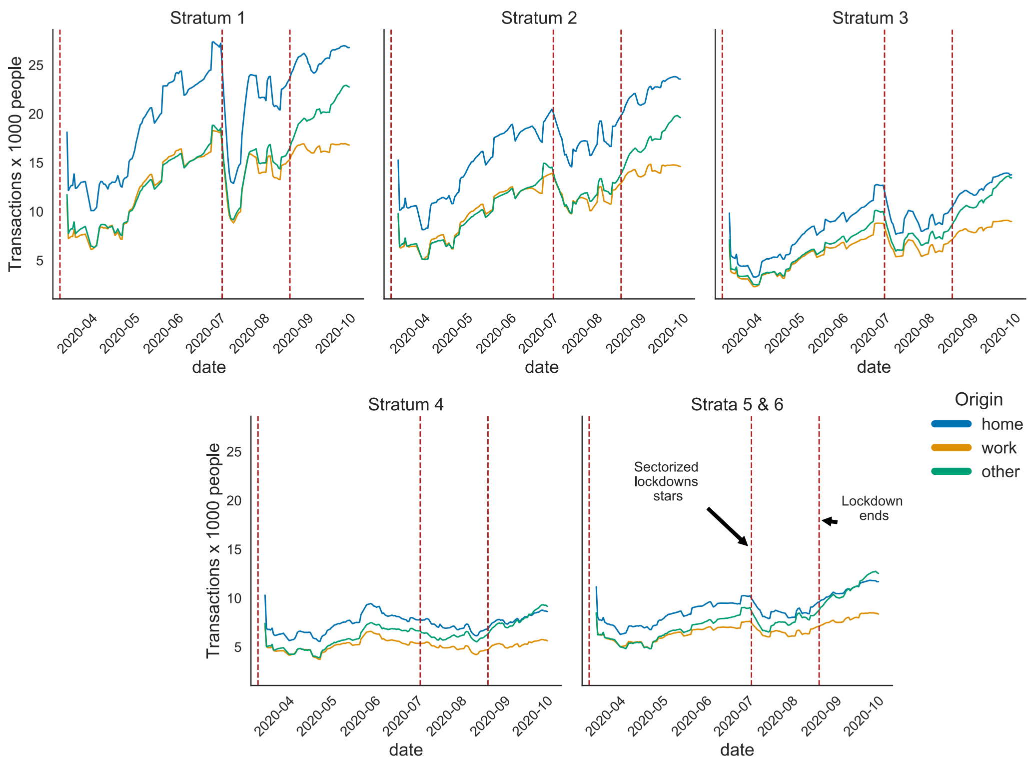

Figure 2.6: Transactions per thousand people by strata and location type in the recovery period. Red dotted lines mark the start of the lockdown (Mar/20), the sectorized lockdown (Jul/13) and the end of the lockdown (Sep/01).

2.4.3 The radius of gyration slightly decreases for lower strata but increases for higher strata

While other studies have shown that the Rg has decreased about 50% after the lockdowns [59], our study shows that when measuring Rg only for users that continue using BRT in Bogotá, their Rg slightly decreased for lower and medium strata and increased for higher strata, as shown in Figure 2.7 and Figure 2.8. Typically, the Rg is calculated with GPS tracks collected even if an individual stays at home. For the GPS case, it is possible to accurately measure an Rg of zero if a person stays at home, and therefore their observation is usually included in the average Rg estimation. As for public transit, the average Rg on a given day can only be calculated if an individual rides transit on a particular day, as it is mode-specific. Therefore, after the lockdown, our measure of Rg represents the average distance reached in public transit by users who remain using transit after the lockdowns. The average Rg slight decreases for strata one to three but significantly increases in other

Table 2.4: OLS estimation. Dependent variable is the transactions per 1000 population. For the categorical variable activity origin type, category home is the base. Estimation period: Sep/01 to Oct/12 (41 days), excluding weekends.

	Stratum 1: Trans x 1000 people	Stratum 2: Trans x 1000 people	Stratum 3: Trans x 1000 people	Stratum 4: Trans x 1000 people	Strata 5 & 6: Trans x 1000 people
constant	24.747***	20.752***	11.151***	7.264***	9.874***
	(0.155)	(0.120)	(0.088)	(0.072)	(0.057)
work	-8.430***	-7.069***	-3.518***	-2.279***	-2.604***
	(0.220)	(0.169)	(0.124)	(0.102)	(0.080)
other	-6.831***	-5.864***	-1.928***	-0.683***	-0.714***
	(0.220)	(0.169)	(0.124)	(0.102)	(0.080)
slope	0.049***	0.080***	0.071***	0.034***	0.052***
	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)
slope * work	-0.037***	-0.053***	-0.034***	-0.018***	-0.020***
	(0.009)	(0.007)	(0.005)	(0.004)	(0.003)
slope * other	0.069***	0.040***	0.035***	0.029***	0.037***
	(0.009)	(0.007)	(0.005)	(0.004)	(0.003)
R-squared	0.985	0.989	0.982	0.972	0.988
Adj. R-squared	0.985	0.989	0.982	0.973	0.989
N	123	123	123	123	123

* $p < .1$, ** $p < .05$, ***$p < .01$

strata. The average Rg seems unaffected by the sectorized lockdowns, but strata five and six have a decreasing trend.

Unlike the previous two analyses, the Rg is more stable for most strata, therefore fitting a line to the trend may not be appropriate because the slope is very likely to be zero. For this reason, we performed a two-sample t-test for mean comparison to test if Rg before and after the lockdown are significantly different. The null hypothesis is that the average Rg difference before and after the pandemic is equal to zero, and the alternative hypothesis is that it is different from zero. The after-lockdown data represents the days after the first day of economy re-opening. We calculated the Welsh t-test and used the Satterthwaite formula for the degrees of freedom approximation; this procedure assumes that each sample variance is unequal. The results show that for strata two, we fail to reject the null hypothesis, and therefore, there is no significant difference in the average Rg before and after the lockdowns. For strata one and three, there is a slight but significant decrease of the Rg. For strata four to six, we reject the null hypothesis in favor of the alternative hypothesis and conclude that the average Rg is higher after the lockdown for this population. The results are shown in

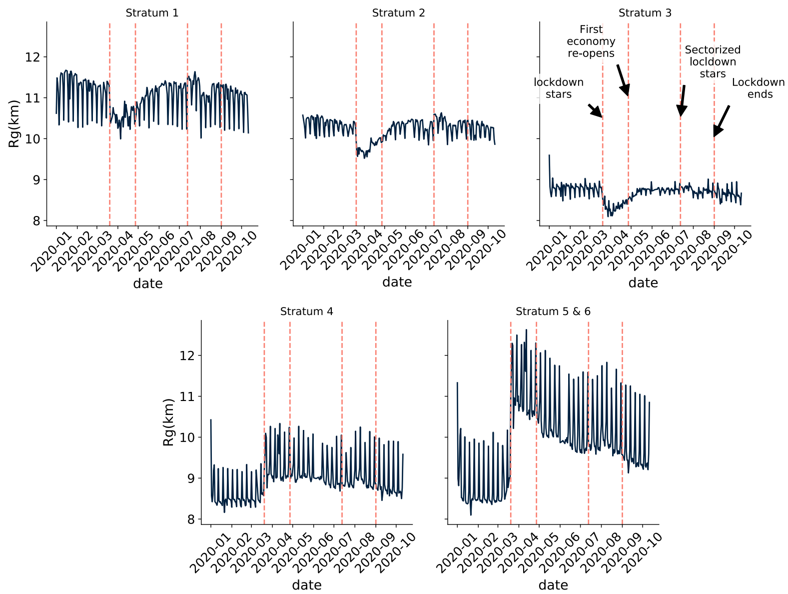

Figure 2.7: Average Radius of Gyration (Rg) in km by strata. Red dotted lines mark the start of the lockdown (Mar/20), the first economy re-opening (Apr/27), the sectorized lockdown (Jul/13) and the end of the lockdown (Sep/01).

Table 2.5.

Table 2.5: Average Radius of Gyration in KM mean difference t-test. Before time period is Jan/01 - Mar/20. After period is Apr/27 - Oct/12.

	Average Rg(Km)							
	Before			After				
	μ	σ	n	μ	σ	n	t	df
Stratum 1	11.19	0.42	79	11.04	0.36	155	2.58**	140.31
Stratum 2	10.32	0.21	79	10.28	0.18	155	1.37	136.86
Stratum 3	8.77	0.16	79	8.71	0.11	155	3.48***	116.68
Stratum 4	8.65	0.44	79	9.04	0.42	155	-6.38***	147.55
Strata 5 & 6	8.95	0.91	79	10.05	0.69	155	-9.39***	124.82

* $p < .1$, ** $p < .05$, ***$p < .01$

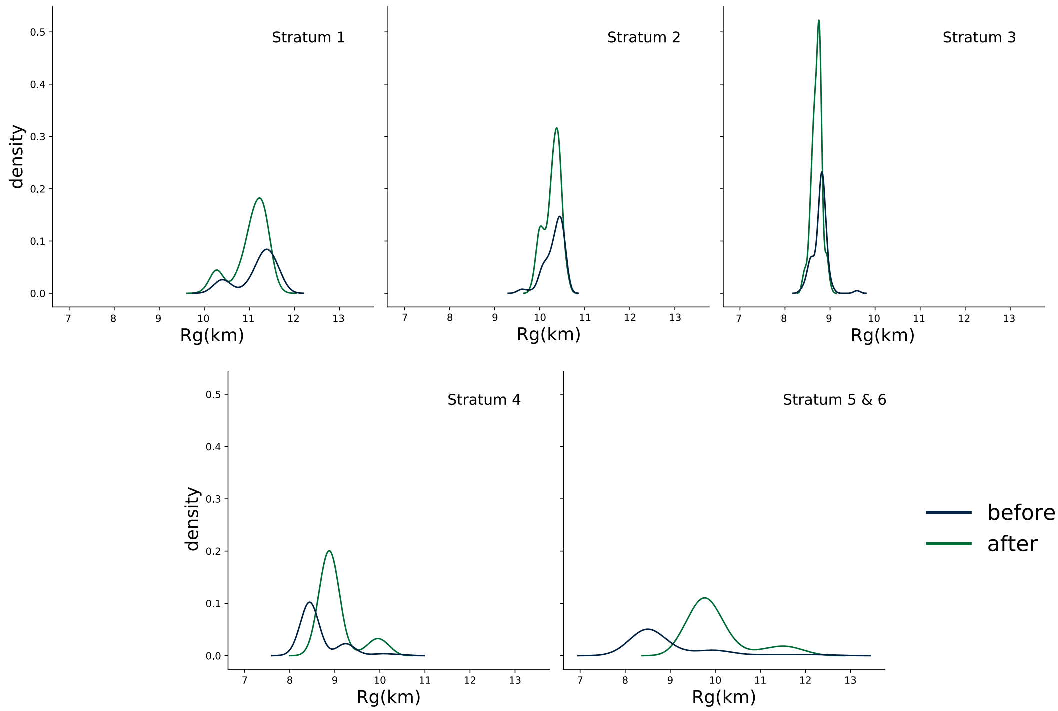

Figure 2.8: Distributions of Average Radius of Gyration (Rg) in km. Blue is the distribution before the lockdown. Green is the distribution after the first economy re-opening.

2.5 Discussion

The results show that strata one and two have different behavior than other strata and that middle and high strata behaved very similarly during the pandemic. Lower strata had the least reduction of public transit use after the lockdowns and returned to transit faster. These results may show that low strata have a more significant proportion of captive users who need public transportation to access jobs. In the context of Bogotá, public transit captivity can be understood in two ways. First, lower strata have a low rate of vehicle ownership. Therefore, the availability of other transportation modes is limited. According to the HTS (2019), strata one and two have 95.5 and 137.7 vehicles per 1.000 people, respectively, while the same rate is 593.3 for strata six. Second, the spatial distribution of the strata in Bogotá makes it challenging to switch to alternative modes such as walk or bike because low-income neighborhoods are further away from the downtown. Therefore, bike or walk may not be a viable alternative. Even bike ownership for lower-strata is lower than higher-strata, with 111 bikes per 1.000 people for strata one and 319 bikes per 1.000 for strata six (HST, 2019). On the contrary, middle and high strata have higher rates for personal and private vehicle ownership and are closer to job areas, and therefore it is easier to switch to sustainable modes such as walking and biking. These results could also be influenced by the fact that new bike lines are highly concentrated in the middle and higher stratum; these bike lanes were implemented to give transit users better and safer options to commute during the COVID-19 pandemic.

The sectorized lockdowns also had different impacts. The first set of neighborhoods affected by the lockdown were dense job areas and low strata neighborhoods, explaining the higher drop in strata one and two. During this first phase, reduction rates for stratum one were similar to middle and higher strata. However, as the restrictions eased for those neighborhoods, stratum one quickly returned to similar levels right before the sectorized lockdowns, while middle and higher strata remained equally low. Lockdowns in other phases seem to have had little effect on the BRT demand. As the economy re-opens, the number of transactions at home and work locations follows a similar trend. However, the growth rate for “other” transactions increased for every stratum. The increase of transactions in “other” locations could be influenced by the eased restrictions and the fact that the public transit remains with low capacity, yet it is still operating with 100% of the fleet. As the BRT has good connectivity and accessibility, the population may be using it more for other activities outside work.

As for the Rg, lower and middle strata had a slightly decreased before and after the lockdowns. This result shows that those users that remain in transit are reaching similar destinations pre- and post-the lockdowns. These trends do not seem to be affected by the sectorized schema. However, for higher strata, the Rg increases during the lockdowns. This result may show that most of the low-distance high strata BRT trips were replaced by other modes, and only long distances remain, consistent with findings in (João Filipe Teixeira, Miguel Lopes, 2020). strata five and six also showed a decline in the average Rg, which might have been influenced by the increased transactions in “other” locations. One contribution of this study is the calculation of the Rg specific to a transportation mode, in this case, the BRT. Other studies calculate the Rg based on the visited locations regardless of the transportation mode used to access those locations. These studies showed a significant drop in the Rg, but they did not compare it for the individuals who remained traveling during the COVID-19 pandemic. For transit operation and even for other transportation services, this information is vital to adapt the supply to current demand needs.

Our study highlighted that the reduction of the BRT system’s demand does not distribute equally among strata. Therefore, the operations of the fleet should not remain equal to the pre-lockdown schedules. To promote equity in the transit services, transit agencies need to move the resources to where they are more needed, such as vulnerable populations that solely depend on transit. A more in-depth analysis of the BRT origin-destination pairs may help identify which pairs to prioritize considering the strata, the transaction volume, and the activity type. While our results show a differential impact based on strata, the methodology only captures frequent users before the pandemic and analyzes their trends after the pandemic. We did not account for frequent users that have become frequent after the lockdowns as we did not have a reference point for them. Additionally, while we assumed that different strata have different rates of taking transit, we did not consider that longer distances to BRT stations may reduce the probability to take BRT. As a result, the current methodology may be overestimating the representation of BRT use in the edges of BRT stations' catchment areas. However, these results provide a general understanding of how transit demand is shifting during the COVID-19 pandemic for different groups.

2.6 Limitations and Future Work

The current methodological approach to infer home and work location did not consider work instability. We defined work instability as a worker with varying working locations such as construction workers or people in the informal job market, which is likely to affect the results for low strata. For instance, one user can have a home, work, and study location, but the methodology will only capture one or the other. Further research in this area may help us understand and classify more complex tours. We did not consider home or work change. This assumption is particularly important because the COVID-19 lockdowns may have also impacted the relocation rates. However, data on home and work re-locations before the 2020 is not available as a baseline. Additionally, we assumed that probability vector of a transaction belonging to one stratum depends only on the strata distribution within the catchment area, and the rates of transit for each stratum. We did not consider other factors such as distance, which may give greater weight to blocks closer to the stations. Lastly, frequent users were estimated using data pre-lockdowns only, and our results showed the trends of such users, however we did not account for new frequent users that may have become frequent and active after the lockdowns. We also assumed that catchment areas are the only source of variations, however it is possible that some home transaction may fall outside the catchment area given some informal services that are used as feeder routes. For instance, bicitaxis, and cars may help collect users for areas that may fall outside the catchment area. Information on informal transit was not available, and therefore was not accounted for this study.

2.7 Conclusion

We classified transactions of frequent users by strata using a probability vector, which is a function of the rate of BRT use and the population in the catchment area of the station. This classification showed a similar aggregate strata distribution compared to the 2019 HTS, which validated our inference methodology. The results from this study showed a differential impact of the COVID-19 pandemic in BRT use by strata. Lower strata showed the least reduction of transit use comparing to middle and higher strata. At the beginning of the lockdown period, lower strata returned to transit faster than any other stratum. By the end of the lockdown restrictions, the number of trips for "other" locations was significantly higher than those of work trips, which adds extra challenges to transit operators. This study also showed that transit users in lower strata were reaching similar destinations as before the lockdown as measured by Rg. However, for higher strata, the average Rg increased during the lockdown, suggesting that those that keep using transit after the lockdown needed it for longer trips. This study's results can help transit agencies better allocate resources to improve the level of service and accessibility to both home and work for vulnerable populations.

Chapter 3

Public Transit Demand Prediction during Highly Dynamic Conditions: A Meta-analysis of state-of-the-art models and open-source benchmarking Infrastructure.

ABSTRACT[1]

Near real-time demand prediction is an important input for dynamic bus routing. While many researchers have developed a myriad of complex methods to predict short-term transit demand, the applications have been limited to short, stable time frames and a small number of stations in one geography. How these methods perform in highly dynamic environments has not been studied, nor has their performance been systematically compared. We built an open-source infrastructure with five common methodologies, including econometric and deep learning approaches, and assess their performance under stable and highly dynamic conditions. Given the sudden demand changes of highly dynamic conditions and the complex spatiotemporal correlation of transit demand, we also implemented and tested the performance of the models with an "adaptive training" and "multi-output" model design, in contrast to traditional training and single-output models. We use time series from smartcard data to predict demand for the following day for the Bus Rapid Transit (BRT) system in Bogotá, Colombia. The dynamic conditions in our time series include a month-long protest and the COVID-19 pandemic. Both conditions have triggered unexpected and uncertain closures of

[1]This chapter has been reproduced with the permission of my co-authors Marta C. González, and Joan L. Walker

multiple transit stations and drastic shifts in demand. The results show that most tested models perform relatively similarly in stable conditions, with forecasting errors varying from 8.5 to 12%. The benchmark showed that all models performed significantly worse in both dynamic conditions relative to stable conditions. In the month-long protest, the increased prediction error ranged between 14 to 24%. Similarly, during the COVID-19 pandemic, the increased prediction error ranged from 12 to 82%. Notably, in the COVID-19 pandemic condition, a long-short-term memory (LSTM) model with an adaptive training and multi-output design stood out by outperforming other models and adapting faster to disruptions. The prediction error stabilized within approximately 1.5 months, whereas other models continued to exhibit higher error rates even a year after the start of the pandemic. With our open-source codebase infrastructure, our aim is to lower the barrier for other researchers to replicate and reproduce models, facilitate a collective effort within the research community to improve the benchmarking process, and accelerate the advancement of short-term ridership prediction models.

3.1 Introduction

The motivation behind dynamic routing is to be responsive to real-time demand and improve transit services and reliability. Short-term demand prediction supports the decision-making process in dynamic routing by providing ridership forecasts to allow the allocation of resources ahead of time. In a highly dynamic condition, a situation that is both unexpected and uncertain, these decisions become more relevant. As the demand changes drastically, public transit services need to adjust their fleet operation to meet the new needs. Examples of highly dynamic conditions are the COVID-19 pandemic and the impacts of long-term protests. The induced lockdowns during the COVID-19 pandemic resulted in drastic changes in travel demand, and the duration of these changes was (and still is) uncertain because of the novelty of the virus. As a result, many transit agencies responded by reducing and modifying transit services. Furthermore, while most protests might be foreseeable in the short term, the aftermath is not; protests might last for weeks or months and result in damages to the public transit infrastructure, forcing temporary station closures and transit services modifications. However, the literature on short-term ridership prediction methods has been limited to short, stable time frames and a small number of stations in one geography. How they perform in highly dynamic environments has not been studied, nor has their performance been systematically compared.

The literature on short-term ridership prediction has had exponential growth of proposed methods over the last decade, including statistical models [70, 29, 101], machine learning [102, 62, 28, 104, 22, 19], deep learning [109, 120, 64, 63, 92, 10, 20, 46, 43, 130, 7], Bayesian approaches [87], optimization [37, 124, 126], and various other methods such as grey theory

[53, 118], gravity models [93], Kalman filtering [127, 55] and combinations of two or more of this approaches [56, 68, 131, 128]. This large number of available methods presents an increasing challenge to researchers in comprehending, implementing, and comparing their performance. This growing pool of models makes it increasingly difficult to conduct reliable comparisons, leading to benchmarks that typically include only a small subset of models. Moreover, the heterogeneity of these subsets across the literature hinders the ability of the research community to identify relevant literature, even for stable conditions. This scenario highlights the need for a standardized framework for implementing and benchmarking the various models, facilitating collaborative efforts towards advancing the field.

The objectives of this research are twofold: first, to provide an open-source codebase infrastructure that uses a publicly available dataset to create standard, reliable, and statistically rigorous benchmarks, and second, to use this infrastructure to compare the performances of state-of-the-art models in highly dynamic conditions. To achieve these objectives, we propose an open-source codebase containing five widely used models, including two statistical models – ARIMA and SARIMA) – and three Deep Learning Models – Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), and the Long-Short-term memory network (LSTM) – which are frequently used in the literature on short-term ridership prediction. These methods are substantially different in how they structure and process data series data and are widely used for short-term ridership prediction models and other time series applications.

We use more than five years of data of the Bus Rapid Transit (BRT) system in Bogotá, Colombia, which is publicly available and can be used by other researchers. This dataset contains two highly dynamic conditions, a month-long protest in November and December 2019 and the COVID-19 pandemic since March 2020. Given the sudden demand changes of highly dynamic conditions and the complex spatiotemporal correlation of transit demand, we also test the performance of the models with an "adaptive training" and "multi-output" model design. The adaptive training model re-trains the model as new information becomes available, which might improve model performance in highly dynamic conditions. The multi-output model aims to consolidate the prediction of individual stations in one model, capturing possible spatial and temporal correlations across stations.

The remainder of the paper is organized as follows. Section 2 summarizes the relevant literature for short-term public transit demand forecasting. Section 3 explains the meta-analysis's methodology, including appropriate research methods and modeling strategies. Section 4 presents the results of the meta-analysis, and section 5 concludes with a discussion and conclusion.

3.2 Literature Review

This section focuses on the five methodological approaches to predicting short-term transit ridership and the two modeling designs. In the context of highly dynamic conditions, we also reference literature pertaining to demand prediction under special events such as holidays, concerts, and sports events. We focus primarily on aggregate prediction, which aggregates demand at the station or bus route level.

Parametric models, such as ARIMA and SARIMA, have been traditionally used to model time series data. The aim is to model a stationary process where the mean is zero and the variance is constant. The models are composed of an auto-regressive part (AR) and a moving average part (MA) to capture historical data and past prediction errors. The "I" and "S" stand for integrated and seasonality, which tries to account for trends and seasonal patterns. The main advantages of these models are the solid statistical background and the default calculation of confidence interval. Additionally, [32, 9] proposed an Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model to relax the constant variance assumption of the stationary process, which allows to model heteroskedasticity in time series analysis. A SARIMA model was used by [70] to forecast monthly demand using ten years of training data. [29] use an ARIMA + GARCH model to model demand volatility, using 15-minute aggregation for one month of data and three transit stations with high passenger demand.

Deep Learning models have also been widely used for time series analysis in the last decade. For short-term transit demand, there have been three main deep learning architectures: Multilayer perceptron (MLP) [109, 120], Convolutional Neural Networks (CNN) [64], and Recurrent Neural Networks (RNN) [63, 92]. The MLP model is a fully connected feed-forward neural network with at least three layers, the input layer, the hidden layer(s), and the output layer. For time series, the input layer is composed of the past observation; however, given that it is a fully connected network, there is no explicit time dependency. Inspired by image classification, the CNN model connects inputs with predefined temporal dependency (e.g., observations within the same week) [10]. RNN allows outputs of some nodes to affect subsequent inputs to the same nodes, which is a more realistic representation of time. In RNN, the most common cells are the Long-short term memory (LSTM) [46] and the Gated Recurrent Unit (GRU) [20], which maintain hidden states to filter relevant information and store long-term dependencies.

There are two main model designs to predict short-term ridership. Single versus multioutput models and the adaptive training strategy. Single-output models train individual transit stations, and multi-output models train one model to predict the output of all stations at once. Parametric and machine learning models have developed multi-output models; the main objective is to find the correlation of one variable in multiple outputs. In the parametric case, they are known as Vector Autoregressive Models (VAR) [58]. Multi-output models have been used in other fields, such as the energy forecast prediction [89] and the air quality prediction [135]. The main advantages of these models are that they capture the spatial and temporal correlation, and the modeler only needs to train one model instead of multiple individual models for each time series. In transportation research, multi-output models have been recently used for the bike-sharing system demand [69], bus travel time prediction [81], and public transit passenger prediction [117].

Static models maintain the estimated parameters for prediction, while adaptive training models (also known as online models or continuous learning) modify the parameters when new information becomes available. In this paper, we refer to adaptive training as ”online training”. The online training idea comes from the problem that artificial neural networks forget past information when trained for a new task, known as catastrophic forgetting [41]. In practice, ARIMA and SARIMA models update the model parameters when new information is available, as suggested by the default settings of multiple implementations [94, 66]. However, this aspect is not underscored in the short-term ridership prediction. In machine learning, this modeling strategy has been explored in other fields, such as energy consumption [1], medical research [60], recommendation systems [83], and spam detection [100], but it has not been applied to the short-term demand prediction problem of public transit. One of the advantages of online training is the adaptability to unpredictable and uncertain changes and the ability of these models to support real-time decision-making, which is relevant for highly dynamic conditions.

Previous methods focus mainly on short-term demand prediction under stable conditions. However, demand prediction during big events has received some attention. [116] develops an early warning system that detects abnormal passenger outflows to predict abnormal passenger inflow in the future and cross reference they're finding with large-scale events. Similarly, [47] proposed a clustering mechanism and a dynamic time wrapping to account for multiple social events such as holidays and sporting events. Lastly, [133] implemented a naïve Bayesian-based transition model to forecast demand under unplanned events. These methods, however, have two main shortcomings. First, they were only validated during known periods of disruption, such as holidays and large-scale events, but not under highly dynamic conditions. Second, these methods do not take full advantage of multi-output models, as they only consider specific stations where special events are likely to happen.

The contributions of this research are twofold. First, we create an open-source codebase that implements major model architectures and establish a common benchmarking platform to compare model implementations reliably and accurately. Second, we use this codebase to perform a meta-analysis of state-of-the-art models and common modeling strategies to compare their performance in terms of model accuracy during stable and highly dynamic conditions in a data set that spans multiple years and stations.

3.3 Methods and Data

The research objective is to compare the prediction accuracy of multiple model implementations currently proposed in the literature for stable and dynamic conditions. In this section, we describe the selection of methods that encompass most of the state-of-the-art methodologies for the short-term transit demand prediction problem. Then we describe the data set we use in our experiments and its two highly dynamic conditions. We also describe the experiments, metrics of analysis, and preprocessing steps.

3.3.1 Method selection

We have included at least one reference for each major methodology in Table 4.1, which has been used to inspire our implementation. We prioritized papers with high citation counts, clear explanations of the model, and relatively canonical model implementations for each category. This approach ensured that we were building upon well-established and rigorously tested methodologies with proven effectiveness with short-term ridership prediction. To ensure the comparability of the models, we applied the same preprocessing and feature engineering techniques across all implementations, even if it meant deviating from the exact methodology described in some of the selected papers. This was necessary to ensure that the methods were appropriately adapted to our specific data, while maintaining consistency in the evaluation process. For instance, [63] implements three LSTM networks to model same day, daily, and weekly demand. In our implementation, demand is aggregated to the day, so we only use one LSTM network. While some studies have found that external data such as weather, land use, and social data improve accuracy, we avoid using them because this research objective is not to test the accuracy of external data but the performance of the methods themselves.

In [70], the authors used the SARIMA model to account for yearly seasonality patterns. [29] proposed a GARCH in addition to the ARIMA model to account for the demand volatility. For this research, we only train the model with the ARIMA part since the GARCH component only affects the estimation of the demand volatility but not the point estimate. These two papers selected a model based on an information criterion. [29] used the Akaike Information Criterion (AIC) score and [70] used the Bayesian Information Criterion (BIC) score for some pre-defined models. Since model selection is a time-consuming task, especially when training a model for an individual station, we use a stepwise algorithm [49] to expedite the model selection using the AIC score.

Reference [120] used a three-layer neural network. The input layer contains as many nodes as input variables, and the output layer size is the forecast window. The hidden layer size was taken from the default implementation of NeuroShell 2, which is the average of the input layer and the output layer size. Our implementation of the MLP model uses the same model structure. [63] proposed many data sources and feature engineering procedures. However,

Table 3.1: Short-term ridership prediction selected References for the open-source codebase infrastructure implementation - Attributes as reported in references

Study	Methods	Number of Stations	Time Frame	Aggregation	Accuracy	Preprocessing
[29]	ARIMA	3	1 Month October 2012	15 mins	MAPE: 3.39%	Log Transformation
[70]	SARIMA	Full System	10 years 2004-2014	Monthly	MAPE: 7.13%	Log Transformation
[120]	MLP	1 Line	1 Month May 2008	15 mins	MAPE: 5.04%	Empirical Mode Decomposition
[64]	CNN	1091 Lines	8 Months Mar-Oct 2016	Hourly	SMAPE: 20.98%	Unclear
[63]	LSTM	3	5 Months	10 mins	SMAPE (DL-N) 16.91%	Min-Max

for the purpose of this research, we select the implementation of their Deep-Learning Nearest Part Architecture (DL-N) model, which only takes the information of previous timesteps. However, as the focus of our research is on daily demand forecasting based on a full week demand, our implementation takes into account daily and weekly cycles. The implementation used a unique LSTM layer with 32 units, which is maintained in our implementation. The CNN architecture consists of one convolutional layer with 256 filters and a dilatation factor of 7 to represent the weekly cycle.

3.3.2 Experiments

To evaluate and compare the performance of the selected models, we conduct four sets of experiments varying the model designs in two dimensions.

To compare the models, we conduct four sets of experiments varying two critical design dimensions that can affect the performance of the models under highly dynamic conditions. The first dimension pertains to the adaptive learning strategy, which we refer to as "online training". In this approach, the model is re-trained as new information becomes available. re this approach with static training, which estimates a model using only the test data, and generates predictions using the latest available information. However, the model is not re-estimated with new data. In the second design dimension, we compare a single-output model to a multi-output model. With the single-output strategy, we train each transit station's time series data individually. In the multi-output strategy, we use the time series data of all stations as input to predict the demand for all transit stations simultaneously.

The set of experiments is summarized in Table 3.2. To train ARIMA and SARIMA models, we use [94] python implementation. However, the implementation is restricted to univariate time series, and always requires updating the model parameters with new information. Given these constraints, we only test the ARIMA and SARIMA models for the single output and online training experiment.

Table 3.2: Experiments

Model Design	Single output	Multioutput
Static Training	MLP CNN LSTM	MLP CNN LSTM
Online Training	MLP CNN LSTM ARIMA SARIMA	MLP CNN LSTM

3.3.3 Data

The data encompasses the daily transactions of 147 stations over five years (August 2015 to May 2021) in the BRT system of Bogotá, Colombia, as shown in Figure 3.1. The BRT system is the city's main mass public transportation system, with over 2.5 million daily transactions during stable conditions (pre-COVID-19). There were two highly dynamic conditions during this period—a major national protest in November and December 2019 and the ongoing COVID-19 pandemic since March 2020. Both events have triggered unexpected and uncertain closures of multiple transit stations. This dataset is publicly available at: https://datosabiertos-transmilenio.hub.arcgis.com.

For the four sets of experiments, we train a baseline model with data from April 2015 to August 2018. The test period starts in September 2018 and until April 2021. We use the baseline mode for the online training strategy and update it every timestep. For all the experiments, we normalize the data using the min-max approach because it normalizes elements that will fall between zero and one and maintains the meaning of zero (zero transactions). Given the data considerations, we use daily aggregation to fit ARIMA and SARIMA. For every experiment, we use the information from the past 21 days to predict the daily demand of the following week. We include common temporal variables to account for cyclical patterns. We use a dummy variable to signal Saturdays and Holidays and a sine and cosine function to encode yearly and weekly trends.

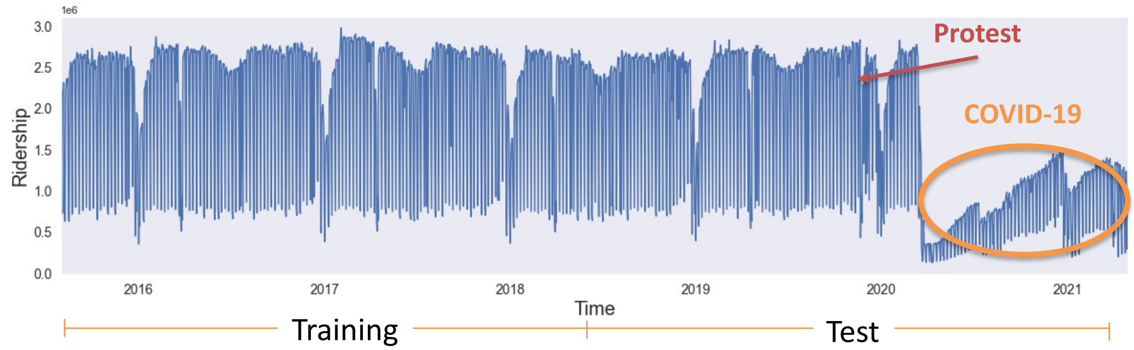

Figure 3.1: BRT daily aggregated demand training period from August 2015 to May 2021. Training Period: Aug 2015 - Jul 2018. Test Period: Aug 2018 - May 2021. Protest: Nov and Dec 2019. COVID-19: March 2020 to May 2021

3.3.4 Metric of Analysis

Most of the performance indicators currently used in the literature for short-term demand forecasting are the Root Mean Squared Error (RMSE), Mean Absolute Value (MAE), and Mean Absolute Percentual Error (MAPE). We will instead use the Mean Arctangent Absolute Percentual Error (MAAPE) [57]. RMSE and MAE measures depend on the magnitude of the observations, making them useful to compare different methods on the same dataset but not as effective when comparing one method across multiple datasets. The MAPE solves that problem by estimating a percentage error instead of an error. In the MAPE, the error can be considered a slope $|\widehat{y}-y|/|y|$, and its limit when y approaches zero is positive infinity. Therefore, the error is bounded from 0 to $+\infty$. The MAAPE limits the bounds of the error by considering the error as an angle. The limit of the angle as y tends to 0 is $\pi/2$. Considering target values of zero is relevant to our problem because a transit station might not be open at all time intervals, especially when bus stations are forced to close due to external forces. Additionally, because not all transit routes/stations are open at the same time, when we train a unique multi-output model, we need to standardize the hours of operations of all stations, which may also result in zero target values.

In this research, we proposed a system-wide level that tests the performance of the entire transit system at every time step. Unlike current trends in the literature that test accuracy at the individual level, this metric serves as a summary metric to test a model's performance. The advantage of having a system-wide metric is that it captures the performance in a wider range of transit behaviors, making the metric more generalizable. Therefore, it can also use to compare performances in multiple geographies. In this research, the average error of the entire transportation system is given by:

$$MAAPE_t = \frac{1}{s \ * w} \sum_{j}^{s} \sum_{i}^{w} \arctan(\frac{\hat{y}_{s,w,t} \ - y_{s,w,t}}{y_{s,t,w}}) \tag{3.1}$$

Where t is the aggregation period, s is the number of stations, and w is the forecast window. Therefore $\hat{y}_{s,w,t}$ represents the predicted value of station s, for w forecast windows in the time period t.

3.4 Results

In this section, we summarize the main results from the experiments. First, we include the system-level results and focus on the models' performance during stable and highly dynamic conditions. Additionally, we include an analysis of the temporal variation of the MAAPE for all models. The statistical analysis results for each experiment can be found in the appendix.

To understand the evolution of the error metric, we estimate a daily system-wide MAAPE for all experiments, as shown in Figure 3.2. The plot is smoothed using the rolling MAAPE for the last seven days. During stable conditions, most models perform similarly in all four sets of experiments, although the CNN and ARIMA models showed consistent underperformance. During stable conditions, some peaks in the MAAPE metric correspond to the end of the year holidays, eastern, and long weekends. These results also show that all models and all experiments experience a sharp increase in the accuracy metric for two highly dynamic conditions, but the initial shock is lower in the month-long protest condition. The online training experiments reveal a faster decrease in the MAAPE during the COVID-19 pandemic; however, there is a noticeable difference in the expected time for the MAAPE to stabilize. On average, models in the single-output and online training experiment take 3 months, while the LSTM model in the multi-output and online training takes 1.5 months. Lastly, while the CNN model usually underperforms in most experiments during stable conditions, this is not true in the COVID-19 condition.

To evaluate the statistical performance of the models and experiments, we regress the $MAAPE_{n,t}$ with respect to the COVID and Protest conditions. These are dummy variables where one represents that the period t belongs to the given condition and zero otherwise. Since some temporal variables are not well captured in the models, we also control for Saturdays and holiday effects.

$$MAAPE_{n,t} = \alpha_n + \ \beta_{1,n}\,(covid_t) + \beta_{2,n}\,(strikes_t) + \beta_{3,n}\,(temporal_t) + \varepsilon_n \tag{3.2}$$

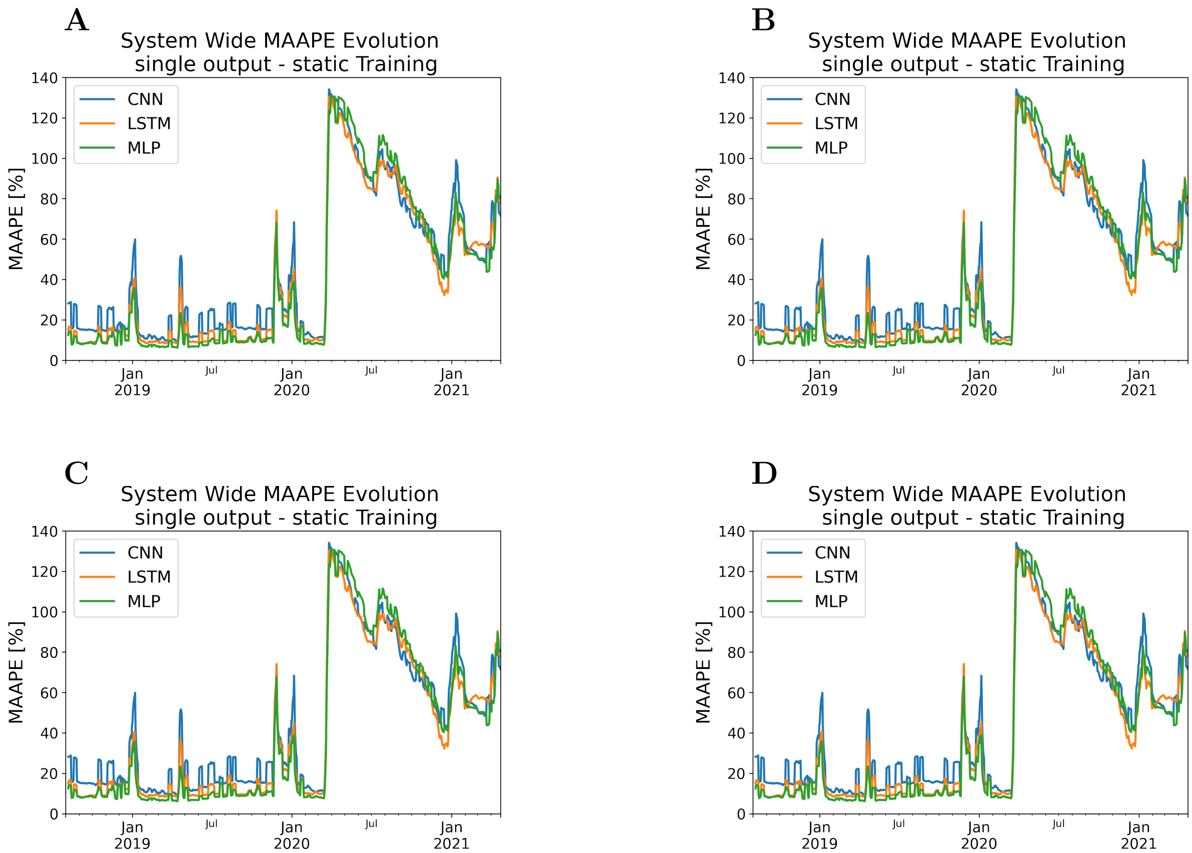

Figure 3.2: Daily System-Wide MAAPE evolution for Test Period. A. Single-Output and Static Training. B. Multioutput and Static Training. C. Single-output and Online Training. D. Multioutput and Online Training

Where n represents the experiment, and t is the aggregation period. The parameter α_n represents the average accuracy of the model during stable conditions, $\beta_{1,n}$ is the added error associated with the COVID-19 pandemic, and $\beta_{2,n}$ is the added error associated with the strikes. The advantage of linear regression is the standard error estimation, which allows comparing if the difference in the MAAPE metric of the two models is statistically significant. The results of this estimated can be found in the appendix.

3.4.1 Stable Condition

For stable conditions, the single-output and online-training LSTM model is the best system performance (benchmark results shown in Figure 3.3). This model outperformed six out of the 12 experiments and has no statistical difference from the other 5. The daily MAAPE is

0.080, which can be interpreted as an 8.1% error, and its standard deviation is 0.007. Other models performed relatively similarly, for example, the single-output and online-training DENSE model. However, the LSTM version is preferred because of the lower magnitude of its standard error. The multi-output version of the online-training LSTM model also performed well with a 9% accuracy. The worse performing model in the CNN model, with the higher error (16%). For eight out of the 12 experiments, this model performed significantly worse.

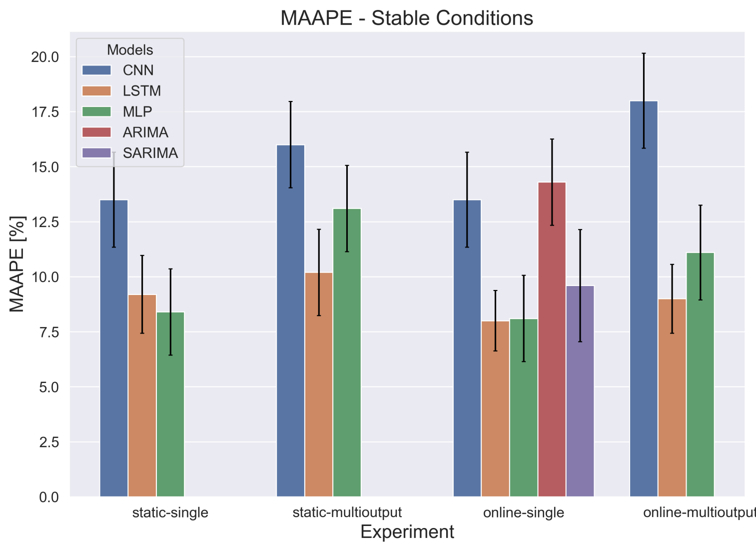

Figure 3.3: MAAPE Performance Benchmark in stable conditions

3.4.2 COVID-19 Condition

For the COVID-19 highly dynamic condition, the best performing model was the multi-output, online training LSTM model, as shown in Figure 3.4. This model outperformed 9 out of the 12 experiments and has no statistically significant difference only with the single-online-LSTM and the multi-static CNN models. The marginal increase in the MAAPE during the COVID-19 pandemic is 11.9%, and a standard error of 0.012. The worse-performing models were the single-output and static training models, with an average MAAPE increase of 75%. From the plots, the online models tend to have a higher spike at the beginning of the COVID-19 condition but leveled up relatively fast after a few months.

3.4.3 Protest Condition

The performance during the one-month-long protest in November-December 2019 are presented in Figure 3.5. The single-output, online-training CNN model performs best, with a MAAPE of 17.5%. However, this accuracy is only significantly better than multi-output LSTM for both static- and online training. The multi-output, online-training LSTM model

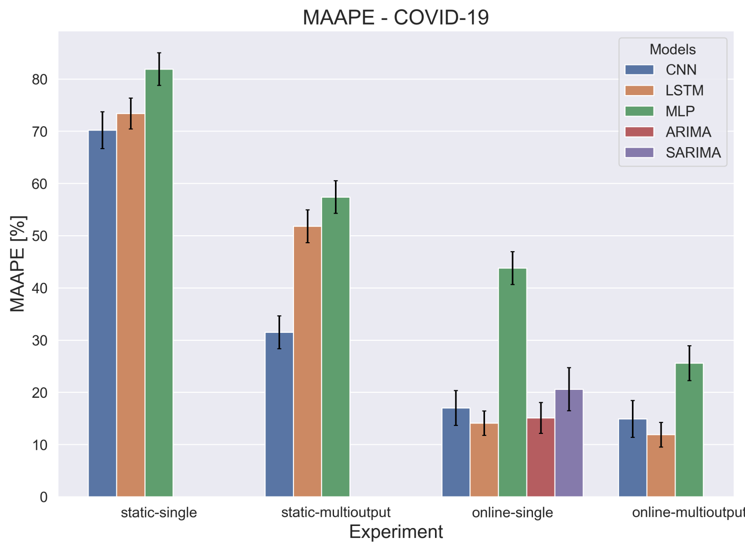

Figure 3.4: MAAPE Performance Benchmark during COVID-19 condition

performs significantly worse than three of the other experiments, and the marginal MAAPE for the protest condition is 24.4%.

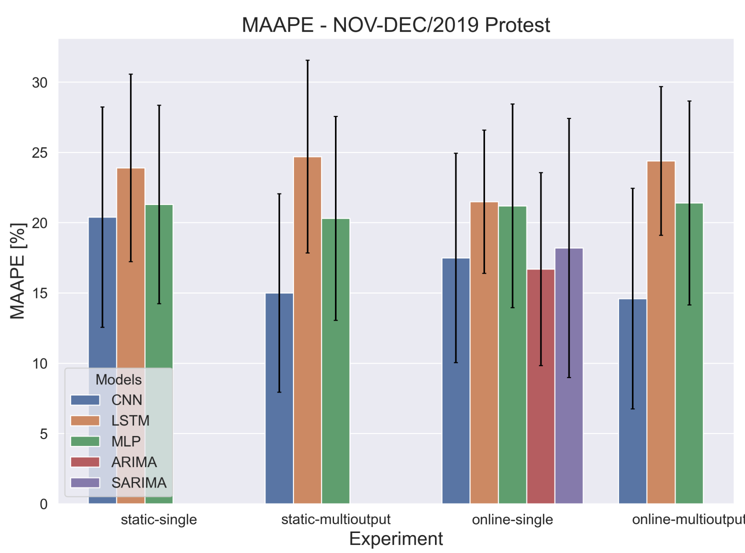

Figure 3.5: MAAPE Performance Benchmark during protest condition

3.4.4 Other temporal Variables

Figure 3.6 shows that the MAAPE is higher during Sundays Holidays, even though these variables are added in the explanatory variables to account for the differentiated behavior. The single-output static-training dense model seems to outperform the predictions. The average MAAPE increase for this model is 12.1%, and the difference is statistically significant for 10 out of the 12 experiments. The only model that performs similarly is its online version, with an increased MAAPE of 12.4%. However, the MAAPE is better for the static-training experiments than the weekdays during the COVID-19 condition. This difference is less obvious in the online-training models.

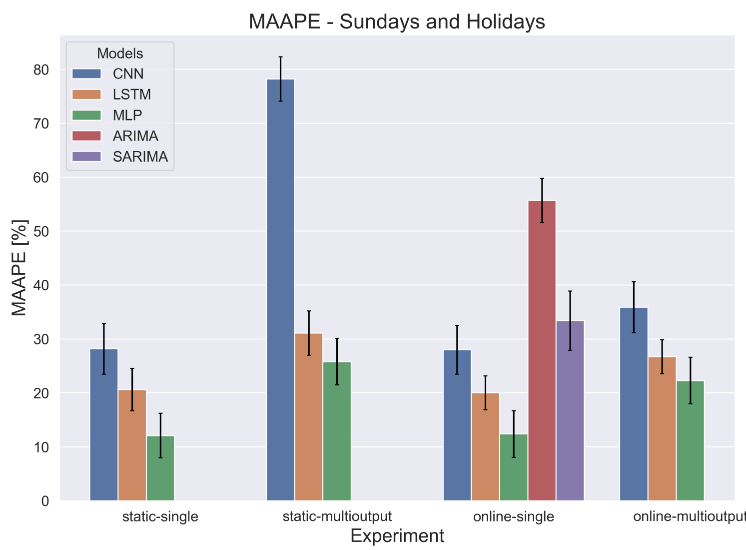

Figure 3.6: MAAPE Performance Benchmark - Sundays

3.4.5 Running times

An important metric to consider is running times, especially as aggregation periods become smaller. While there is no direct comparison for running time within the strategies tested in this research, we try to estimate the average training and prediction time, as shown in Figure 3.7. We plot the average time to train and predict an individual transit station for single-output models. For multi-output models, the result is the time it took to estimate a unique model with all transit stations. From the plot, training times for multioutput models are almost negligible. However, for single-output models, the total training would need to be multiplied by the number of transit stations. We did not plot training times for ARIMA and SARIMA models because they are one order of magnitude different. The LSTM model is almost 5 times slower than the rest of the models; however, this result is expected because the model is a recurrent neural network that is not possible to parallelize. For all models and experiments simulation, running times are 1.0 seconds or less. Multioutput models can take 5 times longer than single-output models, but the extra running time is overweighted by the fact that multioutput models predict the demand of the entire transit system at once.

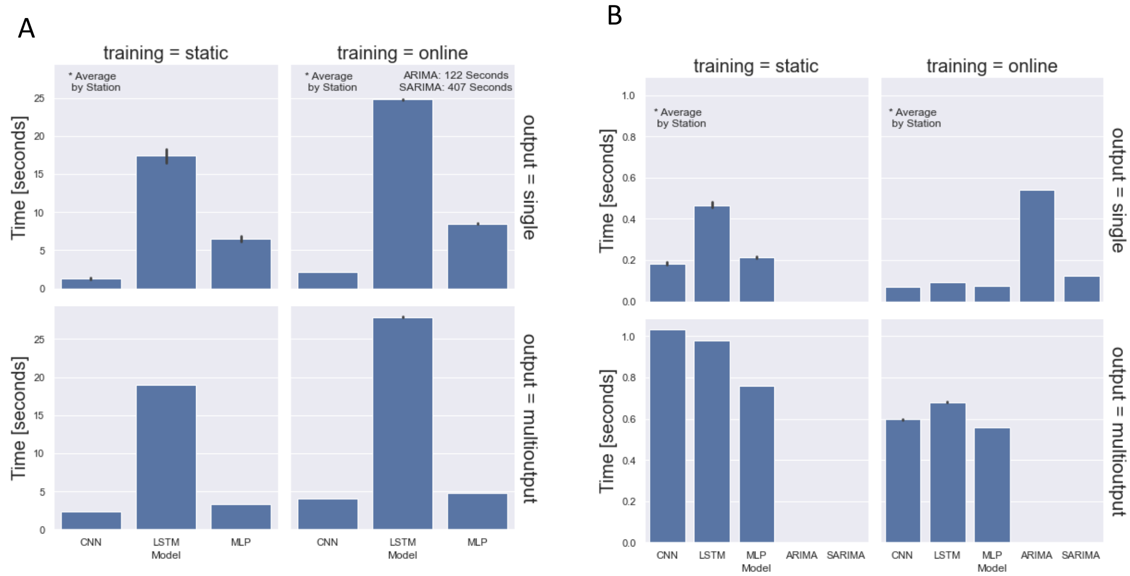

Figure 3.7: Average training and simulation running times A). Training B). Simulation

In summary, multi-output and online training models perform better than single-output static-training models. For stable conditions, LSTM, MLP, and SARIMA outperform other models. Single-output models have better performance; however, there is no statistically significant difference with multi-output models. Similarly, online models perform slightly better than static models, but the difference is not statistically significant. Given that the running times for multi-output models are more convenient, we select multi-output, online LSTM as state-of-the-art performance for stable conditions.

The multi-output, online LSTM model achieves state-of-the-art performance for stable conditions. The multioutput model is more convenient for training times and performs similarly to single-output models (no statistically significant difference). The online strategy, while not significantly different from the static strategy, is preferred because it performs better in one highly dynamic condition. The LSTM model is chosen because its performance is slightly better than the MLP model. For the COVID-19 condition, the best-performing model is the multi-output, online LSTM model. Online models perform significantly better than static models, and the LSTM performs slightly better in the multi-output strategy than in the single-output strategy. Additionally, the LSTM model learns new patterns faster than other models. For the long-month protest, none of the models perform significantly better

Table 3.3: State-of-the-art performances.

Condition	Output	Training	Model	Training Running Time	MAAPE (C.I)
Stable	Multiple	Online	LSTM	27 sec	0.09($\pm$0.016)
COVID-19	Multiple	Online	LSTM	27 sec	0.12($\pm$0.023)
Protest	Multiple	Online	CNN	4 sec	0.20($\pm$0.078)

than the others; however, we select the CNN model, as it tends to perform slightly better than other models. State-of-the-art performances are summarized in Table 3.3.

3.5 Discussion and Conclusion

This paper performed a meta-analysis to compare and critique the performance of main modeling strategies and state-of-the-art model architectures, including both econometric and deep learning approaches, during stable and highly dynamic conditions. The dynamic conditions in our time series included the COVID-19 pandemic and a month-long protest. Both conditions triggered unexpected and uncertain closures of multiple transit stations and shifts in demand that forced changes in service operations. We showed that all models and modeling strategies in this analysis perform significantly worse during highly dynamic conditions, as shown in Figure 3B and Figure 3C. The multioutput online LSTM model not only performs best during COVID-19 conditions, but it also is the model that most rapidly adapts to highly dynamic demand. While the performance of any of the implementations during the month-long protest is significantly better than the others, the multioutput online CNN model performed slightly better.

Our results show that online models improve model accuracies faster, but it takes at least 1.5 months to learn new demand patterns. This result is significant because the duration of a highly dynamic condition is uncertain, yet decisions to modify transit operations need to be fast and made in timely manner. In this research, we show the potential of the online training strategy to learn new patterns and improve model accuracy; however, more research is needed to learn these patterns faster. While models for stable conditions do not necessarily benefit from online training, the unexpected nature of highly dynamic conditions makes a static strategy insufficient. Waiting to implement new models when these conditions happen is usually time-consuming and might not support time-critical decisions.

The research on short-term demand prediction has grown exponentially in the last few years, making selecting state-of-the-art models more difficult for researchers and practitioners. As a result, the benchmarking process sample a few models with no certainty that these implementations produce state-of-the-art performances. In this research, we have selected five common methodologies and modeling strategies to compare and critique state-of-the-art models, but we need a collective effort as a research community to build a common platform to share and benchmark implementations. Our research shows that some implementations have similar performances, and other factors, such as running time, might be more relevant to determine a model strategy and implementation. A common benchmarking infrastructure also allows researchers to identify shortcomings in model implementation and propose new strategies to solve them. For instance, our results also show that predictions during holidays are significantly worse, even if exogenous variables are added to account for these effects. Additionally, when stations are closed, models still predict a non-zero demand. Therefore, a common and systematic benchmark process has the potential to expand the field of short-term ridership prediction faster and more reliably.

This research is an initial step to explore online training strategies, not only in the context of highly dynamic conditions but also in stable conditions. Additionally, we have created an open-source codebase to benchmark models systematically and reliably. Our aim is that future researchers use and contribute so that we can collectively improve the short-term ridership prediction problem. Future research can focus on model implementations or strategies that improve the average time to learn new patterns and improve predictions in highly dynamic conditions. More research is also needed to account for station closures and their effect on surrounding stations. Current implementations assume all stations are open and non-zero values are predicted, even when a station is closed.

The open-source codebase to replicate the results of this research is available at:

`https://github.com/jdcaicedo251/transit_demand_prediction.`

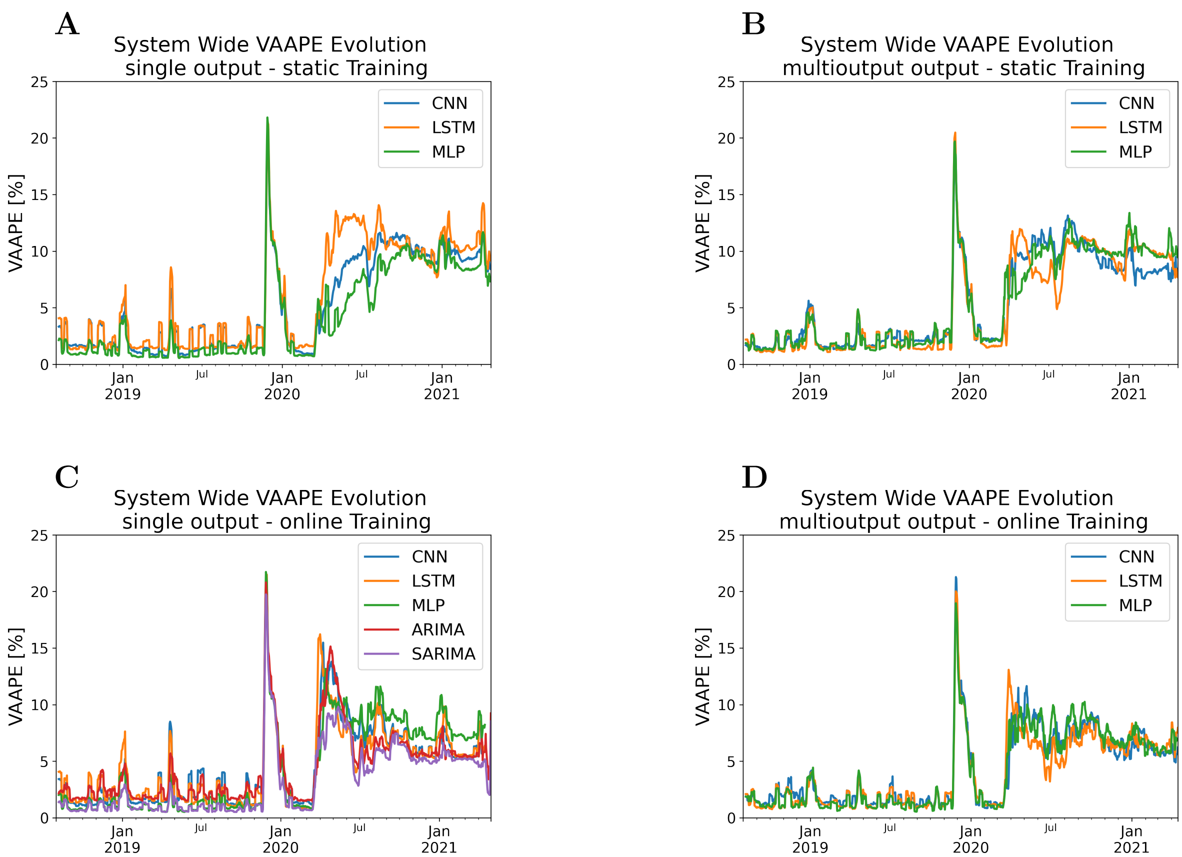

Figure 3.8: Daily System-Wide VAAPE evolution for Test Period. A. Single-Output and Static Training. B. Multioutput and Static Training. C. Single-output and Online Training. D. Multioutput and Online Training

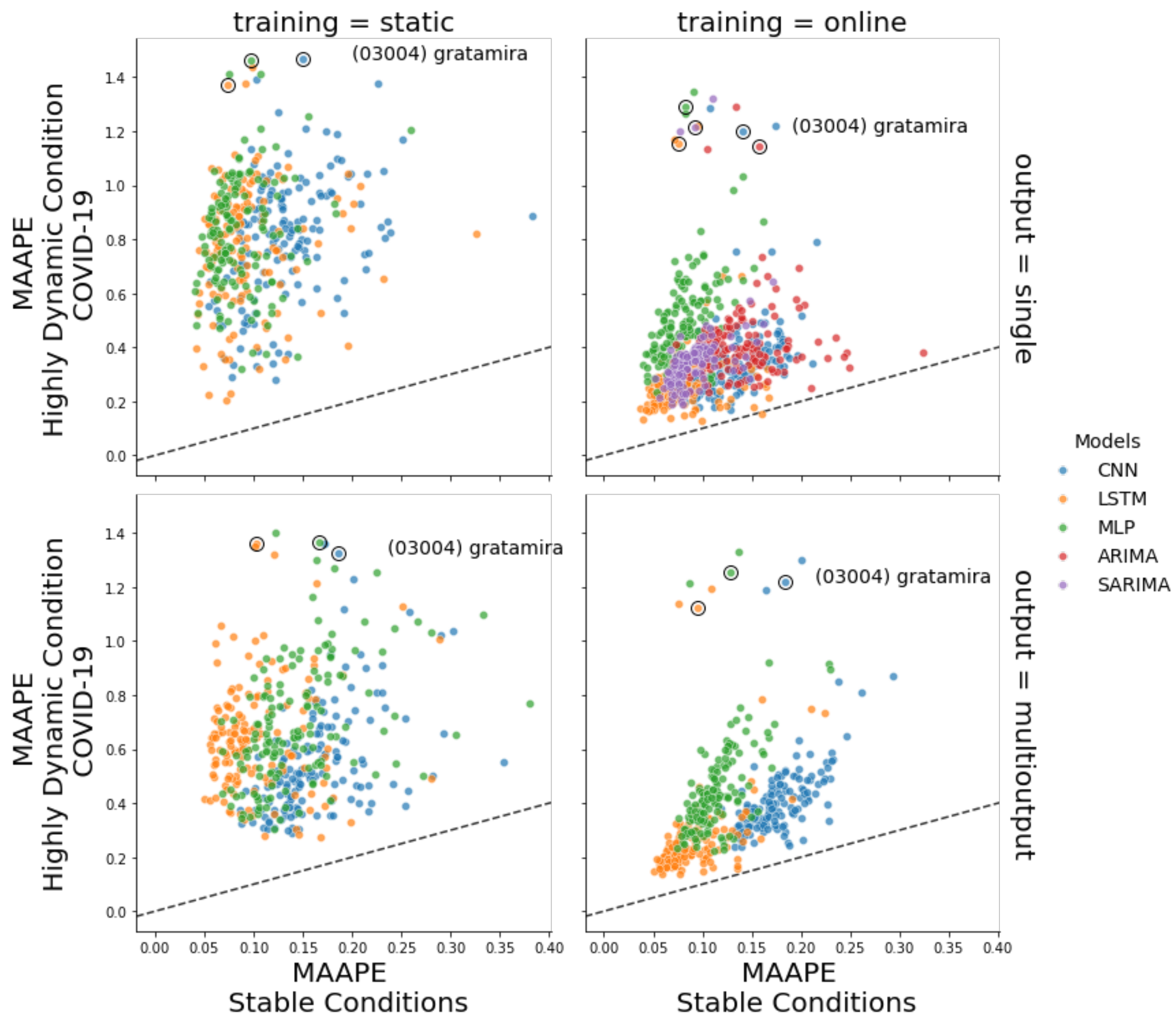

Figure 3.9: MAAPE Stable vs. COVID-19 Condition. Each observation represents a BRT station

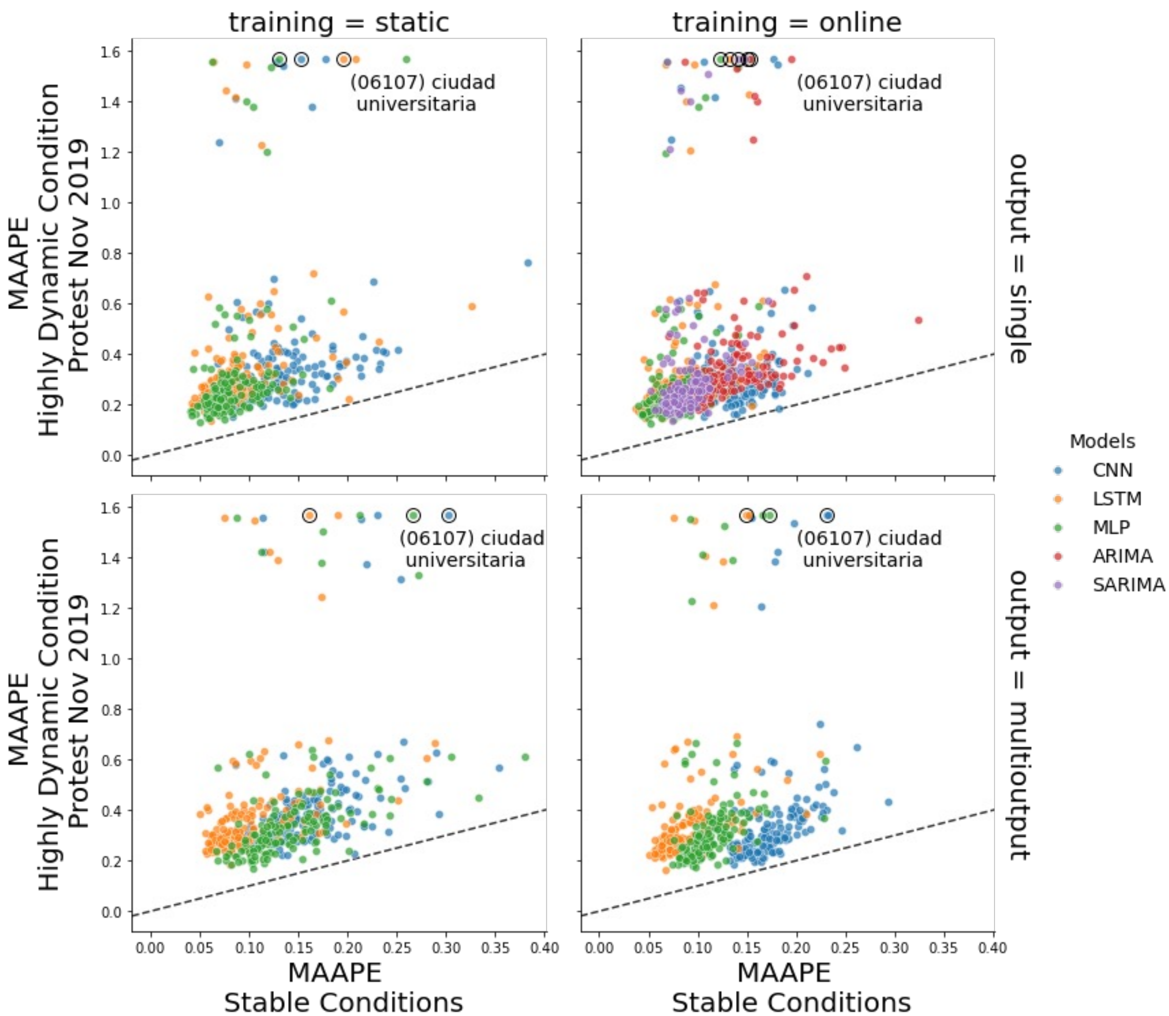

Figure 3.10: MAAPE Stable vs. Protest Condition. Each observation represents a BRT station

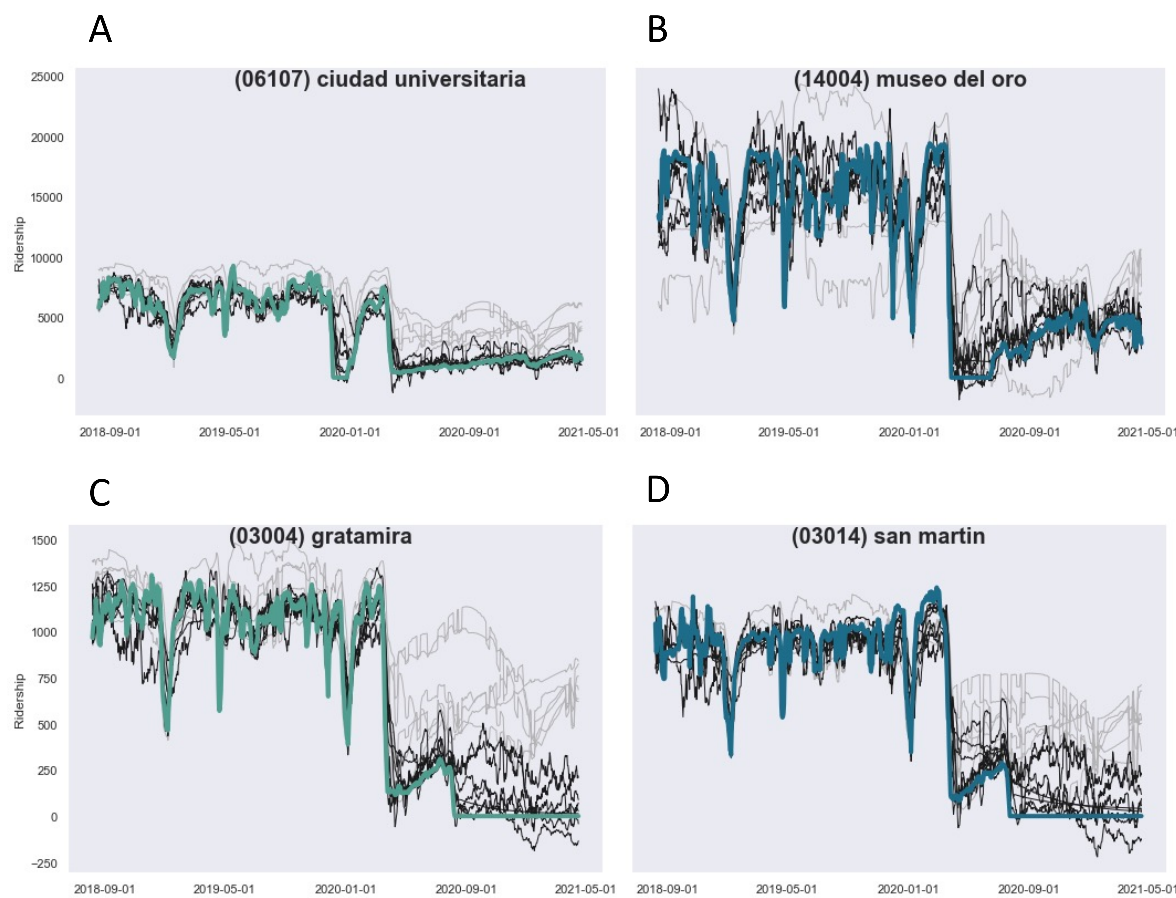

Figure 3.11: Temporary Closed Stations examples. Light-gray: Static Models. Dark-gray: Online Models

Table 3.4: Statistical Analysis - Single-output and Static Training

	CNN	LSTM	MLP
Intercept	0.135***	0.092***	0.084***
	(0.011)	(0.009)	(0.010)
covid	0.702***	0.734***	0.819***
	(0.018)	(0.015)	(0.016)
protest	0.204***	0.239***	0.213***
	(0.040)	(0.034)	(0.036)
saturday	0.004	0.007	0.012
	(0.026)	(0.022)	(0.024)
holidays	0.282***	0.206***	0.121***
	(0.024)	(0.020)	(0.021)
covid:saturday	-0.110***	-0.134***	-0.214***
	(0.042)	(0.035)	(0.038)
covid:holidays	-0.347***	-0.310***	-0.379***
	(0.037)	(0.031)	(0.034)
R-squared	0.667	0.752	0.761
R-squared Adj.	0.665	0.751	0.760
N	997.0000	997.0000	997.0000

Table 3.5: Statistical Analysis - Multioutput and static Training

	CNN	LSTM	MLP
Intercept	0.160***	0.102***	0.131***
	(0.010)	(0.010)	(0.010)
covid	0.315***	0.518***	0.574***
	(0.016)	(0.016)	(0.016)
protest	0.150***	0.247***	0.203***
	(0.036)	(0.035)	(0.037)
saturday	0.047**	0.019	0.081***
	(0.024)	(0.023)	(0.024)
holidays	0.782***	0.311***	0.258***
	(0.021)	(0.021)	(0.022)
covid:saturday	-0.053	-0.055	-0.330***
	(0.037)	(0.037)	(0.039)
covid:holidays	-0.360***	-0.136***	-0.141***
	(0.033)	(0.033)	(0.034)
R-squared	0.669	0.634	0.623
R-squared Adj.	0.667	0.632	0.621
N	997.0000	997.0000	997.0000

Table 3.6: Statistical Analysis - Single-output and Online training

	ARIMA	SARIMA	CNN	LSTM	MLP
Intercept	0.152***	0.109***	0.135***	0.080***	0.081***
	(0.009)	(0.013)	(0.011)	(0.007)	(0.010)
covid	0.139***	0.193***	0.170***	0.141***	0.438***
	(0.014)	(0.020)	(0.017)	(0.012)	(0.016)
protest	0.183***	0.233***	0.175***	0.215***	0.212***
	(0.032)	(0.045)	(0.038)	(0.026)	(0.037)
saturday	0.022	0.023	0.007	0.009	0.013
	(0.021)	(0.030)	(0.025)	(0.017)	(0.024)
holidays	0.547***	0.348***	0.280***	0.200***	0.124***
	(0.019)	(0.027)	(0.023)	(0.016)	(0.022)
covid:saturday	0.059*	-0.037	0.021	0.065**	-0.139***
	(0.034)	(0.048)	(0.040)	(0.028)	(0.039)
covid:holidays	0.045	-0.009	-0.019	0.045*	-0.110***
	(0.030)	(0.042)	(0.036)	(0.025)	(0.034)
R-squared	0.630	0.302	0.289	0.398	0.487
R-squared Adj.	0.628	0.297	0.285	0.394	0.484
N	990.0000	990.0000	990.0000	990.0000	990.0000

Table 3.7: Statistical Analysis - Multioutput and Online training

	CNN	LSTM	MLP
Intercept	0.180***	0.090***	0.111***
	(0.011)	(0.008)	(0.011)
covid	0.149***	0.119***	0.256***
	(0.018)	(0.012)	(0.017)
protest	0.146***	0.244***	0.214***
	(0.040)	(0.027)	(0.037)
saturday	-0.010	0.014	0.023
	(0.026)	(0.018)	(0.025)
holidays	0.359***	0.267***	0.223***
	(0.024)	(0.016)	(0.022)
covid:saturday	0.053	0.065**	0.029
	(0.042)	(0.028)	(0.040)
covid:holidays	0.064*	0.001	0.191***
	(0.037)	(0.025)	(0.035)
R-squared	0.379	0.412	0.453
R-squared Adj.	0.375	0.409	0.450
N	990.0000	990.0000	990.0000

Chapter 4

A Novel Modeling Framework for Short-term Ridership Prediction During Station Closures

ABSTRACT[1]

Temporary station closures create challenges for predicting ridership accurately, as they can suddenly increase demand at other locations. Yet, current short-term ridership prediction models fail to account for these closures, yielding inaccurate forecasts. This research uses station closure information to improve short-term ridership prediction in neighboring stations of a station closure. We propose a novel modeling strategy combining Graph Information and the attention mechanism to capture the spatiotemporal correlation among stations and time simultaneously. We compare the performance with state-of-the-art models in an open-source benchmark codebase using five years of Bus Rapid Transit (BRT) boarding transactions in Bogotá, Colombia. Our analysis includes more than 10,000 station closures lasting longer than two hours. Compared to other state-of-the-art models, the proposed model demonstrates a reduction in prediction error ranging from 3.3 to 23.5% across multiple metrics and modeling strategies when predicting ridership in the nearby stations of a station closure. This research makes three key contributions: recognizing the importance of station closures in short-term ridership prediction models, proposing a new modeling architecture that uses this information to improve overall model accuracy, and adding the model architecture to the open-source benchmark platform for future comparisons. This study offers practical implications for transit agencies in making faster and better decisions during temporary station closures by providing more accurate forecasts, which contributes to developing a resilient public transit

[1]This chapter has been reproduced with the permission of my co-author Joan L. Walker

system capable of adapting quickly to disruptions.

4.1 Introduction

Public transit systems are vulnerable to planned or unplanned disruptions that can lead to temporary closures of transit stations. Such closures can have a significant impact on travel demand patterns, as travelers may be forced to alter their travel plans or use alternative transportation modes[99, 125, 84]. For instance, in Bogotá alone, over a five-year period, there were more than 10,000 closures lasting more than two hours, highlighting the need to account for this factor in short-term demand prediction models. Smartcard transaction data from Bogotá also revealed that station closures led to increased demand at nearby stations. However, to effectively optimize resource allocation, transit agencies require precise knowledge of when, where, and to what extent the demand of neighboring stations will be affected by the closures.

Despite the impact and frequency of station closures on nearby stations, current short-term ridership prediction models do not incorporate this information, resulting in non-zero demand predictions even when stations are closed. Furthermore, it remains unclear whether station closures have an impact on the prediction error of state-of-the-art models in nearby stations, limiting the use of these models to inform decisions during station closures.

This research aims to close this gap by quantify the impact of station closures on the prediction accuracy of nearby stations of state-of-the-art-model and incorporate station information closure to improve the prediction accuracy. We propose a combination of Graphical Neural Networks (GNNs) and the Transformer architecture to improve information transfer between neighboring stations, particularly in the event of station closures. A GNN model is a type of neural network that can process and analyze graph data structures by learning representations of nodes and edges and performing computations on them [38]. The attention mechanism is a technique that assigns weights to different parts of input data to enable a model to focus on the most important information [6]. Given that the transit network has complex spatiotemporal correlations, we use the graph structure to encode spatial relationships and enhance it with the attention mechanism to simultaneously pay attention to the closure of another station for a given time duration. In this research, we also test the efficiency of different model strategies to incorporate station closure information and analyze different data processing techniques that might improve overall model accuracy. We demonstrate the superiority of our model using the Bus Rapid Transit (BRT) boarding transaction data from Bogotá, Colombia, covering almost five years, and comparing the performance with models implemented in an open-source codebase for short-term ridership prediction.

We hypothesize that the short-term ridership predictions for neighboring stations will be significantly worse in these circumstances. Additionally, we hypothesize that the graph structure and the attention mechanism can enhance the temporal and spatial communication between stations and improve the forecasting accuracy of neighboring stations in the event of station closures. To test our hypothesis, we use Bus Rapid Transit (BRT) boarding transaction data from Bogotá, Colombia, covering almost five years.

The contributions of this research are three-fold. Firstly, we highlight the importance of accounting for station closures and suggest modeling strategies to integrate this information into short-term ridership prediction models. Secondly, we develop a novel modeling framework that incorporates station closure information and improves the prediction accuracy of neighbouring stations impacted by a station closure. Thirdly, we implement our modeling approach on an open-source benchmark platform, making it possible for future researchers to reproduce, compare, improve, and build upon our model. The results of this research have the potential to support rapid decision-making processes for transit agencies during station closures, increasing service resilience and enhancing user satisfaction. Specifically, our proposed modeling framework and strategies enable transit agencies to allocate resources proactively and respond promptly to unexpected closures, enhancing system resilience. Additionally, the improved short-term ridership predictions for neighboring stations increase users' confidence in the reliability of the transit system, further enhancing service satisfaction.

The remainder of this paper is organized as follows. Section 2 provides an overview of the relevant literature on short-term ridership prediction. Section 3 describes our proposed modeling framework, including the modeling strategies used to account for station closures and the experimental setup designed to evaluate the performance. In Section 4, we present the results of our research. Finally, in Section 5, we summarize and discuss the research findings and implications.

4.2 Literature Review

Short-term ridership prediction has traditionally been approached as a time series problem, with modeling techniques including statistical methods [29, 70], machine learning [19, 102, 28], and deep learning models [109, 120, 64, 63, 92]. However, these implementations have typically focused on a single or few stations without considering the spatial correlations between adjacent stations.

To address this gap, a combination of Graphical Neural Networks (GNN) and Recurrent Neural Networks (RNN) has been proposed to capture spatial and temporal correlations among transit stations [43]. In this context, the attention mechanism is utilized to assign weights to relevant information for each model component, starting with the temporal component and subsequently with the spatial component, in a sequential rather than a single step manner. [129, 136]. For short-term demand prediction, a sequential approach may not adequately capture the complex interdependence between spatial and temporal correlations of transit demand.By simultaneously applying the attention mechanism to both spatial and temporal components, our approach is able to capture more complex relationships and dependencies between these factors, ultimately leading to improved model accuracy. However, this approach has not been proposed in the context of short-term ridership prediction.

The attention mechanism has also been used to estimate the importance of different model components in the modeling framework. For instance, [67] developed a multigraph model and used an attention mechanism to capture the importance of each graph, and [44] used the attention mechanism to model the importance of multiple temporal bi-LSTM encoders.

Although most studies focus on relatively stable conditions, only a few have attempted to capture irregularities in transit demand. [62] utilized a radial function network to predict non-regular demand 30 minutes prior to special events, such as concerts or holidays. They also found that their model performs well for unplanned events, such as train signal failures, to predict the out-demand of the stations. [39] developed a model to capture abnormal passenger flow but did not explore the causes of these abnormalities. Additionally, [35] focused on failures that caused people to be stuck in transit stations and require evacuation services. While these papers investigate the effects of planned and unplanned disruptions, they focus on the affected station, mostly in alighting predictions, and do not consider the impact on surrounding stations. Furthermore, none of these studies offer modeling strategies to address the possibility that some transit stations might be closed during a certain period.

In summary, GNNs have been effectively employed to capture spatial correlations between transit stations, while RNN-based architectures have been utilized to capture temporal correlations. Attention mechanisms have been introduced to filter relevant information in either one or both components, with the sequential application used when employed in both. Despite the extensive research in this area, current state-of-the-art models fail to account for the possibility of transit stations being closed due to planned or unplanned events, nor have these models sufficiently explored the impact of station closures on neighboring stations. These limitations hinder our understanding of the effects of such disruptions on transit demand, indicating a need for further research in this area.

4.3 Methods

In this section we present the problem statement, a new modeling framework to capture spatiotemproal relationships, modeling strategies to process station closure information, data processing tecniques and the experiments and metrics of analysis that are used to draw the conclusions of this research.

4.3.1 Problem Statement

The objective of the short-term ridership prediction model is to forecast the transit demand for the following time step as a function of historical data, exogenous variables, and metadata if available. In this research we focus in the one-step-ahead model of the following form:

$$y_{i,t+1} = f(y_{i,t-k:t}, X_{i,t-k:t}, u_{i,t-k:t+1}, s_i) \tag{4.1}$$

Where $y_{i,t+1}$ is transit demand at the period $t+1$ for station i. $y_{i,t-k:t}$ is the historical observations of station i, and look back period k. $X_{i,t-k:t}$ are exogenous historical observations, which is this case correspond to observations in other transit stations in the look back period k. $u_{i,t-k:t+1}$ is know past a future data such as date information. Lastly, s_i represents metadata of the station such as the location. Notice that we can also express equation 1 in matrix form, where the dependent variable is a vector of shape (N) representing the ridership of all stations N in $t+1$.Additionally, $y_{i,t-k:t}$ and $X_{i,t-k:t}$ can be concatenated to represent the historical observations for all stations.

4.3.2 Modeling Framework

The objective of the modeling framework is to be able to simultaneously capture relevant spatial and temporal correlations to improve the prediction accuracy during station closures. Inspired in [112], we adopt a similar modeling framework with modifications to incorporate graph data information and to estimate attention scores in two dimensions - the spatial and temporal dimension - as shown in Figure 4.1.

Given the problem statement and the nature of the time series data, we only used the decoder part in [112]. Encoder-decoder architectures are usually used in case where the target and input sequence are of varying lengths, such as language translation tasks, where a sentence in one language can have a different length in another language. In this case, the look back window k, and the step prediction $t+1$ is the same for every prediction target, as defined in the problem statement. Therefore, this approach is appropriate for our problem statement.

The main components for the framework are the positional encoding, the input hidden representation, the 2D Graph masked multi-head cross attention, and the station closure information.

Positional Encoding

In [112], the authors recognized that the attention-based architecture does not naturally encode the sequential nature of the data, therefore they proposed a positional encoding

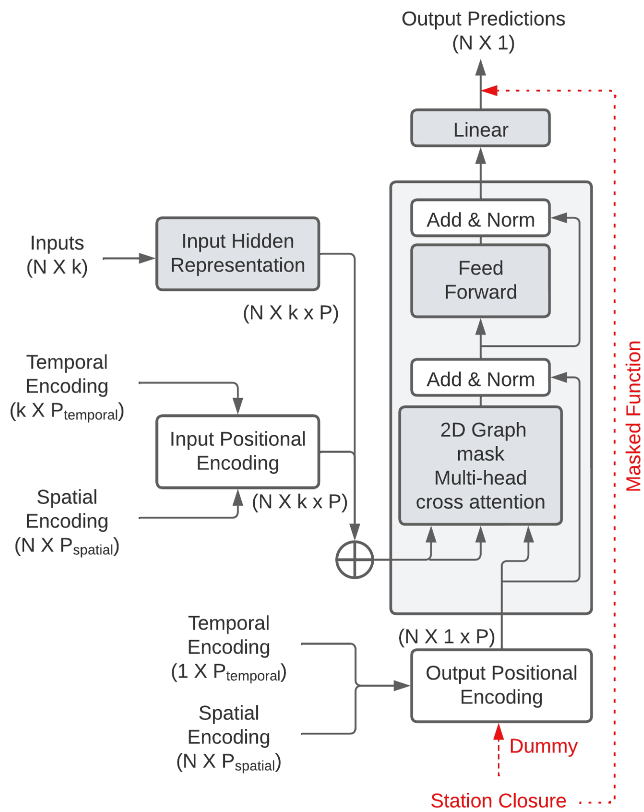

Figure 4.1: G-2DT Modeling Framework. Inspired in [112]

using sine and cosines functions. In this work, we use a similar approach using the date and time information for the temporal encoding, and the coordinates of the stations as the spatial encoding.

The temporal component represents the cycles present in the data, and other time dependent variables. In this research, we consider the annual, weekly and daily cycles. For each cycle we create a sine and cosine function, whose frequency depends on the length of the cycle. Additionally, we include dummy variables for different demand patterns such as weekdays, Saturdays, and Sundays/Holidays. The shape of the temporal positional encoding is $(k, p_{temporal})$, where k is the look back period, and $p_{temporal}$ are the number of temporal positional encoding variables.

The spatial encoding aims to account for spatial differences between stations. For the purpose of this paper, we only include the latitude and longitude coordinates, but more spatial representation can also be added. Given the variation of the coordinates is relatively low (e.g. For BRT station in Bogotá, latitude vary from 4.53 to 4.76, and longitude vary from -74.04 to 74.21), we increase the variation by scaling them to a user-defined min and max value. Multiple experiments showed that re-scaling coordinate data played significant role in the model convergence during training. The shape of spatial encoding tensor is $(N, P_{spatial})$, where N is the number of stations, and $P_{spatial}$ is the number of spatial positional encoding.

To encode this information to the input data, we combine the temporal and spatial positional encoding in one tensor. For each timestamp, the temporal encoding for all stations is the same, therefore we repeat the temporal encoding in the N dimension. Similarly, the spatial temporal for each station is the same for all timestamps, then we repeat the temporal encoding in the K dimension. Lastly, we concatenate the resulting tensors, and obtain the positional encoding tensor of shape (N, k, P), where P is the sum of the number of temporal and spatial encoding.

Notice that the information to build the positional encoding tensor is also known for the prediction target. Therefore, we build an output positional encoding tensor that functions as the query in the attention-mechanism (see section 4.3.2).

Input Hidden Representation

To incorporate the positional encoding information to the input data, we transform the input data to a hidden higher-level abstraction using a fully connected neural network. The number of hidden units is P, which is the same number of positional encoding variables. After the transformation, the shape of high-level abstraction representation of the input data is (N, k, P), where N is the number of stations, k is the look back window, and P is the number of positional encoding variables.

Graph Representation

To capture demand fluctuation during station closure, the graph represents physically close neighboring stations, where riders are likely to walk. The nodes of the graph are the BRT stations, and the links connecting these nodes represent stations that are within a walkable distance (less than 800 meters), and that at least share one service line that connects them. The resulting network representation for the BRT system is shown in Figure 4.2. Figure 4.2 also shows, for instance, that the network connects the stations in along the *Eje Ambiental*, but not with station in *Calle 10*. While these stations are at a walkable distance, there is no service that connects them. Therefore, it is very unlike that riders will walk to stations in *Calle 10* if any of the stations in is *Eje Ambiental* closed. The graph representation is encoded in an adjacency matrix of shape (N, N), where one represents a connection between two stations, and zero otherwise.

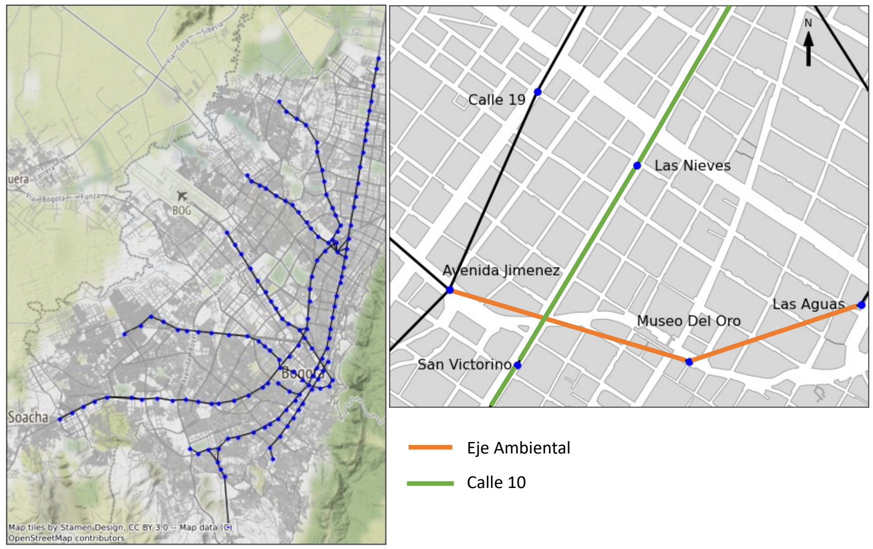

Figure 4.2: BRT Network in Bogotá A) Entire BRT netwrok B) Example Eje Ambiental and Calle 10

2D Graph Masked multi-head cross-Attention

The attention mechanism is a function with three main components: the query (q), the key (k), and the value vector(v), as shown in eq2:attention (also knows as dot-product attention). The output is a transformation of the value vector by the weights define be the query and the key. This weights represent the importance of the input to predict the output.

$$Attention(q,k,v) = softmax(\frac{qk^{\top}}{\sqrt{d_k}})v \tag{4.2}$$

For the purpose of this research, the query is the positional embedding of the target prediction, and the key and value are the input data and the positional encoding information. This type of attention is known as crossed-attention because the query and the key are two different tensors. Additionally, we can pack the vector information to any N-dimensional matrix. In our case, the key and value are 3-dimensional matrices of shape (N, k, P) as mentioned in 4.3.2, where N represents the number of stations (e.g: spatial dimension), and k represents the look back window (e.g.: temporal dimension). To better represent the complex spatiotemporal correlations of the transit demand, we apply the attention function to the first and second dimension simultaneously, therefore the the name "2D attention". Multi-head attention involves running multiple attention layers. The shape of the resulting attention tensor is (h, N, k, N), where h is the number of attention heads (in case of multihead attention), and the sum of the attention coefficient in each (k, N) slice is one.

The resulting attention matrix is fully defined for all stations and look back windows, representing a fully connected graph in time and space. As we are interested in capturing the effect of closures in nearby stations, we mask the attention mechanism to only neighbouring stations, using the adjacency matrix the graph representation.

Other components

As suggested in [112] we maintain the residual connection and the layer normalization, which is a common practice to avoid exploding gradients as more attention layers are added. We also maintain the feed forward layer. The last layer has a linear activation function to predict the demand for all transit stations at once.

4.3.3 Station Closure Modeling Strategies

One of the objectives of this research is to identify some modeling strategies that can read the status of the station to zero out the demand for a given station and time, and use the information to improve ridership prediction in neighbouring stations.

There are multiple ways to process and input station closure information. For this research , we select and test two modeling strategies (1) station state as a dummy variable, and (2) as a masked function layer in the final prediction.

An straightforward signal to include station closures can be adding the information as a dummy variable and treat it as an explanatory variable. To test this model strategy, in the proposed framework, we include this dummy variable in the positional encoding information, as shown in Figure 4.1. Notice that the the station status information varies both in the spatial and temporal dimension (e.g. station A and B are closed in the fifth and sixth time-step). Therefore, the tensor that encodes this information has a shape of $(N, k, 1)$ and can be concatenated with the positional encoding tensor. When we test this model strategy for other model architectures (see 4.3.5), we include station closure dummy variable in the covariates matrix.

A masked function is typically used to prevent overfitting or to disregard non-relevant information while estimating errors. For instance, in [96], some hidden neurons are randomly reduced to zero to avoid overfitting. In language processing, padding is usually masked because it does not add any relevant information to the prediction task. In this research, we test a mask function to explicitly zero out the predicted demand when the stations are closed, also shown in Figure 4.1 with a dotted line. While the mask function can be applied before or after estimating the loss function during training, we apply it before the loss function estimation. Therefore, we expect that the mask serves as a regularization technique to generalize in the case of station closures.

4.3.4 Data Preparation

Data preparation techniques are rarely mention, or briefly discuss. Data Most relevant literature uses the min-max transformation to scale transaction data form zero to one (cite at least 4 papers with more citations).

Data preparation is often overlooked and briefly mentioned in discussions, with papers tending to rely on min-max normalization without sufficient justification [7, 63, 44]. However, the impact of other transformation techniques have not been studied. In this research, we test two data processing techniques: (1) The min-max normalization, and (2) the logarithmic transformation. These two transformations because they maintain the meaning of zero, which is relevant to model station closures, and compress large ranges of data. The min-max normalization scales the data between 0 and 1 while preserving the shape of the original distribution, while the logarithmic transformation convert non-normally distributed data into a more normal distribution, which can improve the accuracy and performance of statistical models.

Figure 4.3 shows the 15-mins transaction distributions for the original data, the min-max normalization, and the logarithmic transformation. The original data distribution shows that the majority of 15-minute intervals have less than 2500 transactions, and the maximum transaction in the 15-mins period is 8138. The min-max normalization scales the data in the range of 0 to 1, maintaining the shape of the original distribution. The logarithmic transformation transforms the data to a more normal distribution and seems to create a gap between zero transactions and non-zero transactions. The range of this transformation is from zero to nine.

4.3.5 Experiments

In order to test the ability of the proposed model framework to improve forecasting prediction in the neighbouring stations when a given station is close, we conducted a total of 30 experiments by testing 5 different modeling architectures, each with 3 closure modes and 2 data transformation techniques, as presented in Table 4.1. The model architectures selected are deep learning models available in an open-source benchmark codebase. We evaluated the effectiveness of the graph information by testing the 2D Transformer (2DT) without graph information, which represents a fully connected graph, and with the graph representation (G-2DT), as described in section 4.3.2.

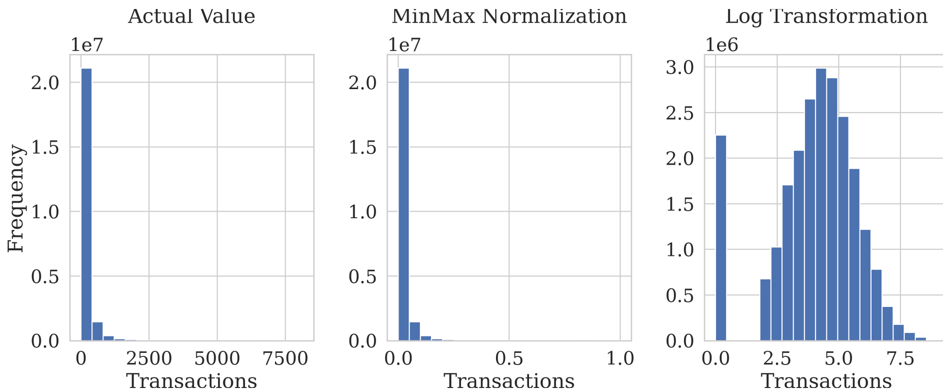

Figure 4.3: 15-mins boarding distribution - Original Values, Min-max normalization and Logarithmic transformation

Table 4.1: Experiment Setting.

Model	Closure Mode	Data Transformation
MLP	None	Min-max
CNN	Masked Function	Logarithmic
LSTM	Dummy Variable	
2DT		
G-2DT		

4.3.6 Metrics of Analysis

We used three metrics of analysis to compare our results. These metrics were chosen for their ability to compare performances between models and stations and handle circumstances where the target value is zero.

Since the modeling framework predicts the demand for each transit station, it is important to compare the performance across stations. However, metrics such as the Root Mean Squared Error (RMSE) depend on the magnitude of the data, therefore not adequate for this analysis. Additionally, since we are comparing the performance in the case when stations are close, many target values are likely to be zero. Metrics such as the Mean Absolute Square Error would be undefined for this conditions. Therefore, we select the Weighted Mean Absolute Percentage error (wMAPE), the symmetric Mean Absolute Percentage Error (sMAPE) and the Mean Arctangent Absolute Error (MAAPE), as follows:

$$wMAPE_t = 100\% * \frac{\sum_i^n |\hat{y}_i \ - y_i|}{\sum_i^n |y_i|} \tag{4.3}$$

$$sMAPE_t = 100\% * \frac{1}{n}\sum_i^n (\frac{|\hat{y}_i \ - y_i|}{|\hat{y}_i| + |y_i|}) \tag{4.4}$$

$$MAAPE_t = 100\% * \frac{1}{n}\sum_i^n \arctan(|\frac{\hat{y}_i \ - y_i}{y_i}|) \tag{4.5}$$

Where y_i is the observed value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

For all experiments, we compared the prediction error in the j neighbouring stations of station i ($\mathcal{N}_i$) when station i is open and stations j are open for all stations i in the set of stations N (Equation 4.6), and when station i is closed, stations j are open for all stations i in the set of stations N (Equation 4.7). For simplicity, we refer to these scenarios as open scenario, and close scenario, respectively.

$$OpenScenarioError = error_{j \in \mathcal{N}_i, i=1, j=1} \quad \forall i \in N \tag{4.6}$$

$$CloseScenarioError = error_{j \in \mathcal{N}_i, i=0, j=1} \quad \forall i \in N \tag{4.7}$$

To ensure accurate comparisons of errors, we re-scaled the prediction and evaluated the error using the original, untransformed target data.

4.4 Results

4.4.1 Data and Station Closures

We used boarding transactions data transactions every 15 minutes of 147 stations of the BRT system in Bogotá, between August 2015 and May 2021. The BRT system is the city's main mass public transportation system, with over 2.5 daily transactions during the study time frame. We train the models with data from August 2015 to August 2018, and test the performance for the remaining period. New stations that start operating after the training date are not considered in the study (e.g. Transmicable stations started operations in Dec 2018). The data set is publicly available at: https://datosabiertos-transmilenio.hub.arcgis.com.

For the purpose of this research, we defined a station closure as a period of at least two consecutive hours with zero transactions in one BRT station. The BRT system operates from 4:30 am to 11:00 p.m., however operation hours vary by stations. To avoid incorrectly labeling a station as 'closed' outside of operating hours, we limited our closure analysis to the time period between 5 a.m. and 10:00 p.m., during which all stations are typically operating. In total, our analysis includes more than 10,000 station closures lasting longer than two hours. Figure 4.4, shows the distribution of the duration of station closures for the five years of transactions.

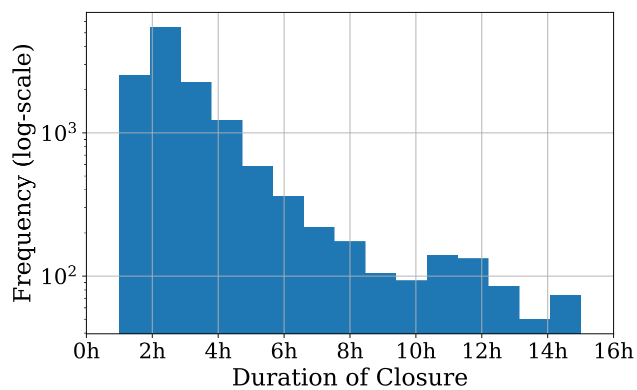

Figure 4.4: Distribution of station closure duration

Based on our definition of station closure, it is not possible to determine the nature of the closure. Therefore, station closures includes both expected and unexpected. However, both types of closures affect the demand in neighboring stations, as shown in the examples in Figure 4.5. The first plot displays the impact of a planned closure for maintenance of Museo del Oro station in September 2016. During this period, the demand for the station appears to have been redistributed to its neighboring stations, Avenida Jimenez and Las Aguas. In the second plot, an unexpected one-week closure of Calle 19 station on November 24, 2019 resulted in a redistribution of demand to Calle 22 station.

4.4.2 Effect of Station Closures on Prediction Accuracy

One of this research is to demonstrate the impact of station closures on the prediction accuracy of neighboring stations and to evaluate the effectiveness of different modeling strategies in improving prediction accuracy under such circumstances. The findings demonstrate that the close scenario tends to yield higher prediction errors than the open scenario, as evidenced by the increasing slope of the lines from open to close in Figure 4.6. However, the extent

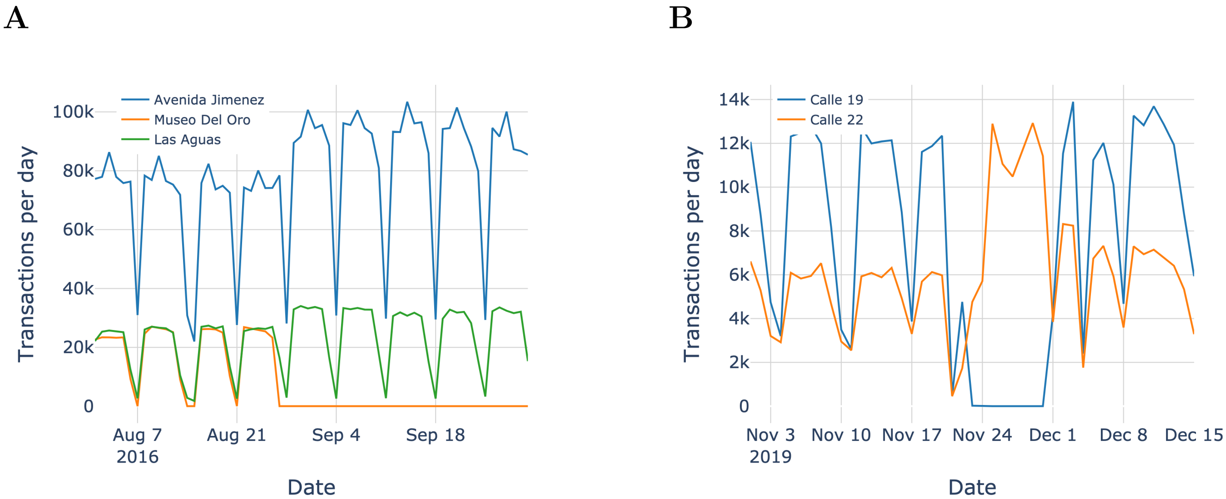

Figure 4.5: Examples of station closures. A) Station Museo del Oro was closed for planned maintenance in September 2016. Neighboring stations Avenida Jimenez and Las Aguas appeared to capture the lost demand. B) Station Calle 19 unexpectedly closed for one week, and demand was redistributed to the neighboring station Calle 22

to which this increase happens significantly varies when considering models, closure modes, and data transformation techniques.

Failure to account for station closures results in a twofold increase in the prediction error between open and closed scenarios. The best-performing models achieved a wMAPE ranging from 14.63% to 15.09% for the open scenario. However, for the closed scenario, the prediction error of surrounding stations increases to at least 31.12%, more than double the error observed in the open scenario.

Likewise, the sMAPE demonstrates a stark increase from 11.29% for the best-performing model in the open scenario to 25.86% in the closed scenario, representing a difference of 14.57%. In addition, for the best-performing model, the MAAPE increases from 21.12% in the open scenario to 36.64% in the closed scenario, revealing a difference of 15.52 points and an increase factor of 1.73.

Moreover, our findings suggest that failing to consider station closures results in no clear trend regarding the data transformation technique that performs best, except for the MAAPE, where logarithmic transformation models outperform min-max normalization models.

The results indicate that the logarithmic transformation outperforms the min-max normalization for almost all models and metrics when station closure information is incorpo-

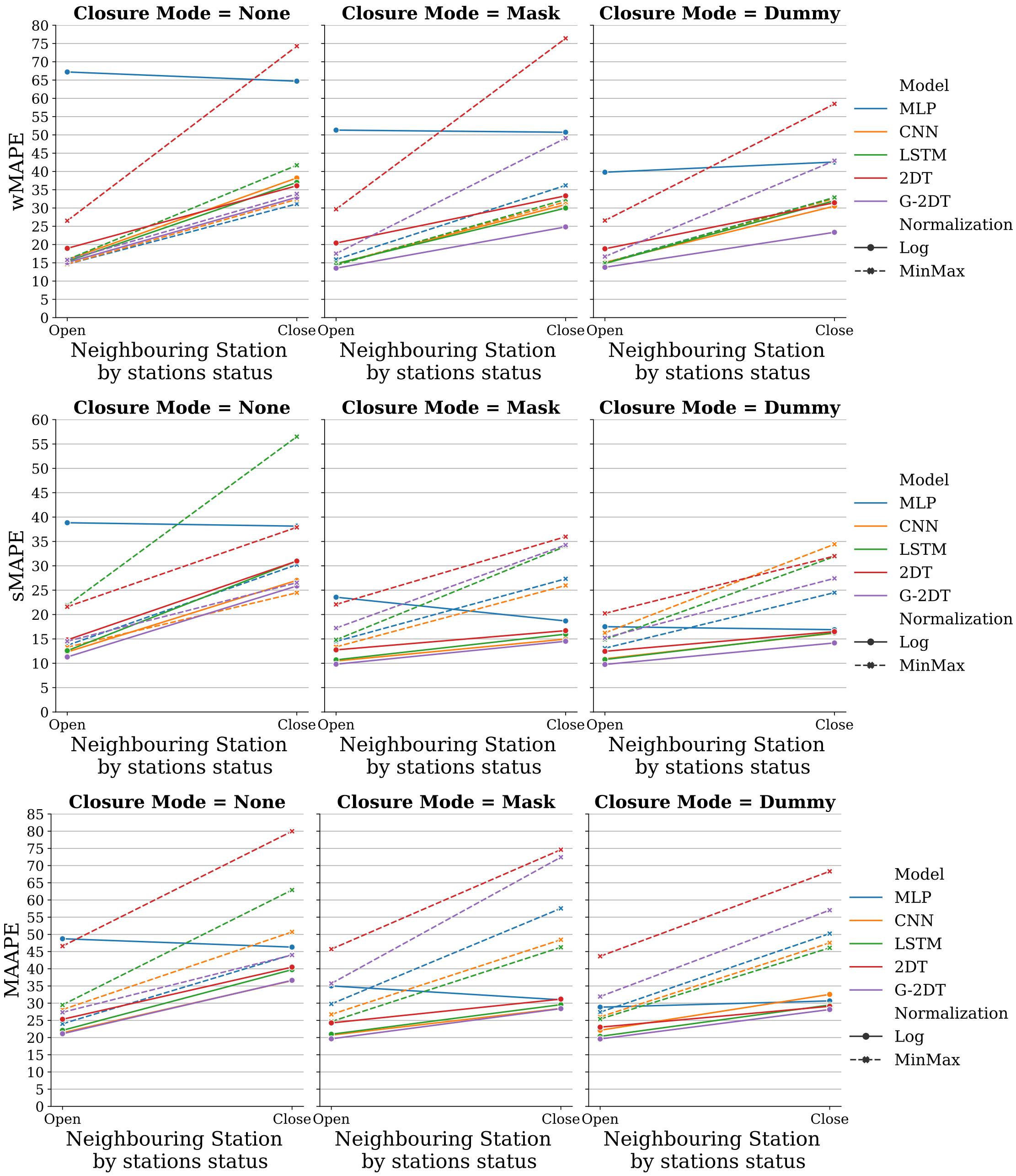

Figure 4.6: Neighbouring stations prediction error by station status (open vs close), by closure mode and data transformations techniques.

	Closure Mode	None		Mask		Dummy	
		Open	Close	Open	Close	Open	Close
Scaling	Model						
MinMax	MLP	**14.63 (±0.42)**	**31.12 (±1.50)**	15.91 (±0.46)	36.18 (±1.72)	14.97 (±0.44)	32.87 (±1.53)
	CNN	**14.66 (±0.43)**	**32.38 (±1.58)**	14.64 (±0.43)	31.69 (±1.53)	14.87 (±0.43)	32.60 (±1.53)
	LSTM	15.87 (±0.47)	41.69 (±2.00)	14.32 (±0.40)	32.35 (±1.60)	14.89 (±0.41)	32.88 (±1.57)
	2DT	26.51 (±0.90)	74.29 (±4.15)	29.67 (±1.03)	76.43 (±4.44)	26.57 (±0.89)	58.49 (±3.12)
	G-2DT	15.79 (±0.52)	33.84 (±1.72)	17.52 (±0.57)	49.12 (±2.24)	16.71 (±0.54)	42.93 (±2.41)
Log	MLP	67.22 (±0.47)	64.70 (±1.07)	51.31 (±0.59)	50.72 (±1.34)	39.79 (±0.66)	42.57 (±1.40)
	CNN	15.85 (±0.49)	38.21 (±1.94)	14.66 (±0.45)	30.95 (±1.60)	15.11 (±0.47)	30.51 (±1.50)
	LSTM	15.40 (±0.47)	37.00 (±1.68)	14.77 (±0.49)	29.98 (±1.46)	14.81 (±0.44)	32.00 (±1.52)
	2DT	18.96 (±0.58)	36.08 (±1.64)	20.41 (±0.70)	33.33 (±1.74)	18.84 (±0.59)	31.44 (±1.49)
	G-2DT	**15.09 (±0.53)**	**32.91 (±2.11)**	**13.53 (±0.46)**	**24.82 (±1.53)**	**13.77 (±0.43)**	**23.35 (±1.40)**
% Impr. w.r.t 2nd best		0.2%	3.89%	5.52%	17.62%	7.02%	23.47%
BM % Impr. w.r.t None		-	-	10.34%	24.58%	8.75%	29.04%
$\Delta_{close,open}$ BM		16.49		11.29		9.58	

Note: Bold indicates the best performing result. If two or more models in the same column are bolded, it indicates that these models are statistically similar. BM = Best Model. Impr. = Improvement. w.r.t = with respect to

Table 4.2: wMAPE for open and close scenarios benchmark by data transformation, model, and closure mode.

	Closure Mode	None		Mask		Dummy	
		Open	Close	Open	Close	Open	Close
Scaling	Model						
MinMax	MLP	13.68 (±0.31)	30.23 (±0.58)	14.48 (±0.28)	27.33 (±0.37)	13.10 (±0.25)	24.50 (±0.35)
	CNN	13.30 (±0.26)	**24.47 (±0.35)**	13.37 (±0.29)	25.97 (±0.44)	16.23 (±0.39)	34.43 (±0.61)
	LSTM	21.90 (±0.53)	56.50 (±0.73)	14.81 (±0.36)	34.01 (±0.63)	14.88 (±0.33)	31.89 (±0.59)
	2DT	21.60 (±0.37)	37.90 (±0.44)	22.08 (±0.38)	35.97 (±0.43)	20.24 (±0.36)	31.99 (±0.39)
	G-2DT	14.52 (±0.28)	26.53 (±0.42)	17.23 (±0.34)	34.27 (±0.41)	15.23 (±0.29)	27.42 (±0.43)
Log	MLP	38.84 (±0.31)	38.14 (±0.44)	23.56 (±0.25)	18.68 (±0.28)	17.51 (±0.22)	16.89 (±0.25)
	CNN	12.32 (±0.24)	27.01 (±0.50)	10.47 (±0.19)	14.96 (±0.24)	10.93 (±0.19)	16.17 (±0.25)
	LSTM	12.58 (±0.28)	31.00 (±0.59)	10.70 (±0.19)	15.97 (±0.26)	10.77 (±0.19)	16.26 (±0.26)
	2DT	14.79 (±0.31)	30.97 (±0.51)	12.75 (±0.22)	16.69 (±0.26)	12.45 (±0.21)	16.49 (±0.27)
	G-2DT	**11.29 (±0.24)**	25.86 (±0.47)	**9.81 (±0.19)**	**14.51 (±0.24)**	**9.76 (±0.18)**	**14.17 (±0.22)**
% Impr. w.r.t 2nd best		8.32%	5.37%	6.30%	3.00%	9.37%	12.37%
Best % Impr. w.r.t None		-	-	13.11%	43.89%	13.55%	45.20%
$\Delta_{close,open}$ Best Model		14.57		4.70		4.41	

Note: Bold indicates the best performing result. If two or more models in the same column are bolded, it indicates that these models are statistically similar. BM = Best Model. Impr. = Improvement. w.r.t = with respect to

Table 4.3: sMAPE for open and close scenarios benchmark by data transformation, model, and closure mode

	Closure Mode	None		Mask		Dummy	
Scaling	Model	Open	Close	Open	Close	Open	Close
MinMax	MLP	24.01 (±0.44)	44.11 (±0.65)	29.78 (±0.57)	57.56 (±0.83)	27.38 (±0.54)	50.23 (±0.71)
	CNN	28.13 (±0.55)	50.71 (±0.72)	26.74 (±0.51)	48.44 (±0.71)	26.01 (±0.44)	47.54 (±0.63)
	LSTM	29.50 (±0.53)	62.90 (±0.72)	24.66 (±0.48)	46.25 (±0.66)	25.39 (±0.46)	46.08 (±0.63)
	2DT	46.59 (±0.79)	79.98 (±0.84)	45.71 (±0.76)	74.62 (±0.86)	43.64 (±0.79)	68.33 (±0.83)
	G-2DT	27.35 (±0.53)	44.00 (±0.62)	35.72 (±0.70)	72.44 (±0.85)	31.93 (±0.61)	57.04 (±0.78)
Log	MLP	48.73 (±0.29)	46.30 (±0.37)	35.02 (±0.30)	30.98 (±0.40)	28.84 (±0.34)	30.66 (±0.44)
	CNN	**21.40 (±0.35)**	**36.55 (±0.47)**	20.72 (±0.37)	**28.47 (±0.46)**	22.13 (±0.41)	32.55 (±0.50)
	LSTM	22.10 (±0.40)	39.74 (±0.52)	20.97 (±0.39)	29.57 (±0.44)	20.35 (±0.36)	29.47 (±0.44)
	2DT	25.31 (±0.42)	40.48 (±0.49)	24.26 (±0.40)	31.18 (±0.45)	23.02 (±0.38)	29.10 (±0.42)
	G-2DT	**21.12 (±0.38)**	**36.64 (±0.50)**	**19.64 (±0.36)**	**28.41 (±0.48)**	**19.60 (±0.37)**	**28.15 (±0.46)**
% Impr. w.r.t 2nd best		1.31%	0.03%	5.21%	0.21%	3.69%	3.26%
Best % Impr. w.r.t None		-	-	7.01%	22.46%	7.20%	23.17%
$\Delta_{close,open}$ Best Model		15.52		8.77		8.55	

Note: Bold indicates the best performing result. If two or more models in the same column are bolded, it indicates that these models are statistically similar. BM = Best Model. Impr. = Improvement. w.r.t = with respect to

Table 4.4: MAAPE for open and close scenarios benchmark by data transformation, model, and closure mode

rated, as shown in Figure 4.6. For instance, wMAPE for the open scenarios in the G-2DT model with logarithmic transformation is approximately 13.5%, compared to around 17% for the min-max transformation. In the close scenario, the wMAPE also increases from roughly 24% in the log transformation to approximately 45% in the min-max normalization. The sMAPE for the same model increases from about 9.7% to 16% in the close scenario and from 14% to 30% in the close scenario. Similarly, the MAAPE increases from 19.6% in the log transformation to 32% in the min-max transformation for the close scenario and from 28% to roughly 65% in the open scenario.

Regarding the CNN and LSTM models, the wMAPE performs relatively similarly for the open scenario when comparing the data transformation techniques. However, there is a notable reduction in prediction error for the closed scenario. Overall, the logarithmic transformation seems to reduce the prediction error when considering station closure for most models and most metrics of analysis.

When incorporating station closure information, the log version of our proposed model is the best-performing model, regardless of the strategy used to process station closure information. For the wMAPE, our model with the masked function outperforms the best second model by 5.52% and 17.62% for the open and close scenarios, respectively. Similarly, our model shows 7.02% and 23.47% improvements for the dummy variable strategy compared to the second-best performing model for the open and close scenarios.

For other metrics, our model also shows improvements ranging from 3.69% to 9.37% for the open scenario and from 0.21% to 12.37% for the close scenario. These findings highlight the robustness and effectiveness of our proposed model, which consistently outperforms other models across a range of metrics and scenarios.

Considering station closures also decreases prediction error for both open and closed scenarios, but the reduction is greater for the closed scenario. The masked function and dummy variable strategies decrease the wMAPE by 10.34% and 8.75% for the open scenario, respectively, whereas for the close scenario, the reduction is 24.58% and 29.04%, respectively. Similarly, the sMAPE error in the open scenario decreases by 13.11% and 13.55% for the masked function and dummy variable strategies, respectively. Still, the improvement in the close scenario is more significant at 43.89% and 45.20%, respectively. Finally, the MAAPE for the open scenario shows an improvement of approximately 7%, while the close scenario improves by 23%. Overall, including station closure information significantly reduces prediction error for both open and close scenarios, with greater improvements observed in the close scenario.

In addition, for the best-performing model, the difference in wMAPE between the close and open scenarios was initially 16.49 when station closures were not considered. However, the masked function and the dummy variable strategy reduced the difference to 11.29 and

9.58, respectively, indicating an average reduction of 36%. Similarly, for sMAPE, the difference decreased from 14.57 to 4.70 and 4.41 for the masked function and dummy variable strategies, respectively. For the MAAPE, the difference decreased from 15.52 to 8.77 and 8.55, indicating an approximate reduction of 44%. These results suggest that incorporating station closure information can effectively reduce the prediction error, particularly for the close scenario, and that the masked function and the dummy variable strategy are effective approaches to consider station closures in the proposed model.

4.5 Conclusion and Discussion

In this study, we reduced prediction errors in neighboring stations when a station closure occurs by incorporating station closure information. We present a novel modeling framework that utilizes the graph structure information and a two-dimensional multi-head attention mechanism to capture spatiotemporal correlations simultaneously, improving overall performance compared to other state-of-the-art models. Our approach has also been implemented on an open-source benchmark platform, facilitating reproducibility and enabling future researchers to compare, refine, and extend our model.

Furthermore, the results suggest that the logarithmic transformation yields better results compared to min-max normalization when station closure information is added, despite the latter being commonly used in deep learning implementations of short-term ridership prediction models. In Figure 4.7, we present an example of the closure of station Ciudad Universitaria with a duration of 3.5 hours, and plot the predicted demand by closure mode, and data transformation techniques. When incorporating station closure information as a dummy variable, the min-max normalization can lead to non-zero predictions, which does not reflect the demand patterns during a closure. On the other hand, the log-transformation approach can effectively reduce the demand to zero, thereby reducing noise and improving model performance. As observed in Figure 4.3, the log normalization method produces a distinct gap in the histogram for zero transactions, which may aid in the detection of station closures. In contrast, the min-max normalization method results in most values being relatively small and closer to zero, potentially making it more difficult for the model to accurately distinguish a station closure. The results suggest that a log-transform normalization method may be preferred over the commonly used min-max normalization method in the literature.

Regarding the strategy to include information on station closures, our results show that the dummy variable and masked function yield similar results. The masked function serves as a form of regularization by reducing noise and irrelevant data. This regularization technique may be preferred over the use of station closure information as a dummy variable, as the masked function does not require prior knowledge of station closures for prediction. As a result, the use of the masked function may lead to more robust and generalizable models for short-term transit demand prediction.

The attention mechanism has the potential to provide valuable information through its ability to estimate attention scores. Using the same station closure example shown in Figure 4.7, Figure 4.8 illustrates a sample of the attention scores of a masked function strategy estimated for a 15-minute period at station Av. El Dorado, which is a neighboring station of Ciudad Universitaria. The temporal dimension is represented in rows, while the three neighboring stations (Ciudad Universitaria, Universidad Nacional, and CAD), as well as the station itself, are represented in columns. During this period, the attention scores for the nearby closed station (Ciudad Universitaria) exhibited significant reductions across all attention heads, providing further evidence that the masked function helps to reduce noise and eliminate irrelevant data.

In Figure 4.9, we present the evolution of attention scores for each closure mode and time period. We observe that for the masking strategy, the same day attention score of the close station decreases to almost zero, indicating that the model correctly disregards its influence during the closure period. Intriguingly, the attention score for station Universidad Nacional increases for the same day, as well as the daily and weekly periods during the closure, indicating the ability of the model to capture long-range temporal and spatial interdependence. These findings provides additional evidence of the concurrent spatiotemporal correlations that exist in transit demand and underscores the significance of incorporating such correlations to enhance model performance during station closures. Nevertheless, further investigation is required to fully understand the interpretability and practical utility of the attention scores.

By addressing the challenge of predicting neighboring station performance during a station closure, our work contributes to the advancement of transportation system resilience and user satisfaction. [82] suggest that improving the ability of transit agencies to make rapid decisions during disruptions is an effective strategy for enhancing the resilience of public transit systems. Our short-term prediction approach facilitates transit agencies to anticipate the impact of station closures in advance and take appropriate actions to mitigate the impact, thereby improving their decision-making abilities during temporary station closures. Transit disruptions can also significantly impact user satisfaction [21, 85]. Mitigating the impact of these disruptions can help improve overall user confidence in the system, which in turn may help reduce the negative effects of disruptions on user satisfaction.

Despite evidence of a correlation between the built environment and transit ridership in various regions [132, 57, 97, 113], current short-term ridership models often do not account for this information. Our research only incorporated latitude and longitude data, but our modeling framework can easily incorporate additional built environmental variables to improve model performance.

In summary, our research demonstrates the relevance of including station closure information in short-term ridership prediction models, as it can effectively reduce the prediction error of neighboring stations impacted by a closure. By incorporating this information, transit agencies can better anticipate and mitigate the effects of closures, and improve their overall response to disruptions.

The open-source codebase to replicate the results of this research is available at:

`https://github.com/jdcaicedo251/transit_demand_prediction.`

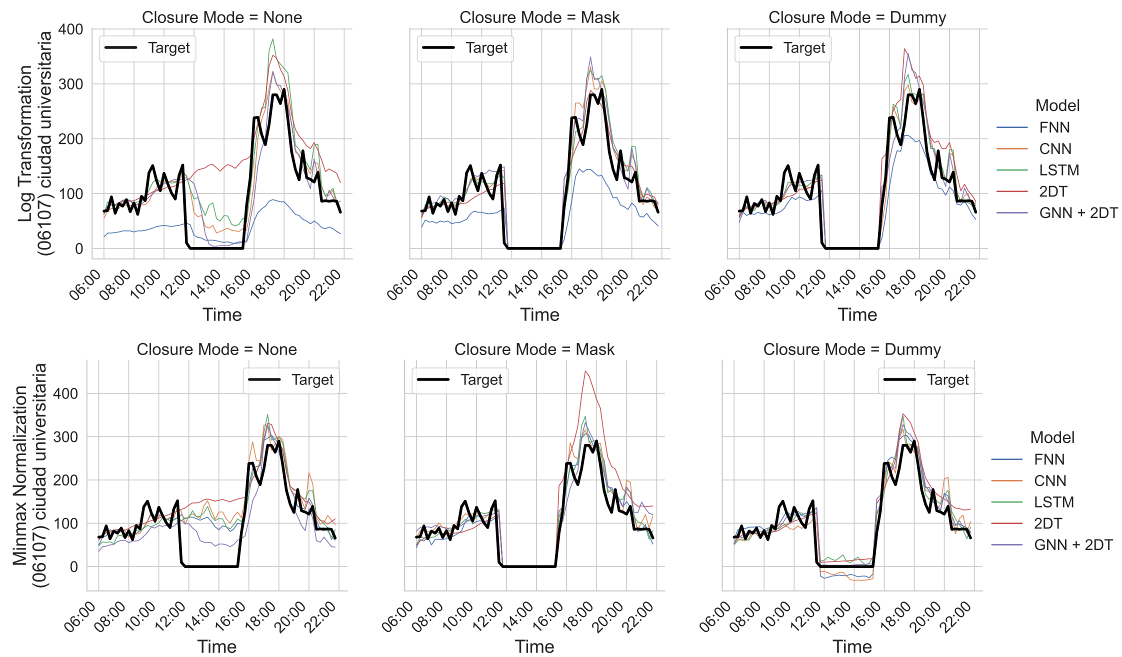

Figure 4.7: Example prediction by closure mode and data processing techniques during a 3.5h closure for Station Ciudad Universitaria. Closure on August 15, 2018 at 11:45 a.m.

Figure 4.8: Example of Spatiotemporal Attention scores (in %) for station Av. El Dorado during unexpected closure of Station Ciudad Universitaria. Timestamp: 2018-08-15 13:00

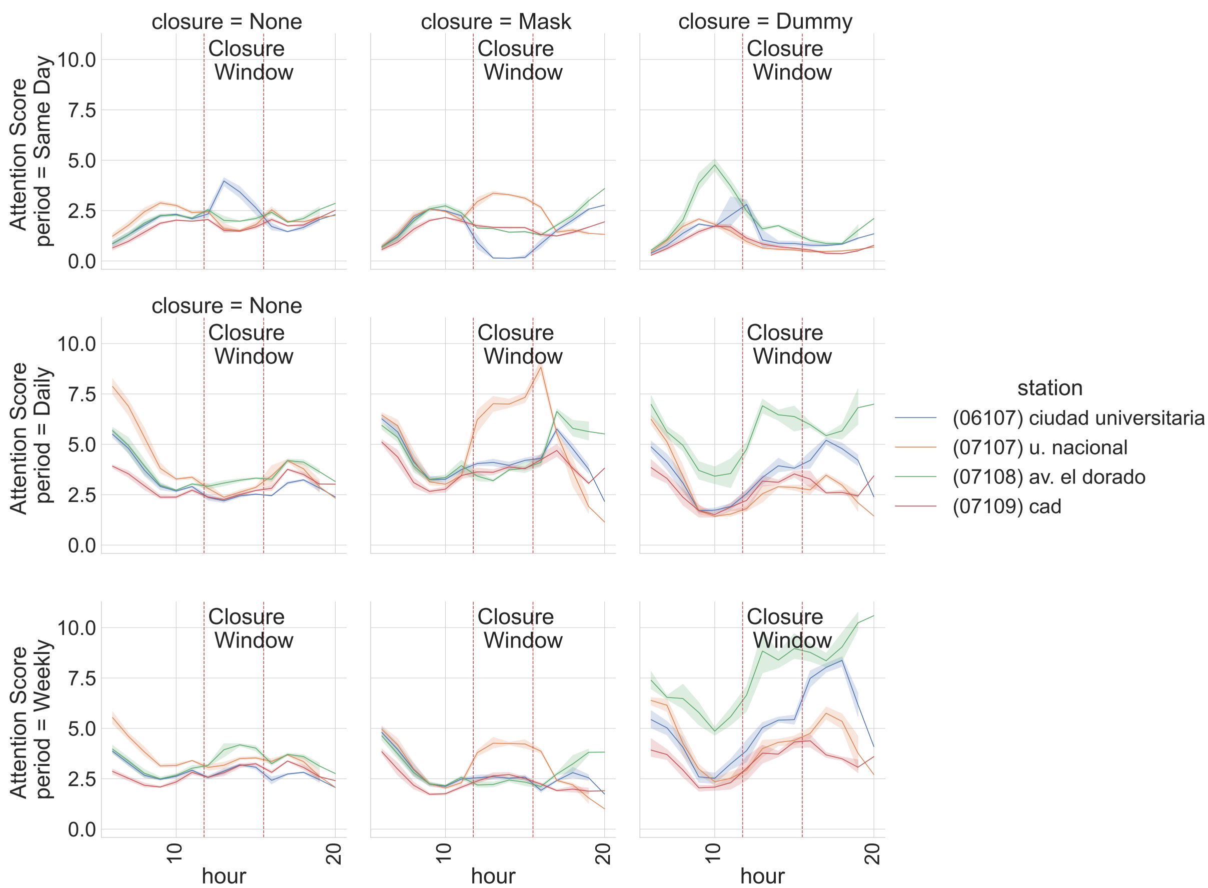

Figure 4.9: Evolution of attention scores (in %) for station Av. El Dorado during unexpected closure of Station Ciudad Universitaria.

Chapter 5

Conclusions

"I compare open source to science. Science took this whole notion of developing ideas in the open and improving on other peoples' ideas. It made science what it is today and made the incredible advances that we have had possible."

– Linus Torvalds.

5.1 Summary

We motivated this dissertation by stating the critical role of public transit in providing access to opportunities, generating economic benefits, promoting sustainable transportation, and improving public health, particularly for low-income populations. However, public transit systems are vulnerable to various anticipated and unforeseen disruptions, including maintenance operations, construction activities, accidents, natural disasters, technical malfunctions, security threats, protests, and public health emergencies, which can drastically change the travel needs of the population, and jeopardize the benefits of public transit. To mitigate and reduce the impact of the disruption, transit agencies need to make informed decisions fast, including vehicle scheduling, fleet size per type of vehicle, driver schedules, and driver roster.

In recent years, transit agencies have increasingly adopted smart cards to collect fares, making valuable data readily available for analysis, and can potentially be used to inform decisions during disruptions. In response, the research community has proposed many inference methods and predictive models to unravel transit patterns and behaviors, information that supports the decisions mentioned above. Despite the abundance of literature, we identified that the research community has yet to generate practical knowledge for decision-making, and the use of current methodologies needs to account for the possibility of disruptions.

In this dissertation, we used inference methods and prediction models to empirically show the potential of smartcard data to inform decision-making during disruptions. We perform a systematic and rigorous statistical benchmark of the most common methods to compare current model strategies. We made it available for other researchers to reduce research barriers, improve benchmarking practices and create collective wisdom. We used this benchmark to propose a new model framework that accounts for station closure and improves short-term ridership prediction under these circumstances.

Using smartcard data, and publicly available datasets such as census and land use data, we infer the strata of frequent transit users and track the evolution during the COVID-19 pandemic. We showed that the transaction reduction for all strata at the beginning of the pandemic was relatively similar, at around 85 to 90%, but lower strata were returning to transit at a pace five times faster than higher strata.

We compare common model architecture for short-term ridership prediction in stable conditions. The results show that most tested models perform relatively similarly in stable conditions, with forecasting errors varying from 8.5 to 12%. However, for highly dynamic conditions, it was shown that all models performed significantly worse. In the protest condition, the prediction error increases for all models in ranges from 14 to 24%., and in the COVID-19 pandemic, the prediction error increases from 12 to 82%, on average. Notably, in the COVID-19 pandemic scenario, the LSTM model stood out by outperforming other models and adapting faster. The prediction error stabilized within approximately 1.5 months, whereas other models continued to exhibit higher error rates even a year after the start of the pandemic.

By incorporating station closure information into our proposed model, we reduced prediction errors by 3.3 to 23.5% under temporary closures compared to existing models in the open-source codebase infrastructure. These improvements were observed across multiple metrics and modeling strategies.

5.2 Contributions

The contributions of this research are summarized as follows:

1. Develop a method to enrich smartcard data with socio-economic information for highly dynamic conditions.
2. Expand the understanding of the change in transit use by different socio-demographic groups during the COVID-19 pandemic in Bogotá.

3. Create an open-source codebase infrastructure to consolidate short-term ridership prediction research to perform systematic, reliable and statistically rigorous benchmarks tests to accelerate the advancements of the models.

4. Systematically compare the performance of state-of-the-art methods for short-term ridership prediction during both stable and highly dynamic conditions.

5. Demonstrate the importance of including station closures to improve the prediction accuracy for nearby stations.

6. Propose a novel modeling framework that captures spatiotemporal correlations of the transit network, improving overall performance and the prediction accuracy for stations impacted by other station closures.

7. Integrate the new modeling framework in the open-source codebase infrastructure for other researchers to replicate, reproduce, compare, improve, and build upon our model.

5.3 Recommendations for Future Research

In terms of inference methods, future research could build upon the limitations of the present study by exploring the impact of home and work relocation on the estimation of the transit usage rate. During the COVID-19 pandemic, changes in the home location were very likely given the stay-at-home orders and the option to telecommute, which might affect the estimate of transit usage rate. Moreover, the estimation of the probability vector could be enhanced by incorporating additional factors, such as the distance to the BRT station and the spatial correlation among stations. Furthermore, future studies may benefit from a more aggregated level of validation, such as at the station level, using other publicly available datasets.

Future research could conduct a more extensive analysis of the metrics used for performance evaluation to ensure that they are aligned with the objectives of the transit agency. While the metrics used in this study were chosen to enable comparisons across stations and potentially across regions, it is important to note that the primary objective may vary depending on the needs of the transit agency. Therefore, future studies could examine other performance indicators and evaluate their relevance to specific transit agency goals.

Future research could aim to better align short-term ridership prediction algorithms with the needs of bus scheduling methods. While current research has focused on predicting ridership up to four steps ahead and every 15 minutes, it is unclear how well these predictions align with the data requirements of bus scheduling methods. Therefore, future studies could investigate the data needs of bus scheduling methods and evaluate how ridership prediction algorithms can be tailored to meet those needs. This could help to improve the applicability of short-term ridership prediction models for practical transportation operations. In the same line of thought, there needs to be more empirical evidence that combines short-term ridership prediction, and bus scheduling methods, even in stable conditions. Further exploration combining these methods can help to identify potential challenges and opportunities for improving transportation system performance and enhancing overall resilience.

Future research could investigate the potential applications and benefits of the attention mechanism scores used in short-term ridership prediction. While this research has shown an example of the attention scores for one station closure and their evolution over time frames, it is unclear how these scores can be interpreted and applied in practice. Therefore, further research could focus on validating the usefulness of these scores and identifying potential areas for analysis and interpretation. This could provide a deeper understanding of the underlying factors that drive ridership patterns, especially under disruptions.

5.4 Conclusion

Overall, this dissertation highlights the immense potential of readily available smartcard data to support decision-making processes rapidly and reliably during transit disruptions, thereby enhancing the overall resilience of transportation systems. While the methods and models presented here are demonstrated empirically with the BRT system in Bogotá, they have broad applicability to other regions. Additionally, developing an open-source codebase for short-term ridership prediction models is a significant step toward standardizing and enhancing the efficiency of future research. We believe that this collaborative effort will foster increased transparency and contribute to collective wisdom that will drive needed improvements in our public transit systems. I am excited to see what innovations lie ahead!

Bibliography

[1] Tanveer Ahmad and Huanxin Chen. "A review on machine learning forecasting growth trends and their real-time applications in different energy systems". In: *Sustainable Cities and Society* 54 (2020), p. 102010. ISSN: 2210-6707. DOI: https://doi.org/10.1016/j.scs.2019.102010. URL: https://www.sciencedirect.com/science/article/pii/S2210670719335516.

[2] Erik Almlöf et al. "Who Is Still Travelling by Public Transport during COVID-19? Socioeconomic Factors Explaining Travel Behaviour in Stockholm Based on Smart Card Data". In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3689091.

[3] Titan Alon et al. *The Impact of COVID-19 on Gender Equality*. Working Paper 26947. National Bureau of Economic Research, Apr. 2020. DOI: 10.3386/w26947. URL: http://www.nber.org/papers/w26947.

[4] Nilufer Sari Aslam, Tao Cheng, and James Cheshire. "A high-precision heuristic model to detect home and work locations from smart card data". In: *Geo-spatial Information Science* 22.1 (2019), pp. 1–11. DOI: 10.1080/10095020.2018.1545884. eprint: https://doi.org/10.1080/10095020.2018.1545884. URL: https://doi.org/10.1080/10095020.2018.1545884.

[5] Sebastian Astroza et al. "Mobility Changes, Teleworking, and Remote Communication during the COVID-19 Pandemic in Chile". In: *Findings* (2020). DOI: 10.32866/001c.13489.

[6] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. 2015.

[7] Yun Bai et al. "A multi-pattern deep fusion model for short-term bus passenger flow forecasting". In: *Applied Soft Computing* 58 (2017), pp. 669–680. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2017.05.011. URL: http://www.sciencedirect.com/science/article/pii/S1568494617302648.

[8] Juan Pablo Bocarejo S. and Daniel Ricardo Oviedo H. "Transport accessibility and social inequities: a tool for identification of mobility needs and evaluation of transport investments". In: *Journal of Transport Geography* 24 (2012). Special Section on

Theoretical Perspectives on Climate Change Mitigation in Transport, pp. 142–154. ISSN: 0966-6923. DOI: https://doi.org/10.1016/j.jtrangeo.2011.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0966692311002286.

[9] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3 (1986). ISSN: 03044076. DOI: 10.1016/0304-4076(86)90063-1.

[10] Anastasia Borovykh, Sander M Bohté, and Cornelis W Oosterlee. "Conditional Time Series Forecasting with Convolutional Neural Networks". In: *arXiv: Machine Learning* (2017).

[11] Rebecca Brough, Matthew Freedman, and David Phillips. "Understanding Socioeconomic Disparities in Travel Behavior during the COVID-19 Pandemic". In: *SSRN Electronic Journal* (2020). ISSN: 1556-5068. DOI: 10.2139/ssrn.3624920. URL: https://dx.doi.org/10.2139/ssrn.3624920.

[12] Barbara B. Brown and Carol M. Werner. "Before and After a New Light Rail Stop: Resident Attitudes, Travel Behavior, and Obesity". In: *Journal of the American Planning Association* 75.1 (2008). ISSN: 0194-4363. DOI: 10.1080/01944360802458013.

[13] Juan D. Caicedo, Joan L. Walker, and Marta C. González. "Influence of Socioeconomic Factors on Transit Demand During the COVID-19 Pandemic: A Case Study of Bogotá's BRT System". In: *Frontiers in Built Environment* 7 (2021). ISSN: 22973362. DOI: 10.3389/fbuil.2021.642344.

[14] R. Camporeale et al. "Quantifying the impacts of horizontal and vertical equity in transit route planning". In: *Transportation Planning and Technology* 40.1 (2017), pp. 28–44. DOI: 10.1080/03081060.2016.1238569. eprint: https://doi.org/10.1080/03081060.2016.1238569. URL: https://doi.org/10.1080/03081060.2016.1238569.

[15] Victor Cantillo-Garcia, Luis Guzman, and Julián Arellana. "Socioeconomic strata as proxy variable for household income in transportation research. Evaluation for Bogotá, Medellín, Cali and Barranquilla". In: *Dyna (Medellin, Colombia)* 86 (Dec. 2019), pp. 258–267. DOI: 10.15446/dyna.v86n211.81821.

[16] Zhichao Cao and Avishai (Avi) Ceder. "Autonomous shuttle bus service timetabling and vehicle scheduling using skip-stop tactic". In: *Transportation Research Part C: Emerging Technologies* 102 (2019). ISSN: 0968090X. DOI: 10.1016/j.trc.2019.03.018.

[17] Robert Cervero. "Linking urban transport and land use in developing countries". In: *Journal of Transport and Land Use* 6.1 (2013). ISSN: 19387849. DOI: 10.5198/jtlu.v6i1.425.

[18] Artem Chakirov and Alexander Erath. "Activity identification and primary location modelling based on smart card payment data for public transport". In: 2012.

[19] Qian Chen, Wenquan Li, and Jinhuan Zhao. “The use of LS-SVM for short-term passenger flow prediction”. In: *Transport* 26.1 (Apr. 2011), pp. 5–10. ISSN: 1648-4142. DOI: 10.3846/16484142.2011.555472. URL: https://doi.org/10.3846/16484142.2011.555472.

[20] Kyunghyun Cho et al. “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches”. In: *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111. DOI: 10.3115/v1/W14-4012. URL: https://aclanthology.org/W14-4012.

[21] Graham Currie and Carlyn Muir. “Understanding Passenger Perceptions and Behaviors During Unplanned Rail Disruptions”. In: *Transportation Research Procedia* 25 (2017), pp. 4392–4402. ISSN: 2352-1465. DOI: https://doi.org/10.1016/j.trpro.2017.05.322.

[22] Xiaoqing Dai, Lijun Sun, and Yanyan Xu. “Short-Term Origin-Destination Based Metro Flow Prediction with Probabilistic Model Selection Approach”. In: *Journal of Advanced Transportation* 2018 (2018). ISSN: 20423195. DOI: 10.1155/2018/5942763.

[23] DANE. *Análisis Geoespacial del CNPV 2018*. 2018. URL: https://geoportal.dane.gov.co/geovisores/territorio/analisis-cnpv-2018/ (visited on 06/09/2020).

[24] DANE. *Metodologia de estratificacion socioeconomica urbana para servicios publicos domiciliarios - Enfoque Conceptual*. 2015. URL: https://www.dane.gov.co/files/geoestadistica/estratificacion/EnfoqueConceptual.pdf (visited on 03/2021).

[25] Anirban DasGupta. “Normal approximations and the central limit theorem”. In: *Fundamentals of probability: A first course*. Springer, 2010, pp. 213–242.

[26] Elizabeth Cahill Delmelle and Irene Casas. “Evaluating the spatial equity of bus rapid transit-based accessibility patterns in a developing country: The case of Cali, Colombia”. In: *Transport Policy* 20 (2012), pp. 36–46. ISSN: 0967-070X. DOI: https://doi.org/10.1016/j.tranpol.2011.12.001. URL: https://www.sciencedirect.com/science/article/pii/S0967070X11001338.

[27] Flavio Devillaine, Marcela Munizaga, and Martin Trépanier. “Detection of Activities of Public Transport Users by Analyzing Smart Card Data”. In: *Transportation Research Record* 2276.1 (2012), pp. 48–55. DOI: 10.3141/2276-06. eprint: https://doi.org/10.3141/2276-06. URL: https://doi.org/10.3141/2276-06.

[28] Chuan Ding et al. “Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees”. In: *Sustainability (Switzerland)* 8.11 (2016). ISSN: 20711050. DOI: 10.3390/su8111100.

[29] Chuan Ding et al. "Using an ARIMA-GARCH Modeling Approach to Improve Subway Short-Term Ridership Forecasting Accounting for Dynamic Volatility". In: *IEEE Transactions on Intelligent Transportation Systems* 19.4 (2017). ISSN: 15249050. DOI: 10.1109/TITS.2017.2711046. URL: https://ieeexplore.ieee.org/document/7954025.

[30] Marco Dueñas, Mercedes Campi, and Luis Olmos. *Changes in mobility and socioeconomic conditions in Bogotá city during the COVID-19 outbreak*. Working papers 62. Red Investigadores de Economía, Sept. 2020. URL: https://ideas.repec.org/p/rie/riecdt/62.html.

[31] Mohamed El Mahrsi et al. "Understanding Passenger Patterns in Public Transit Through Smart Card and Socioeconomic Data: A case study in Rennes, France". In: Aug. 2014.

[32] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4 (1982). ISSN: 00129682. DOI: 10.2307/1912773.

[33] Robert W Fairlie. *The Impact of COVID-19 on Small Business Owners: The First Three Months after Social-Distancing Restrictions*. Working Paper 27462. National Bureau of Economic Research, July 2020. DOI: 10.3386/w27462. URL: http://www.nber.org/papers/w27462.

[34] M. Flórez et al. "Measuring the Impacts of Economic Well Being in Commuting Networks - A Case Study of Bogota, Colombia". In: 2017.

[35] Hongwei Gao et al. "Forecast of Passenger Flow Under the Interruption of Urban Rail Transit Operation". In: *Lecture Notes in Electrical Engineering*. Vol. 639. 2020. DOI: 10.1007/978-981-15-2866-8_27.

[36] Liping Ge, Stefan Voß, and Lin Xie. "Robustness and disturbances in public transport". In: *Public Transport* 14.1 (2022). ISSN: 16137159. DOI: 10.1007/s12469-022-00301-8.

[37] Nataša Glišović et al. "A hybrid model for forecasting the volume of passenger flows on Serbian railways". In: *Operational Research* 16.2 (2016). ISSN: 18661505. DOI: 10.1007/s12351-015-0198-5.

[38] Marco Gori, Dipartimento Ingegneria, and Gabriele Monfardini. "A new model for learning in graph domains". In: *IEEE International Joint Conference on Neural Networks*. 2005.

[39] Jianyuan Guo et al. "Short-Term Abnormal Passenger Flow Prediction Based on the Fusion of SVR and LSTM". In: *IEEE Access* 7 (2019). ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2907739.

[40] Luis A. Guzman and Juan P. Bocarejo. "Urban form and spatial urban equity in Bogota, Colombia". In: *Transportation Research Procedia* 25 (2017). World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016, pp. 4491–4506. ISSN: 2352-1465. DOI: https://doi.org/10.1016/j.trpro.2017.05.345. URL: https://www.sciencedirect.com/science/article/pii/S235214651730652X.

[41] Raia Hadsell et al. "Embracing Change: Continual Learning in Deep Neural Networks". In: *Trends in Cognitive Sciences* 24.12 (2020), pp. 1028–1040. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2020.09.004. URL: https://www.sciencedirect.com/science/article/pii/S1364661320302199.

[42] Gain Han and Keemin Sohn. "Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model". In: *Transportation Research Part B: Methodological* 83 (2016), pp. 121–135. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2015.11.015. URL: https://www.sciencedirect.com/science/article/pii/S0191261515002593.

[43] Yong Han et al. "Predicting Station-Level Short-Term Passenger Flow in a Citywide Metro Network Using Spatiotemporal Graph Convolutional Neural Networks". In: *ISPRS International Journal of Geo-Information* 8.6 (2019). ISSN: 2220-9964. DOI: 10.3390/ijgi8060243. URL: https://www.mdpi.com/2220-9964/8/6/243.

[44] Siyu Hao, Der-Horng Lee, and De Zhao. "Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system". In: *Transportation Research Part C: Emerging Technologies* 107 (2019), pp. 287–300. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2019.08.005. URL: https://www.sciencedirect.com/science/article/pii/S0968090X19300245.

[45] Samiul Hasan et al. "Spatiotemporal Patterns of Urban Human Mobility". In: *Journal of Statistical Physics* 151.1 (2013), pp. 304–318. ISSN: 1572-9613. DOI: 10.1007/s10955-012-0645-0. URL: https://doi.org/10.1007/s10955-012-0645-0.

[46] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

[47] Hainan Huang et al. "Identifying Subway Passenger Flow under Large-Scale Events Using Symbolic Aggregate Approximation Algorithm". In: *Transportation Research Record* 2676.2 (2022), pp. 800–810. DOI: 10.1177/03611981211047835. URL: https://doi.org/10.1177/03611981211047835.

[48] Jizhou Huang et al. "Understanding the Impact of the COVID-19 Pandemic on Transportation-Related Behaviors with Human Mobility Data". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 3443–3450. ISBN: 9781450379984. DOI: 10.1145/3394486.3412856. URL: https://doi.org/10.1145/3394486.3412856.

[49] Rob J Hyndman and Yeasmin Khandakar. "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software* 27.3 (2008), pp. 1–22. DOI: 10.18637/jss.v027.i03. URL: https://www.jstatsoft.org/index.php/jss/article/view/v027i03.

[50] O.J. Ibarra-Rojas et al. "Planning, operation, and control of bus transport systems: A literature review". In: *Transportation Research Part B: Methodological* 77 (2015), pp. 38–75. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2015.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0191261515000454.

[51] Institute for Transportation and Development Policy. *Post-Pandemic, Chinese Cities Gradually Reopen Transport Networks*. 2020. URL: https://www.itdp.org/2020/03/26/post-pandemic-chinese-cities-gradually-reopen-transport-networks/ (visited on 06/15/2020).

[52] Inter-American Development Bank. *Coronavirus Impact Dashboard*. 2020. URL: https://www.iadb.org/en/topics-effectiveness-improving-lives/coronavirus-impact-dashboard (visited on 06/15/2020).

[53] Yunjian Jia et al. "A Combined Forecasting Model for Passenger Flow Based on GM and ARMA". In: *International Journal of Hybrid Information Technology* 9.2 (2016). ISSN: 17389968. DOI: 10.14257/ijhit.2016.9.2.19.

[54] Yang Jiang, P. Christopher Zegras, and Shomik Mehndiratta. "Walk the line: Station context, corridor type and bus rapid transit walk access in Jinan, China". In: *Journal of Transport Geography* 20.1 (2012). ISSN: 09666923. DOI: 10.1016/j.jtrangeo.2011.09.007.

[55] Pengpeng Jiao et al. "Three Revised Kalman Filtering Models for Short-Term Rail Transit Passenger Flow Prediction". In: *Mathematical Problems in Engineering* 2016 (2016). ISSN: 15635147. DOI: 10.1155/2016/9717582.

[56] Yun Jing et al. "Short-Term Prediction of Urban Rail Transit Passenger Flow in External Passenger Transport Hub Based on LSTM-LGB-DRS". In: *IEEE Transactions on Intelligent Transportation Systems* 22.7 (2021). ISSN: 15580016. DOI: 10.1109/TITS.2020.3017109.

[57] Sungil Kim and Heeyoung Kim. "A new metric of absolute percentage error for intermittent demand forecasts". In: *International Journal of Forecasting* 32.3 (2016). ISSN: 01692070. DOI: 10.1016/j.ijforecast.2015.12.003.

[58] Gebhard Kirchgässner and Jürgen Wolters. *Introduction to modern time series analysis*. 2007. DOI: 10.1007/978-3-540-73291-4.

[59] Brennan Klein et al. "Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak". In: (2020).

[60] Cecilia S Lee and Aaron Y Lee. "Clinical applications of continual learning machine learning". In: *The Lancet Digital Health* 2.6 (2020), e279–e281. ISSN: 2589-7500. DOI: https://doi.org/10.1016/S2589-7500(20)30102-3. URL: https://www.sciencedirect.com/science/article/pii/S2589750020301023.

[61] Lu Li, Hong K. Lo, and Feng Xiao. "Mixed bus fleet scheduling under range and refueling constraints". In: *Transportation Research Part C: Emerging Technologies* 104 (2019). ISSN: 0968090X. DOI: 10.1016/j.trc.2019.05.009.

[62] Yang Li et al. "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks". In: *Transportation Research Part C: Emerging Technologies* 77 (2017), pp. 306–328. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2017.02.005. URL: http://www.sciencedirect.com/science/article/pii/S0968090X17300451.

[63] Yang Liu, Zhiyuan Liu, and Ruo Jia. "DeepPF: A deep learning based architecture for metro passenger flow prediction". In: *Transportation Research Part C: Emerging Technologies* 101 (2019), pp. 18–34. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2019.01.027. URL: http://www.sciencedirect.com/science/article/pii/S0968090X18306806.

[64] Yang Liu et al. "Automatic Feature Engineering for Bus Passenger Flow Prediction Based on Modular Convolutional Neural Network". In: *IEEE Transactions on Intelligent Transportation Systems* 22.4 (2021), pp. 2349–2358. DOI: 10.1109/TITS.2020.3004254.

[65] Ying Long and Jean-Claude Thill. "Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing". In: *Computers, Environment and Urban Systems* 53 (2015). Special Issue on Volunteered Geographic Information, pp. 19–35. ISSN: 0198-9715. DOI: https://doi.org/10.1016/j.compenvurbsys.2015.02.005. URL: https://www.sciencedirect.com/science/article/pii/S0198971515000356.

[66] Markus Löning et al. "sktime: A Unified Interface for Machine Learning with Time Series". In: (Sept. 2019). DOI: 10.48550/arxiv.1909.07872. arXiv: 1909.07872. URL: http://arxiv.org/abs/1909.07872.

[67] Yuhuan Lu et al. "Dual attentive graph neural network for metro passenger flow prediction". In: *Neural Computing and Applications* 33.20 (2021). ISSN: 14333058. DOI: 10.1007/s00521-021-05966-z.

[68] Zhenliang Ma et al. "Predicting short-term bus passenger demand using a pattern hybrid approach". In: *Transportation Research Part C: Emerging Technologies* 39 (2014), pp. 148–163. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2013.12.008. URL: https://www.sciencedirect.com/science/article/pii/S0968090X13002738.

[69] Aliasghar Mehdizadeh Dastjerdi and Catherine Morency. "Bike-Sharing Demand Prediction at Community Level under COVID-19 Using Deep Learning". In: *Sensors* 22.3 (2022). ISSN: 1424-8220. DOI: 10.3390/s22031060. URL: https://www.mdpi.com/1424-8220/22/3/1060.

[70] Miloš Milenković et al. "SARIMA modelling approach for railway passenger flow forecasting". In: *Transport* 33.5 (2018), pp. 1113–1120. DOI: https://doi.org/10.3846/16484142.2016.1139623. URL: https://journals.vgtu.lt/index.php/Transport/article/view/201.

[71] Harvey J. Miller et al. "Public transit generates new physical activity: Evidence from individual GPS and accelerometer data before and after light rail construction in a neighborhood of Salt Lake City, Utah, USA". In: *Health and Place* 36 (2015). ISSN: 18732054. DOI: 10.1016/j.healthplace.2015.08.005.

[72] MinSalud. *Determinantes sociales: elementos a tener en cuenta para la toma de decisiones en el país 2020*. Apr. 2020. URL: https://www.minsalud.gov.co/Paginas/Determinantes-sociales-elementos-a-tener-en-cuenta-para-la-toma-de-decisiones-en-el-pais.aspx (visited on 03/15/2020).

[73] Joseph Molloy et al. "MOBIS-COVID19/29. Results as of 30/11/2020 (second wave)". 2020.

[74] Ramon Munoz-Raskin. "Walking accessibility to bus rapid transit: Does it affect property values? The case of Bogotá, Colombia". In: *Transport Policy* 17.2 (2010), pp. 72–84. ISSN: 0967-070X. DOI: https://doi.org/10.1016/j.tranpol.2009.11.002.

[75] Duy Q. Nguyen-Phuoc et al. "Transit user reactions to major service withdrawal – A behavioural study". In: *Transport Policy* 64 (2018). ISSN: 1879310X. DOI: 10.1016/j.tranpol.2018.01.004.

[76] Sergio Arturo Ordóñez Medina. "Inferring weekly primary activity patterns using public transport smart card data and a household travel survey". In: *Travel Behaviour and Society* 12 (2018), pp. 93–101. ISSN: 2214-367X. DOI: https://doi.org/10.1016/j.tbs.2016.11.005. URL: https://www.sciencedirect.com/science/article/pii/S2214367X16300588.

[77] Alfonso Orro et al. "Impact on City Bus Transit Services of the COVID–19 Lockdown and Return to the New Normal: The Case of A Coruña (Spain)". In: *Sustainability* 12.17 (Sept. 2020), p. 7206. ISSN: 2071-1050. DOI: 10.3390/su12177206. URL: https://www.mdpi.com/2071-1050/12/17/7206.

[78] Luca Pappalardo et al. "scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data". In: (July 2019). arXiv: 1907.07062. URL: https://arxiv.org/abs/1907.07062.

[79] Ray Paternoster et al. "Using the Correct Statistical Test for Equality of Regression Coefficients". In: *Criminology* 36 (Nov. 1998), pp. 859–866. DOI: 10.1111/j.1745-9125.1998.tb01268.x.

[80] Ignacio Pérez-Messina et al. "Modalflow: Cross-Origin Flow Data Visualization for Urban Mobility". In: *Algorithms* 13.11 (Nov. 2020), p. 298. ISSN: 1999-4893. DOI: 10.3390/a13110298. URL: https://www.mdpi.com/1999-4893/13/11/298.

[81] Niklas Christoffer Petersen, Filipe Rodrigues, and Francisco Camara Pereira. "Multi-output bus travel time prediction with convolutional LSTM neural network". In: *Expert Systems with Applications* 120 (2019). ISSN: 09574174. DOI: 10.1016/j.eswa.2018.11.028.

[82] Daniel Piner and Benjamin Condry. "International best practices in managing unplanned disruption to suburban rail services". In: *Transportation Research Procedia*. Vol. 25. 2017. DOI: 10.1016/j.trpro.2017.05.331.

[83] Ivens Portugal, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review". In: *Expert Systems with Applications* 97 (2018), pp. 205–227. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2017.12.020. URL: https://www.sciencedirect.com/science/article/pii/S0957417417308333.

[84] Ehsan Rahimi et al. "Analysis of Transit Users' Response Behavior in Case of Unplanned Service Disruptions". In: *Transportation Research Record* 2674.3 (2020). ISSN: 21694052. DOI: 10.1177/0361198120911921.

[85] Ehsan Rahimi et al. "Analysis of transit users' waiting tolerance in response to unplanned service disruptions". In: *Transportation Research Part D: Transport and Environment* 77 (2019). ISSN: 13619209. DOI: 10.1016/j.trd.2019.10.011.

[86] Chris Rissel et al. "Physical activity associated with public transport use-a review and modelling of potential benefits". In: *International Journal of Environmental Research and Public Health* 9.7 (2012). ISSN: 16604601. DOI: 10.3390/ijerph9072454.

[87] Jérémy Roos, Gérald Gavin, and Stéphane Bonnevay. "A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data". In: *Transportation Research Procedia* 26 (2017), pp. 53–61. ISSN: 23521465. DOI: 10.1016/j.trpro.2017.07.008. URL: http://dx.doi.org/10.1016/j.trpro.2017.07.008.

[88] Daniela Roque and Houshmand E. Masoumi. "An Analysis of Car Ownership in Latin American Cities: a Perspective for Future Research". In: *Periodica Polytechnica Transportation Engineering* 44.1 SE - (Apr. 2016), pp. 5–12. DOI: 10.3311/PPtr.8307. URL: https://pp.bme.hu/tr/article/view/8307.

[89] Muhammad Sajjad et al. "Towards efficient building designing: Heating and cooling load prediction via multi-output model". In: *Sensors (Switzerland)* 20.22 (2020). ISSN: 14248220. DOI: 10.3390/s20226419.

[90] Secretaria Distrital de Movilidad Bogotá. *Household travel survey 2019*. 2019. URL: https://www.simur.gov.co/portal-simur/datos-del-sector/encuestas-de-movilidad/ (visited on 03/15/2020).

[91] Secretaría Distrital de Planeación Bogotá. *Rutas Zonales del SITP*. 2020. URL: https://datosabiertos.bogota.gov.co/dataset/manzana-estratificacion-bogota-d-c (visited on 06/13/2020).

[92] Shouwei Sha et al. "RNN-Based Subway Passenger Flow Rolling Prediction". In: *IEEE Access* 8 (2020), pp. 15232–15240. DOI: 10.1109/ACCESS.2020.2964680.

[93] Loutao Shen et al. "Hybrid Approach Combining Modified Gravity Model and Deep Learning for Short-Term Forecasting of Metro Transit Passenger Flows". In: *Transportation Research Record* 2675.1 (2021), pp. 25–38. DOI: 10.1177/0361198120968823.

[94] Taylor G Smith et al. *pmdarima: ARIMA estimators for Python*. 2017. URL: http://www.alkaline-ml.com/pmdarima.

[95] Christian Spreafico and Davide Russo. "Exploiting the scientific literature for performing life cycle assessment about transportation". In: *Sustainability (Switzerland)* 12.18 (2020). ISSN: 20711050. DOI: 10.3390/su12187548.

[96] Nitish Srivastava et al. "Dropout: A simple way to prevent neural networks from overfitting". In: *Journal of Machine Learning Research* 15 (2014). ISSN: 15337928.

[97] Todor Stojanovski. "How density, diversity, land use and neighborhood type influences bus mobility in the swedish city of karlstad: Mixing spatial analytic and typo-morphological approaches to assess the indirect effect of urban form on travel". In: *Journal of Transport and Land Use* 11.1 (2018). ISSN: 19387849. DOI: 10.5198/jtlu.2018.1089.

[98] Andy Sumner, Chris Hoy, and Eduardo Ortiz-Juarez. "Estimates of the impact of COVID-19 on global poverty". In: *Unuwider* 2020.April (Apr. 2020), pp. 1–9. DOI: https://doi.org/10.35188/UNU-WIDER/2020/800-9. URL: https://doi.org/10.35188/UNU-WIDER/2020/800-9.

[99] Huijun Sun et al. "Estimating the influence of common disruptions on urban rail transit networks". In: *Transportation Research Part A: Policy and Practice* 94 (2016). ISSN: 09658564. DOI: 10.1016/j.tra.2016.09.006.

[100] Nan Sun et al. "Near real-time twitter spam detection with machine learning techniques". In: *International Journal of Computers and Applications* 44.4 (2022), pp. 338–348. DOI: 10.1080/1206212X.2020.1751387. URL: https://doi.org/10.1080/1206212X.2020.1751387.

[101] Yujuan Sun, Guanghou Zhang, and Huanhuan Yin. "Passenger flow prediction of subway transfer stations based on nonparametric regression model". In: *Discrete Dynamics in Nature and Society* 2014 (2014). ISSN: 1607887X. DOI: 10.1155/2014/397154.

[102] Yuxing Sun, Biao Leng, and Wei Guan. “A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system”. In: *Neurocomputing* 166 (2015). ISSN: 18728286. DOI: 10.1016/j.neucom.2015.03.085.

[103] João Filipe Teixeira and Miguel Lopes. “The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York’s Citi Bike”. In: *Transportation Research Interdisciplinary Perspectives* 6 (2020), p. 100166. ISSN: 2590-1982. DOI: https://doi.org/10.1016/j.trip.2020.100166. URL: http://www.sciencedirect.com/science/article/pii/S2590198220300774.

[104] Florian Toque et al. “Short long term forecasting of multimodal transport passenger flows with machine learning methods”. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. Vol. 2018-March. 2018. DOI: 10.1109/ITSC.2017.8317939.

[105] TransitApp. *How Coronavirus is disrupting public transit*. 2020. URL: https://transitapp.com/coronavirus (visited on 06/15/2020).

[106] TransitApp. *Who’s left riding public transit? A COVID data deep-dive.* 2020. URL: https://medium.com/transit-app/whos-left-riding-public-transit-hint-it-s-not-white-people-d43695b3974a.

[107] Transmilenio S.A. 2020. URL: https://datosabiertos-transmilenio.hub.arcgis.com (visited on 10/13/2020).

[108] Transmilenio S.A. *Estadísticas de oferta y demanda del Sistema Integrado de Transporte Público - SITP - Febrero*. 2020. URL: https://www.transmilenio.gov.co/publicaciones/151672/estadisticas-de-oferta-y-demanda-del-sistema-integrado-de-transporte-publico-sitp-febrero-2020/ (visited on 03/2021).

[109] Tsung Hsien Tsai, Chi Kang Lee, and Chien Hung Wei. “Neural network based temporal feature models for short-term railway passenger demand forecasting”. In: *Expert Systems with Applications* 36.2 PART 2 (2009). ISSN: 09574174. DOI: 10.1016/j.eswa.2008.02.071.

[110] Union Internationale des Transports Publics (UITP). *Management of COVID-19 guidelines for public transport operators*. 2020. URL: https://www.uitp.org/publications/management-of-covid-19-guidelines-for-public-transport-operators/ (visited on 02/2020).

[111] N.Job A. van Exel and Piet Rietveld. “Public transport strikes and traveller behaviour”. In: *Transport Policy* 8.4 (2001), pp. 237–246. ISSN: 0967-070X. DOI: https://doi.org/10.1016/S0967-070X(01)00022-1. URL: https://www.sciencedirect.com/science/article/pii/S0967070X01000221.

[112] Ashish Vaswani et al. “Attention Is All You Need”. In: (June 2017). DOI: 10.48550/arxiv.1706.03762. arXiv: 1706.03762. URL: http://arxiv.org/abs/1706.03762.

[113] C. Erik Vergel-Tovar and Daniel A. Rodriguez. *The ridership performance of the built environment for BRT systems: Evidence from Latin America*. 2018. DOI: 10.1016/j.jtrangeo.2018.06.018.

[114] Chengpeng Wan et al. "Resilience in transportation systems: a systematic review and future directions". In: *Transport Reviews* 38.4 (2018). ISSN: 14645327. DOI: 10.1080/01441647.2017.1383532.

[115] Ding Wang et al. "Impact of COVID-19 Behavioral Inertia on Reopening Strategies for New York City Transit". In: *CoRR* abs/2006.13368 (2020). arXiv: 2006.13368. URL: https://arxiv.org/abs/2006.13368.

[116] Haiyang Wang et al. "Early warning of burst passenger flow in public transportation system". In: *Transportation Research Part C: Emerging Technologies* 105 (2019), pp. 580–598. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2019.05.022.

[117] Xueqin Wang et al. "An effective spatiotemporal deep learning framework model for short-term passenger flow prediction". In: *Soft Computing* 26.12 (2022), pp. 5523–5538. ISSN: 1433-7479. DOI: 10.1007/s00500-022-07025-8. URL: https://doi.org/10.1007/s00500-022-07025-8.

[118] Yonggang Wang, Jingfeng Ma, and Jun Zhang. "Metro passenger flow forecast with a novel Markov-grey model". In: *Periodica Polytechnica Transportation Engineering* 48.1 (2019). ISSN: 15873811. DOI: 10.3311/PPtr.11131.

[119] Yuan Wang et al. "A Data-Driven and Optimal Bus Scheduling Model with Time-Dependent Traffic and Demand". In: *IEEE Transactions on Intelligent Transportation Systems* 18.9 (2017). ISSN: 15249050. DOI: 10.1109/TITS.2016.2644725.

[120] Yu Wei and Mu-Chen Chen. "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks". In: *Transportation Research Part C: Emerging Technologies* 21.1 (2012), pp. 148–162. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2011.06.009.

[121] Michael Wilbur et al. "Impact of COVID-19 on Public Transit Accessibility and Ridership". In: (Aug. 2020). arXiv: 2008.02413. URL: http://arxiv.org/abs/2008.02413.

[122] World Health Organization. *Supporting healthy urban transport and mobility in the context of COVID-19*. 2020. URL: https://www.who.int/publications/i/item/9789240012554 (visited on 11/02/2020).

[123] Weitiao Wu et al. "Stochastic bus schedule coordination considering demand assignment and rerouting of passengers". In: *Transportation Research Part B: Methodological* 121 (2019). ISSN: 01912615. DOI: 10.1016/j.trb.2019.01.010.

[124] Binglei Xie et al. "Travel characteristics analysis and passenger flow prediction of intercity shuttles in the pearl river delta on holidays". In: *Sustainability (Switzerland)* 12.18 (2020). ISSN: 20711050. DOI: 10.3390/SU12187249.

[125] Hannah Younes et al. “How transit service closures influence bikesharing demand; lessons learned from SafeTrack project in Washington, D.C. metropolitan area”. In: *Journal of Transport Geography* 76 (2019). ISSN: 09666923. DOI: 10.1016/j.jtrangeo.2019.03.004.

[126] Yuan Yuan et al. “Urban rail transit passenger flow forecasting method based on the coupling of artificial fish swarm and improved particle swarm optimization algorithms”. In: *Sustainability (Switzerland)* 11.24 (2019). ISSN: 20711050. DOI: 10.3390/SU11247230.

[127] Chun Hui Zhang, Rui Song, and Yang Sun. “Kalman filter-based short-term passenger flow forecasting on bus stop”. In: *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology* 11.4 (2011). ISSN: 10096744.

[128] Jinlei Zhang, Feng Chen, and Qing Shen. “Cluster-Based LSTM Network for Short-Term Passenger Flow Forecasting in Urban Rail Transit”. In: *IEEE Access* 7 (2019). ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2941987.

[129] Jinlei Zhang et al. “Deep Learning Architecture for Short-Term Passenger Flow Forecasting in Urban Rail Transit”. In: *IEEE Transactions on Intelligent Transportation Systems* 22.11 (2021). ISSN: 15580016. DOI: 10.1109/TITS.2020.3000761.

[130] Jinlei Zhang et al. “Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method”. In: *Transportation Research Part C: Emerging Technologies* 124 (2021), p. 102928. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2020.102928. URL: https://www.sciencedirect.com/science/article/pii/S0968090X20308275.

[131] Wenjie Zhang et al. “Short-Time Prediction of Subway Inbound Passenger Flow Based on K-means Clustering Combination Model”. In: *Lecture Notes in Electrical Engineering*. Vol. 901 LNEE. 2022. DOI: 10.1007/978-981-19-2259-6_62.

[132] Jinbao Zhao et al. “What influences Metro station ridership in China? Insights from Nanjing”. In: *Cities* 35 (2013). ISSN: 02642751. DOI: 10.1016/j.cities.2013.07.002.

[133] Yangyang Zhao and Zhenliang Ma. “Naïve Bayes-Based Transition Model for Short-Term Metro Passenger Flow Prediction under Planned Events”. In: *Transportation Research Record* 2676.9 (2022), pp. 309–324. DOI: 10.1177/03611981221086645. URL: https://doi.org/10.1177/03611981221086645.

[134] Zhan Zhao, Haris N Koutsopoulos, and Jinhua Zhao. “Individual mobility prediction using transit smart card data”. In: *Transportation Research Part C: Emerging Technologies* 89 (2018), pp. 19–34. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2018.01.022. URL: http://www.sciencedirect.com/science/article/pii/S0968090X18300676.

[135] Yanlai Zhou et al. "Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting". In: *Science of the Total Environment* 651 (2019), pp. 230–240. ISSN: 18791026. DOI: 10.1016/j.scitotenv.2018.09.111. URL: https://www.sciencedirect.com/science/article/pii/S0048969718335538.

[136] Yirong Zhou et al. "A spatiotemporal hierarchical attention mechanism-based model for multi-step station-level crowd flow prediction". In: *Information Sciences* 544 (2021). ISSN: 00200255. DOI: 10.1016/j.ins.2020.07.049.