Unveiling Urban Mobility Patterns: A Data-Driven Analysis of Public Transit

Oluwaleke Yusuf, Adil Rasheed
Department of Engineering Cybernetics
Norwegian University of Science
and Technology
7034 Trondheim, Norway
Email: oluwaleke.u.yusuf@ntnu.no, adil.rasheed@ntnu.no

Frank Lindseth
Department of Computer Science
Norwegian University of Science
and Technology
7034 Trondheim, Norway
Email: frankl@ntnu.no

Martin Slaastuen
Market Insight
AtB AS
7011 Trondheim, Norway
Email: martin.slaastuen@atb.no

Abstract—The expansion of urban centers necessitates enhanced efficiency and sustainability in their transportation infrastructure and mobility systems. The big data obtainable from various transportation modes potentially offers critical insights for urban planning. This study presents analysis of detailed historical public transit data, enriched with relevant temporal and geospatial metadata, as a precursor to injecting dynamism into digital twins of mobility systems via ML/DL-based predictive modeling. A data preprocessing framework was implemented to refine the raw data for effective historical analysis and predictive modeling. This paper examines public transit data for patterns and trends—incorporating factors such as time, geospatial elements, external influences, and operational aspects. From a technical standpoint, this research helps to assess the quality of the available transit data and identify important information for use in digital twins. Such digital twins foster educated decisions for efficient, sustainable urban mobility systems by anticipating infrastructure demand, identifying service gaps, and understanding mobility dynamics.

Index Terms—Urban Mobility Analysis, Public Transportation Data, Digital Twins, Smart Mobility Solutions.

I. INTRODUCTION

Trondheim, the third most populous municipality in Norway, is an ever-growing urban center in Trøndelag county, home to approximately 206,000 inhabitants as of 2023. This growth—fueled by its status as an educational and research hub with institutions like the Norwegian University of Science and Technology (NTNU), SINTEF, and St. Olav's University Hospital—places increasing demands on its transportation infrastructure. With the resultant increase in car traffic and air pollution, the focus of city planners has shifted towards sustainable transport, with initiatives like Miljøpakken and MobilitetsLab Stor-Trondheim emphasizing public transport and active mobility over driving.

The increasing reliance on public transit in urban areas like Trondheim has spurred significant academic interest in analyzing public transit data. Researchers utilize varied data sources — including Automated Passenger Counting (APC) [1], General Transit Feed Specification (GTFS) [1]–[4], Integrated Circuit (IC) card [5]–[7] — to glean insights into urban transit efficiency, punctuality, and service reliability.

These analyses inform infrastructure planning and policy-making, contributing to the development of more efficient and sustainable transportation infrastructure and mobility systems.

Recent literature has explored diverse aspects of public transit. Studies like Zhang et al. [2] have identified key transit corridors using smart card data, while Yao et al. [8] introduced models to measure public transit systems' performance, considering multiple stakeholders and environmental factors. Huang et al. [5] proposed frameworks for analyzing bus network resilience, and Li et al. [6] examined passenger travel behavior in urban transit corridors. These studies, along with others [1], [3], [4], [7], provide a foundation for understanding the complexities of urban mobility and the need for data-driven approaches for infrastructure planning and policy-making to enhance public transit systems.

Within this context, a Digital Twin (DT) provides a comprehensive technological framework for aggregating real-time data from transportation infrastructure and mobility systems into a unified digital representation. By integrating real-world information with a visual representation of the entire system, the DT achieves a realistic mimicry of the system's state and behavior. This level of digitalization allows for in-depth analysis of historical trends for decision-making, real-time analytics for automated responses and emergency coordination, predictive modelling for forecasting and advanced simulations for exploratory "What-if?" analyses.

However, the intricate spatiotemporal dynamics of urban mobility systems and the interplay between various modes of transportation make analyzing past data and forecasting future trends a daunting task. As a foundation for traffic modeling similar to [9] for applications in a DT of Trondheim's mobility system, this study seeks to tackle these challenges by initially focusing on the historical analysis of public transit data. This analysis assesses the utility of such data in understanding the system's dynamics and interdependencies to facilitate its automation through a DT.

II. DATA

This study is based on data obtained from AtB, Trøndelag's public transport administrator, via APC systems installed on
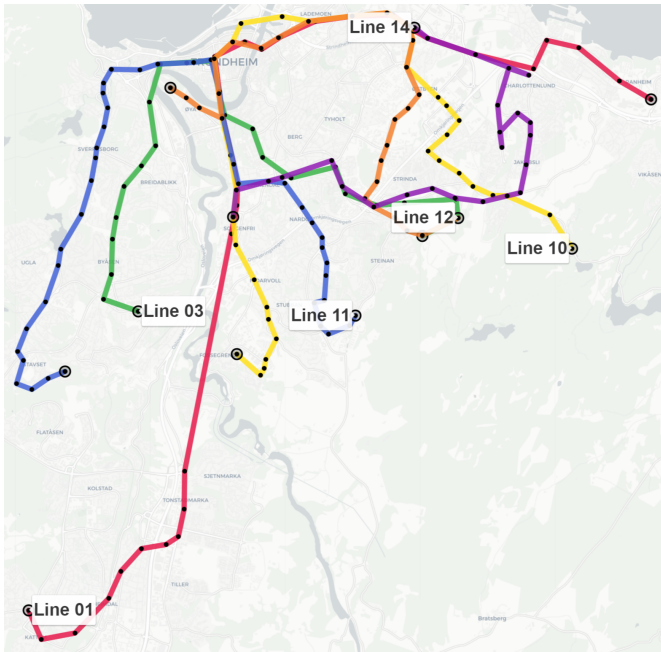
Fig. 1. Satellite map of the AtB bus lines in the dataset.

buses—which employ DILAX optical sensors with 3D stereoscopic vision technology—for passenger counting. Despite occasional technical issues leading to counting discrepancies [10], the APC data, after preprocessing, reliably reflects passenger trends and patterns across the bus transit system.

A. Data Description

The study's dataset covers the period from 1st May 2020 to 30th November 2023, comprising 1,179,770 unique trips over 1,295 days, across 6 lines and 204 stops in Trondheim Municipality. As shown in Figure 1, the bus lines in the data cover a wide section of the city, with most of them routing through the city center. The raw dataset features a comprehensive set of 63 attributes, offering insights into the temporal and spatial aspects of bus trips, passenger flow metrics, and various operational parameters.

- **Temporal Information**: Includes date and time features, such as *TripScheduledDeparture*, *TripDepartureHour*, *StopScheduledArrival* and *StopActualArrival*. These capture the scheduled and actual timings for departures and arrivals for specific stops and entire trips.
- **Operational Flags**: A series of conditional flags such as *StopTime*, *FLAG_Trip_15minDeviation* and *FLAG_Stop_IsDelayed* offer insights into operational issues.
- **Spatial Information**: Encoded via *StopName*, *Longitude*, and *Latitude*, this provides a spatial context to each trip.
- **Passenger Flow Metrics**: The *Boarding*, *Alighting*, and *Onboard* fields represent data from the APC system, which records the volume of passenger traffic.
- **Vehicle and Service Information**: Details such as *BusType* and *CarriageNumber* provide data on the assets used and the .

- **Special Dates**: The dataset is enriched with flags such as *FLAG_Holiday_Restday* and *FLAG_SchoolVacation* and descriptive fields such as *DayComment* and *HolidayName* to denote public holidays and school breaks, allowing for nuanced analyses of service variations on special days.

B. Data Preprocessing

As previously mentioned, the dataset's foundation is the APC data, enriched with pertinent operational and spatiotemporal details. The raw dataset was preprocessed to address missing values across the dataset and correct negative values in the APC and temporal data, as outlined below:

1) *APC Features*: The dataset primarily includes the *Boarding* and *Alighting* features from the APC system, representing passengers getting on and off the bus at each stop. These figures are aggregated from counts at each door. Derived from these, new metrics like passenger volume on the bus between stops and the change in passenger count throughout the trip are calculated.
   Data cleaning involved: *(i)* Setting missing values in *Boarding* and *Alighting* to zero. *(ii)* Discarding erroneous derivative APC values and engineering three new features *busVolume*, *tripSumVolume*, and *stopVolume* which denote the passenger count onboard, cumulative boardings since the trip's start, and count at each stop, respectively. *(iii)* Trips with negative *Boarding* or *Alighting* values or where flags *FLAG_Trip_APCActive* or *FLAG_Trip_APCExtremeValue* were *True* were excluded from the dataset.
2) *Temporal Information*: For bus schedule features *TripScheduledArrival*, *StopActualArrival*, and *StopActualDeparture*, empty values were replaced with corresponding scheduled values. Instances where *StopActualDeparture* was earlier than *StopActualArrival* were corrected and the *StopTime* subsequently recalculated to reflect this correction.
3) *Missing Values*: Missing values for the conditional operational flags were determined using expert knowledge and set accordingly. Missing *BusType*, *DayComment*, and *HolidayName* values were filled with "No Information", while modal values were used for missing *Municipality* and *StopType* data.
4) *Unique Trip Identifiers*: The *Trip* feature, a daily unique identifier, was combined with the *Date* feature to create a dataset-wide unique identifier *dateTrip*.

In the final preprocessing step, we removed features deemed irrelevant, outdated (as confirmed by expert consultation), or redundant. The resultant dataset includes 45 features and 1,130,100 unique trips, representing a 4.21% reduction from the original dataset. This reduction primarily addresses inaccuracies in the APC data, notably from May to September 2020, a period shortly after AtB deployed the APC system, and its bugs were being ironed out.
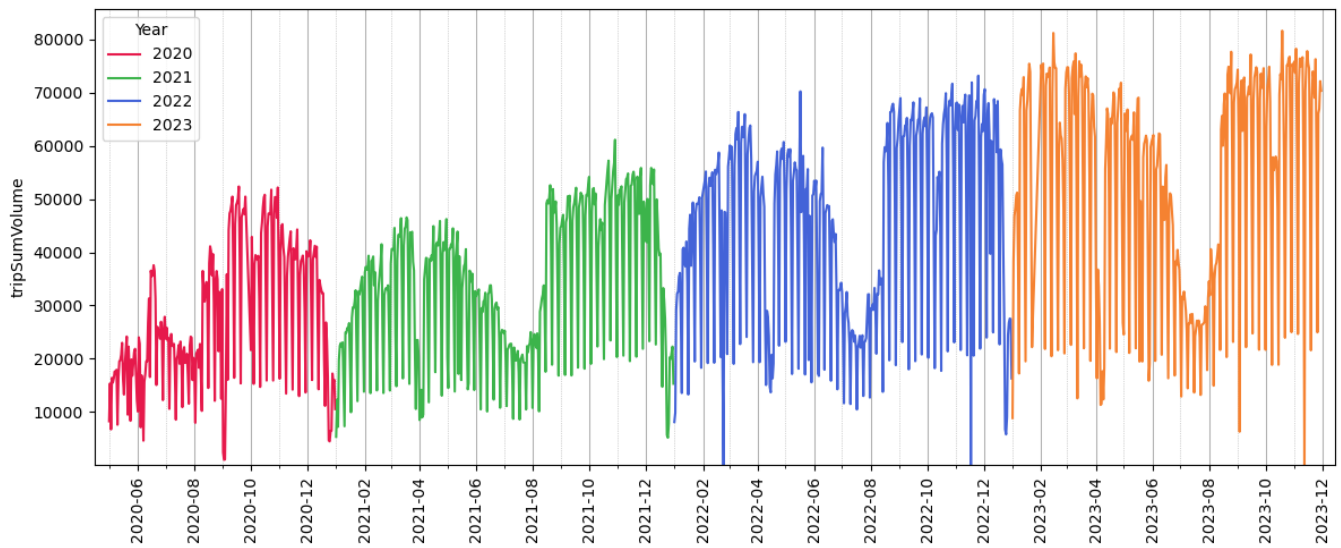
Fig. 2. Yearly breakdown of passenger volume for the entire dataset from May 2020 to November 2023.
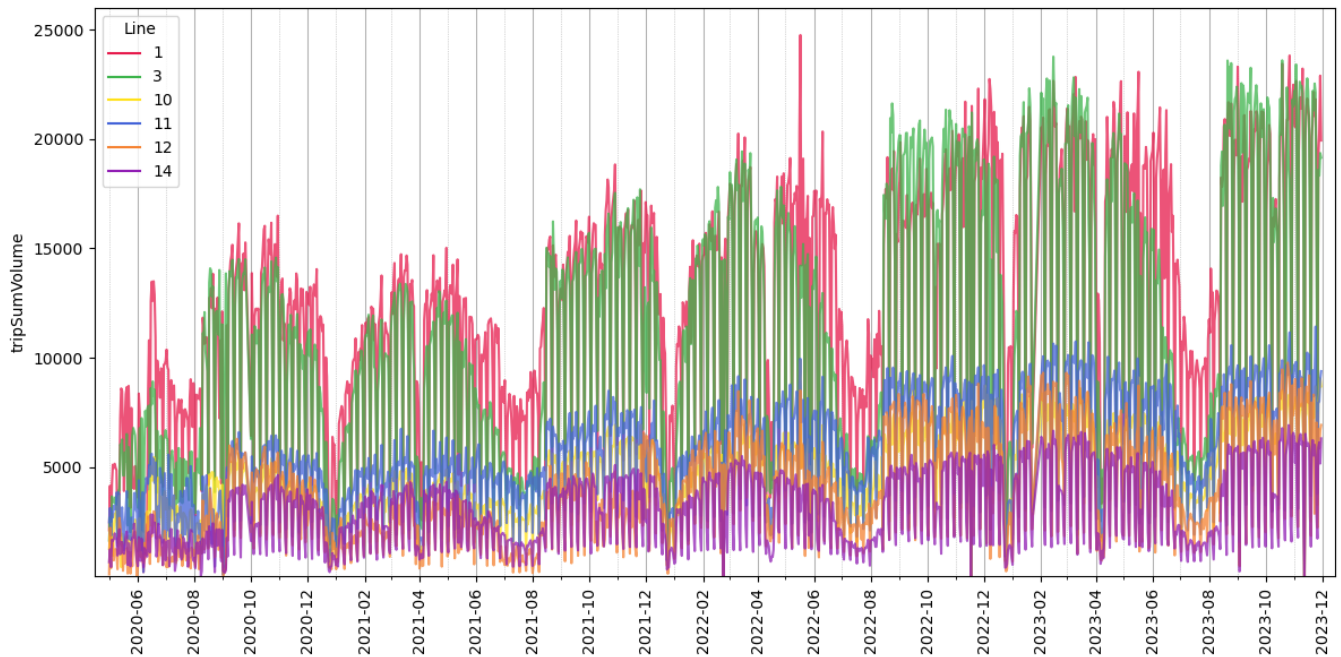

Fig. 3. Detailed overview of passenger volume across different bus lines for the entire dataset.

III. ANALYTICS AND DISCUSSION

This section interprets the intricate dynamics of Trondheim's bus transit system through APC data, focusing on passenger volume over time. The analysis encompasses temporal fluctuations, the impact of external factors, and operational efficiencies within the network.

A. Temporal Patterns

The temporal analysis of the data is performed at various levels—hourly, daily, weekly, and monthly—using the mean value for each level. This analysis specifically considers data from 1st March 2022 to 30th November 2023, excluding the Covid-19 period, which is examined separately in Section III-B.

1) Total Daily Volume: Analyzing the total daily passenger volume for the entire system from 1st May 2020 until 30th November 2023, as illustrated in Figure 2, reveals a general upward trend in passenger volume, influenced by Covid-19 recovery and population increases in Trondheim. The data shows notable dips in volume during holidays such as Easter, Summer, Christmas, and New Year. Additionally, the highest daily volumes correspond to specific events like the annual UKA Student Festival and St. Valentine's Day.

Further breaking down the daily passenger volume by line, as shown in Figure 3, highlights variations in the passenger volume experienced by each bus line over time. Line 1 handled a total volume of 16,050,047 passengers, closely followed by Line 3 at 14,282,754 passengers with Line 11 a distant
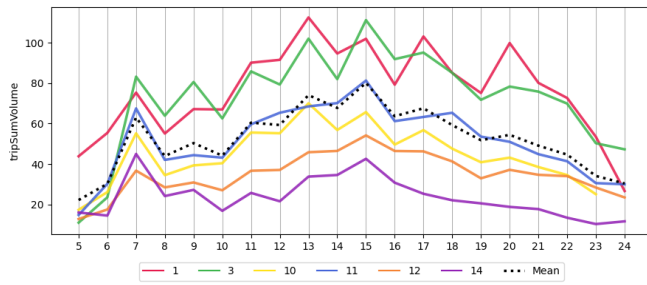
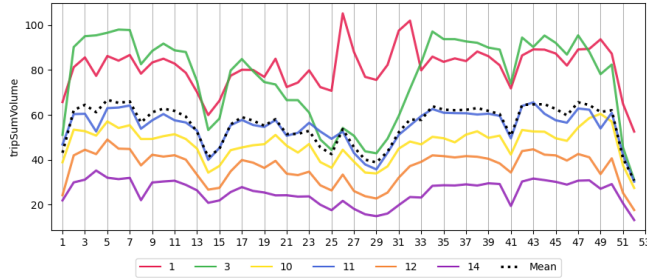Fig. 4. Hourly variation of passenger volume across bus lines.

Fig. 5. Daily variation of passenger volume across bus lines.

Fig. 6. Weekly variation of passenger volume across bus lines.

Fig. 7. Monthly variation of passenger volume across bus lines.

third at 7,169,532 passengers. Lines 1 and 3 exhibit distinct patterns, particularly around major holiday periods, with Line 3's volume affected by the NTNU academic calendar.

2) Average Hourly and Daily Volumes: The average hourly passenger volumes, displayed in Figure 4, show a consistent morning peak at 07:00 across all lines, with an additional surge in Line 3 due to NTNU traffic. The afternoon peaks are dispersed between 14:00 and 16:00, and a third peak at 20:00 is more pronounced for Line 1, likely due to its proximity to dining and entertainment areas.

The daily passenger volumes, as depicted in Figure 5, are relatively stable throughout the week with a noticeable dip on Sundays. Lines 1 and 3 display contrasting weekend patterns, possibly reflecting a shift in student traffic on weekends.

3) Average Weekly and Monthly Volumes: The average weekly and monthly passenger volumes, as detailed in Figures 6 and 7, generally mirror each other, with weekly data offering more granular insights. Lines 10, 11, 12, and 14 exhibit similar trends at both the weekly and monthly levels. Seasonal trends are evident, such as dips during holidays and spikes in travel activity at specific times of the year. Lines 1 and 3 show deviations from these patterns due to their unique passenger demographics and the seasonal impact of NTNU's academic schedule.

B. External Influences

This subsection delves into how external events, including the COVID-19 pandemic and major holidays, influence the daily passenger volume across Trondheim's bus transit system.

1) Covid-19 Pandemic: The pandemic's impact on passenger volumes is evident in Figures 2 and 3. The Norwegian government's Covid-19 policies, when mapped against the bus transit data as displayed in Figure 8, show a clear correlation
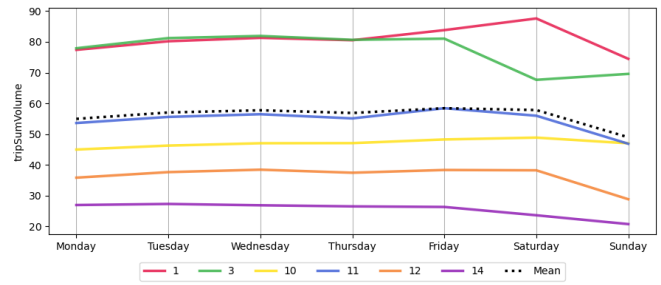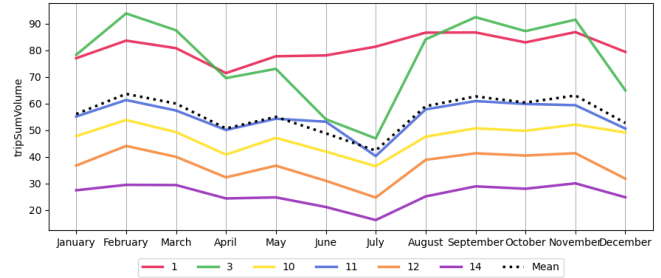
between policy changes and passenger volumes, justifying the exclusion of this period from earlier analyses.

2) Major Holidays and Local Events: To understand the effects of holidays and local events on average daily passenger volumes, the data for 2022 and 2023 is plotted with overlaid colored bands marking significant dates, as seen in Figure 9. Generally, passenger volume drops on these dates, with the notable exception of "Constitution Day," which causes a surge in volume. The comparison between 2022 and 2023 shows initial discrepancies, likely due to COVID-19's lingering effects, but a gradual alignment in patterns as the year progresses.

C. Operational Trends

1) Trip Counts and Passenger Volumes: Figure 10 presents the total bus trips and passenger volumes for each line. Echoing findings from Section III-A1, Line 1 manages the highest number of trips and passengers, closely followed by Line 3. A noteworthy observation is the contrast between the trip counts and passenger volumes for Lines 10, 11, 12, and 14. Lines 11 and 10, despite similar trip counts, see Line 11 carrying more passengers. A similar disparity exists between Lines 12 and 14, with Line 14 transporting more passengers. This indicates a potential misalignment between the scheduling supply and actual passenger demand on Lines 10 and 14, often resulting in underutilized buses.

Table I ranks the top 5 bus types used in Trondheim by AtB, based on passenger volume and trip count. Dominating this list are the Van Hool ExquiCity buses, primarily servicing the high-demand Lines 1 and 3. This table also sheds light on AtB's efforts to enhance the sustainability of the bus transit system with hybrid, electric, and biofuel buses, though fossil fuels remain predominant. This reliance is an area of concern for future adaptation to the growing population and increasing demand placed on the bus system.
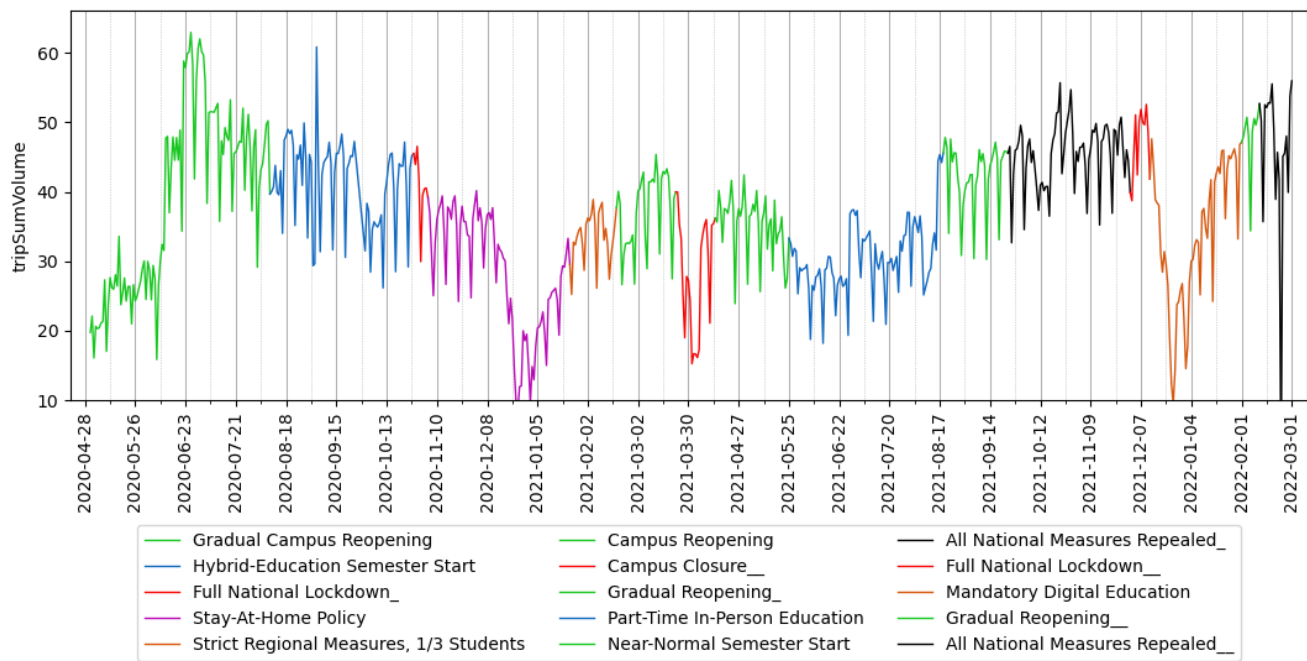
Fig. 8. Effect of Norwegian Covid-19 policies on passenger volume from May 2020 to March 2022.
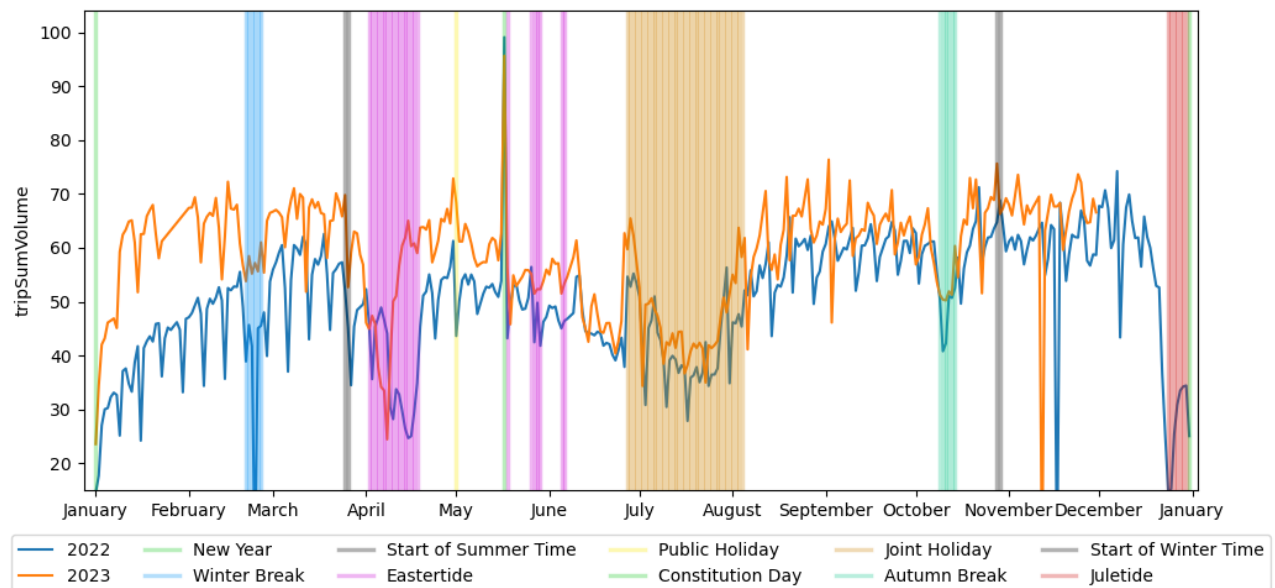

Fig. 9. Variation of passenger volume with major holidays and local events in 2022 and 2023.

Further analysis of individual vehicles in the AtB fleet shows the top-10 chassis are all Van Hool ExquiCity 24m Diesel-Hybrid models, predominantly operated on Line 3. These buses, averaging 15,925 trips and conveying 1,035,355 passengers, are the fleet's workhorses. Their significant usage underscores the importance of rigorous maintenance schedules to extend their operational lifespan and ensure system reliability.

2) Trip Deviations and System Efficiency: A key aspect of assessing the bus transit system's efficiency is examining the deviations between scheduled trip completion times and actual end times, i.e., the bus's arrival at the last stop. These trip-level

TABLE I
COMPARISON OF TRIP COUNTS AND PASSENGER VOLUMES HANDLED BY DIFFERENT BUS TYPES.

Bus Type	Fuel	Trips	Volume
Van Hool ExquiCity	Diesel-Hybrid	457,964	30,250,790
MAN Lion's City 18	Diesel	334,790	11,381,406
MAN Lion's City G A23	Biodiesel/CNG	129,001	5,023,050
Heuliez GX 437	Electric	154,053	4,440,077
MAN Lion's City A21	Biogas/CNG	22,241	746,516

deviations are an aggregate of cumulative stop-level deviations along the route. Positive deviations indicate late arrivals, while
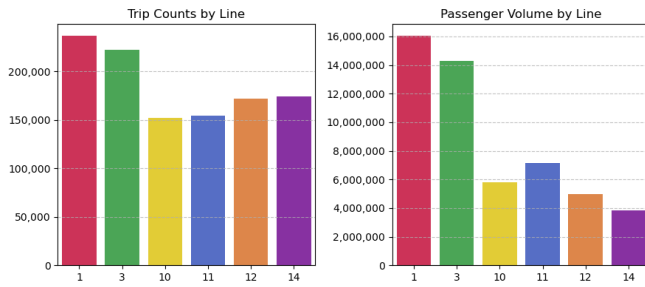
Fig. 10. Comparison of trip counts and passenger volumes handled by different bus lines.

Fig. 11. Trip scheduling deviations for different bus lines.

negative ones signify early arrivals. Figure 11 illustrates the distribution of these trip deviations in seconds for each line. The system-wide average deviation stands at 176 seconds, with Line 14 performing best at an average of 120 seconds, and Line 10 lagging with an average deviation of 217 seconds.

The root causes of trip deviations can vary, with passenger volume and road traffic being prime factors. Line 14, which has the least passenger volume, consequently spends less time at stops, leading to its lower deviation. However, an interesting anomaly arises with Line 10. As discussed in Section III-C1, despite having a lower passenger volume compared to Line 11, Line 10's average deviation is significantly higher. This discrepancy suggests that other factors, possibly specific to Line 10's route or operational dynamics, contribute to its higher deviation, warranting further investigation to fully understand these patterns.

IV. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper analyzed mobility data from Trondheim's public bus transportation system spanning from May 2020 until November 2023. It sheds light on the shifting dynamics of passenger volume against the backdrop of the city's growing population and escalating demands on the bus system. The study of temporal patterns across specific bus lines offers insights into the characteristics of various city areas and their demographics, particularly as said areas and demographics evolve over time. In addition, our analysis reveals the impact of external factors, notably the varied government responses to the COVID-19 pandemic, on mobility trends. This research contributes to a broader research project that aims to create a multi-modal Digital Twin (DT) of Trondheim's mobility system, instilled with dynamism via real-world data, for enhancing the efficiency and sustainability of future urban mobility solutions.

B. Future Work

Future work aims at developing automated AI-based solutions for mobility systems, incorporating various machine (and deep) learning techniques for geospatial analysis, dimensionality reduction, predictive modelling and synthetic data generation for use in DTs. This initiative will involve the use of multi-modal mobility data, thereby augmenting the DT with extensive data-driven historical analyses and
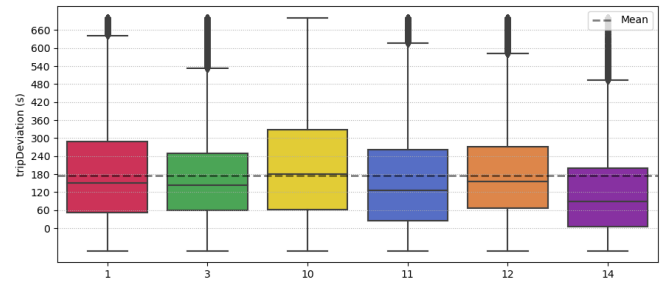
predictive models across the entire transportation infrastructure and mobility system. Additionally, enhancements will be made to the accuracy of the APC data by integrating onboard camera footage for more precise passenger counts. This enhanced data will underpin trip-level and stop-level predictive models for passenger volume at various temporal scales. The resulting insights and models will be instrumental in strengthening the resilience, sustainability, and efficiency of Trondheim's bus transit system as it adapts to an increasing population.

REFERENCES

[1] L. Liu, H. J. Miller, Measuring risk of missing transfers in public transit systems using high-resolution schedule and real-time bus location data, *Urban Studies*, 58:3140–3156 (2021).

[2] T. Zhang, Y. Li, H. Yang, C. Cui, J. Li, Q. Qiao, Identifying primary public transit corridors using multi-source big transit data, *International Journal of Geographical Information Science*, 34:1137–1161 (2020).

[3] L. Liu, A. Porr, H. J. Miller, Realizable accessibility: Evaluating the reliability of public transit accessibility using high-resolution real-time data, *Journal of Geographical Systems*, 25:429–451 (2023).

[4] Y. Park, J. Mount, L. Liu, N. Xiao, H. J. Miller, Assessing public transit performance using real-time data: Spatiotemporal patterns of bus operation delays in Columbus, Ohio, USA, *International Journal of Geographical Information Science*, 34:367–392 (2020).

[5] L. Huang, H. Huang, Y. Wang, Resilience Analysis of Traffic Network under Emergencies: A Case Study of Bus Transit Network, *Applied Sciences*, 13:8835 (2023).

[6] X. Li, Y. Gao, H. Zhang, Y. Liao, Passenger Travel Behavior in Public Transport Corridor After the Operation of Urban Rail Transit: A Random Forest Algorithm Approach, *IEEE Access*, 8:211303–211314 (2020).

[7] B. Cottreau, A. Adraoui, O. Manout, L. Bouzouina, Spatio-temporal patterns of the impact of COVID-19 on public transit: An exploratory analysis from Lyon, France, *Regional Science Policy & Practice*, 15:1702–1721 (2023).

[8] D. Yao, L. Xu, J. Li, Evaluating the Performance of Public Transit Systems: A Case Study of Eleven Cities in China, *Sustainability*, 11:3555 (2019).

[9] M. Saki, M. Abolhasan, J. Lipman, A Novel Approach for Big Data Classification and Transportation in Rail Networks, *IEEE Transactions on Intelligent Transportation Systems*, 21:1239–1249 (2020).

[10] S. Liabø, Modeling Passenger Count Data Based on Automatic Counting, Master's thesis, NTNU (2023).