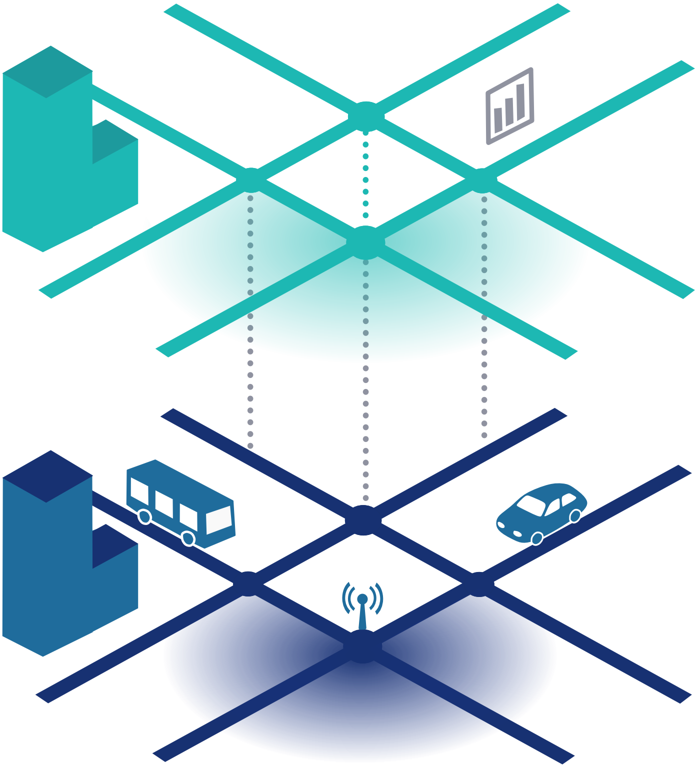
Licentiate Thesis in Transport Science

Digital Twin for Urban Transportation

Architecture, Technology, Modeling and Applications in Stockholm

JONAS JOSTMANN

KTH ROYAL INSTITUTE OF TECHNOLOGY



Digital Twin for Urban Transportation

Architecture, Technology, Modeling and Applications in Stockholm

JONAS JOSTMANN

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Licentiate of Engineering on Monday the 8th of June 2026, at 2:00 p.m. in W25, Teknikringen 78A, Stockholm.

Licentiate Thesis in Transport Science
KTH Royal Institute of Technology
Stockholm, Sweden 2026

Cover page photo: Jonas Jostmann

ISBN 978-91-8106-604-3
TRITA-ABE-DLT-2616

Printed by: Universitetsservice US-AB, Sweden 2026

Abstract

This licentiate thesis develops and prototypically demonstrates Digital Twin (DT) architectures for urban transportation with a focus on road-traffic emissions, Public Transportation (PT) and data-driven path flow estimation to support DT simulation workflows. To advance current transport DTs from isolated sensor dashboards and disconnected simulations, the thesis addresses three research objectives: 1) to formulate a transferable DT architecture for network-level road-traffic emission nowcasting, forecasting, and retrospective analysis, 2) to design an open, automated, and extendable DT development pipeline for PT, and 3) to investigate the robustness and transferability of Partial Least Squares Regression (PLSR) for reconstructing path flows from link flow observations.

Paper I develops and demonstrates an emission-oriented DT framework that integrates camera-based sensing, demand estimation, hybrid traffic simulation, and interactive 3D visualization. Traffic cameras are processed using a computer-vision pipeline that detects and classifies vehicles and extracts speed and acceleration used to nowcast emissions at camera locations. To further estimate network-level emissions, the camera data is used to estimate dynamic OD demand as input for microscopic simulation. A Unity-based 3D platform integrates sensor and simulation output using MQTT-based data streaming. This platform enables joint exploration of near-real-time emissions at sensor locations and simulated emission estimates in the surrounding network. A case study in Kista, Stockholm, illustrates the framework's ability to support both emission nowcasting and scenario analysis, for example by assessing changes in network-level emissions under reduced parking availability.

Since DTs integrate multiple complex components, their development is often resource-intensive and time-consuming. To lower entry barriers Paper II proposes an open, automated DT development pipeline for PT that relies on open data, open standards, and open source software. PT operations are represented using GTFS data, where its static component serves as input for microscopic traffic simulation to enable joint simulation of PT vehicles and private traffic interactions. The GTFS real-time feeds enable both monitoring of current PT operations through low-latency visualization and retrospective analysis of events using a database storing historic observations. The link- and vehicle-level traffic data is displayed together with automatically derived OpenStreetMap building models in a Cesium-based web platform. This interactive visualization allows users to switch between nowcasting, scenario-based forecasting, and playback of historical operations within a 3D spatial context facilitating informed decision making. A case study in Kista, Stockholm demonstrates the pipeline's technical feasibility by showcasing real-time PT operations and simulation-based scenarios visualized in the 3D interactive DT platform.

Paper III formulates and evaluates a PLSR-based path flow estimator as a data-driven alternative to conventional OD matrix estimation (ODME) within DT simulation workflows. Estimating a larger number of OD pairs from a smaller number of link counts, as well as collinearities in the observations render the path flow estimation problem ill-posed. PLSR learns a low-dimensional latent representation that maximizes the covariance between observed link flows and path flows, providing implicit regularization for the ill-posed inverse problem. While the method was used for similar problem settings, it has not yet been used for path flow estimation. Thus, this study evaluates its suitability by assessing its stability and transferability using a synthetic test network and controlled data-generating processes that reflect practically relevant OD and route choice structures. The experiments indicate that PLSR achieves the lowest reconstruction errors when variability in path flows is dominated by OD demand fluctuations. Increased path choice randomness, however, is reducing recoverability at first, but the performance stabilizes once strong path competition regimes are entered. The experiments further indicate that PLSR transfers reliably when fitting and deployment share the same correlation regimes, but performance deteriorates near regime boundaries where OD-driven correlations give way to path choice competition-based correlations. These findings suggest that PLSR can serve as a fast, data-driven path flow estimator in DT contexts. Though, to ensure reliable estimates over time it requires continuous monitoring of the underlying correlation regimes to detect shifts and retrain the model when needed.

Keywords: Digital twin, Urban transportation, Traffic emission, Public transport, Path flow estimation

Sammanfattning

Denna licentiatuppsats utvecklar och prototypiskt demonstrerar Digital Twin (DT)-arkitekturer för urbana transportsystem med fokus på vägtrafikutsläpp, kollektivtrafik och datadriven skattning av vägflöden för att stödja simuleringsbaserade DT-arbetsflöden. För att föra dagens transportrelaterade DT-lösningar vidare från isolerade sensordashboards och fristående simuleringar behandlar avhandlingen tre forskningsmål: 1) att formulera en överförbar DT-arkitektur för nätverksomfattande skattning i realtid, prognoser och retrospektiv analys av vägtrafikutsläpp, 2) att utforma en öppen, automatiserad och utbyggbar utvecklingspipeline för DT inom kollektivtrafik, samt 3) att undersöka robustheten och överförbarheten hos Partial Least Squares Regression (PLSR) för rekonstruktion av vägflöden från länktrafikobservationer.

Artikel I utvecklar och demonstrerar ett utsläppsorienterat DT-ramverk som integrerar kamerabaserad datainsamling, efterfrågeskattning, hybrid trafiksimulering och interaktiv 3D-visualisering. Trafikkameror behandlas med en datorseendepipeline som detekterar och klassificerar fordon samt extraherar hastighet och acceleration, vilka används för att skatta utsläpp i realtid vid kamerapositionerna. För att vidare uppskatta utsläpp på nätverksnivå används kameradata för att skatta dynamisk OD-efterfrågan som sedan används som indata till mikrosimulering. En Unity-baserad 3D-plattform integrerar sensor- och simuleringsdata via MQTT-baserad dataströmning. Plattformen möjliggör gemensam analys av nästan realtida utsläpp vid sensorplatser och simulerade utsläppsskattningar i det omgivande nätverket. En fallstudie i Kista, Stockholm, visar ramverkets förmåga att stödja både realtidsskattning av utsläpp och scenarioanalys, exempelvis genom att bedöma förändringar i nätverksomfattande utsläpp vid minskad parkeringstillgång.

Eftersom DT-system integrerar flera komplexa komponenter är deras utveckling ofta resurskrävande och tidsintensiv. För att sänka trösklarna föreslår Artikel II en öppen och automatiserad DT-utvecklingspipeline för kollektivtrafik som bygger på öppna data, öppna standarder och programvara med öppen källkod. Kollektivtrafikens drift representeras med GTFS-data, där den statiska delen används som indata till mikroskopisk trafiksimulering för att möjliggöra gemensam simulering av kollektivtrafikfordon och interaktioner med privat trafik. GTFS realtidsflöden möjliggör både övervakning av aktuell kollektivtrafikdrift genom visualisering med låg latens och retrospektiv analys av händelser genom en databas som lagrar historiska observationer. Trafikdata på länk- och fordonsnivå visualiseras tillsammans med automatiskt genererade OSM-baserade byggnadsmodeller i en Cesium-baserad webbplattform. Denna interaktiva visualisering gör det möjligt för användare att växla mellan realtidsövervakning, scenariobaserad prognostisering och uppspelning av historisk drift i en rumslig 3D-kontext som underlättar välgrundat beslutsfattande. En fallstudie i Kista, Stockholm, demonstrerar pipeline-lösningens tekniska genomförbarhet genom att visa realtida kollektivtrafikdrift och simuleringsbaserade scenarier i den interaktiva 3D-baserade DT-plattformen.

Artikel III formulerar och utvärderar en PLSR-baserad skattare av vägflöden som ett datadrivet alternativ till konventionell skattning av OD-matriser inom DT-baserade simuleringsarbetsflöden. Att skatta ett större antal OD-par utifrån ett mindre antal länkräkningar, i kombination med kollinearitet i observationerna, gör problemet att skatta vägflöden ill-posed. PLSR lär sig en lågdimensionell latent representation som maximerar kovariansen mellan observerade länkflöden och vägflöden, vilket ger en implicit regularisering av det ill-posed inversa problemet. Även om metoden har använts i liknande problemsammanhang har den ännu inte tillämpats för skattning av vägflöden. Studien utvärderar därför dess lämplighet genom att analysera dess stabilitet och överförbarhet med hjälp av ett syntetiskt testnätverk och kontrollerade datagenererande processer som speglar praktiskt relevanta OD- och vägvalsstrukturer. Experimenten visar att PLSR uppnår lägst rekonstruktionsfel när variationen i vägflöden domineras av variation i OD-efterfrågan. Ökad slumpmässighet i vägval minskar däremot initialt rekonstruerbarheten, men prestandan stabiliseras när systemen går in i regimer med stark konkurrens mellan vägar. Experimenten indikerar vidare att PLSR är pålitligt överförbar när träning och tillämpning delar samma korrelationsregimer, men att prestandan försämras nära regimgränser där OD-drivna korrelationer övergår i korrelationer som orsakas av konkurrens mellan vägval. Resultaten tyder på att PLSR kan fungera som en snabb och datadriven skattare av vägflöden i DT-sammanhang. För att säkerställa tillförlitliga skattningar över tid krävs dock kontinuerlig övervakning av de underliggande korrelationsregimerna för att upptäcka skiften och träna om modellen vid behov.

List of Papers

Papers included in the thesis

I **Jostmann, J.**, Hu, S., Gustafsson, A., Santi, P., Ratti, C., & Ma, Z. (2026). “Digital twin for urban car traffic emission: A case study in Kista, Stockholm”. Journal of Intelligent and Connected Vehicles, `https://doi.org/10.26599/JICV.2026.9210079`. Presented at the 104th Transportation Research Board (TRB) Annual Meeting (2025).

II **Jostmann, J.**, Mo, T., Wang, H., & Ma, Z. (2026). ”Open source Digital Twin development pipeline for Public Transport: A case study in Stockholm”. Journal of Public Transportation, 28, 100149. `https://doi.org/10.1016/j.jpubtr.2026.100149`

III **Jostmann, J.**, Flötteröd, G., & Ma, Z. (2026). ”Partial Least Squares Regression for Ill-Posed Path Flow Estimation from Partial Link Traffic Counts”. Submitted to IEEE Conference on Intelligent Transport Systems (ITSC) 2026.

Related publications not included in the thesis

a) Wiemers, M., Jordan Jones, J., Cicic, M., **Jostmann, J.**, Ma, Z., & Delle Monache, M. L. ”Digital Twin-Enabled Reinforcement Learning for Fault-Resilient Urban Traffic Signal Control”. Accepted for presentation at the 2026 European Control Conference (ECC).

b) **Jostmann, J.**, Flötteröd, G., & Ma, Z. (2024). ”Machine-Learning informed simulation-based dynamic traffic assignment with SUMO”. SUMO User Conference 2024, Berlin, Germany, 2024.

Declaration of Contribution

The ideas for the included papers were developed in collaboration with my co-authors. For all included papers, I was responsible for the coordination between the authors and was the primary contributor to the paper writing, the experimental set-up, the analysis, the interpretation, and the visualization of the results.

For Paper I, the development of the included components was distributed among my co-authors. I led the development of the hybrid-simulation framework, Songhua Hu led the computer vision detection pipeline, and Anton Gustafsson led the development of the visualization platform. Zhenliang Ma, Paolo Santi, and Carlo Ratti supported the conceptualization of the study through supervision and by reviewing the final manuscript.

For Paper II, the implementation work was shared between my co-author Tong Mo and me. Haoyan Wang supported the methodology development and review of the paper. Zhenliang Ma provided critical input for the conceptualization, methodology development, and review of the paper writing.

For Paper III, Gunnar Flötteröd and Zhenliang Ma contributed through conceptualization, interpretation of results, supervision, and reviewing the paper writing.

Declaration of AI Usage

During the preparation of this thesis, I used Large Language Models (LLMs) to improve the language of the article. After using these tools, I reviewed and edited the content as needed and take full responsibility for the content.

Acknowledgement

First, I would like to express my gratitude to my supervisors for their guidance throughout my licentiate. Thank you, Zhenliang, for your continuous support and many insightful discussions we shared. Your advice was always helpful and allowed me to improve in my academic journey. Thank you, Gunnar, for your thoughtful and well-informed feedback. Your attention to detail was especially helpful in structuring my thoughts and guiding my research. I would also like to thank Erik for his support, particularly during the search for funding for the continuation of my PhD.

Thank you also, Wilco, for acting as the advance reviewer for my thesis and providing helpful comments and guidance on how to further improve it.

At this point I would also like to thank the funders of my research. This includes the Stockholm Seseable Lab, Digital Futures and Trafikverket.

Apart from the academic support, I want to thank my colleagues and friends in the division! Thank you for all the nice lunches, fikas, gym sessions, basketball games and other activities we shared together! I especially enjoyed travelling with you to conferences and having atomic pauses!

I would also like to give a special thank you to my friends for sticking by me over the last years. Thank you for your support and understanding when things became busier and I had less time to spare. I am glad that we were able to catch up via podcast-lengthy voice memos, Zoom calls, camping trips and shared vacations.

With that, I would also like to thank my family for always believing in me and supporting me from the very beginning. Even though I know you would like me to live closer to home, you have always backed my decision to move to Sweden. I am so glad to have such a wonderful and loving family!

Finally, I want to thank Anna. I am very happy and grateful that we can share this journey together. You were always there for me during the wonderful moments we had, but you were also the one who probably had to listen most to me complaining when something did not work as I hoped. Thank you for your love, patience, and understanding over the years!

Contents

List of Papers **v**

Declaration of Contribution **vii**

Declaration of AI Usage **ix**

Acknowledgement **xi**

Contents **1**

1 Introduction **3**
- 1.1 Digital Twin Definition . . . 3
- 1.2 Digital Twins in Transportation . . . 5
- 1.3 Digital Twins for Road Traffic Emissions . . . 6
- 1.4 Digital Twins for Public Transportation . . . 7
- 1.5 Transport Digital Twin Modeling - Path Flow Estimation . . . 9
- 1.6 Aim of the Thesis . . . 10
- 1.7 Thesis outline . . . 11

2 Research Objectives **13**

3 Research Methodology **15**
- 3.1 Data and Sensing . . . 17
- 3.2 Traffic Demand Estimation . . . 19
- 3.3 Simulation . . . 21
- 3.4 System Integration and Visualization . . . 22
- 3.5 Case Study . . . 25

4 Thesis Contributions **29**
- 4.1 Paper I . . . 29
- 4.2 Paper II . . . 30
- 4.3 Paper III . . . 31

5 Summary, Discussion and Future Directions **33**

5.1 Summary 33
5.2 Discussion 34
5.3 Future Directions 36

References **39**

Chapter 1

Introduction

Nowadays, transport infrastructure and services are increasingly equipped with sensing and communication technologies that passively collect data from a variety of sources, such as traffic cameras, loop detectors, and GPS-equipped vehicles. Such data enables efficient monitoring of current conditions, supports the analysis of past events, and provides a basis for the evaluation of future scenarios. To make these data streams actionable for operations and planning, they must be embedded in an architecture that 1) reliably acquires, harmonizes, and stores heterogeneous observations, 2) integrates them with models and simulations to extract insights, and 3) presents results in a form that enables informed decision-making. The resulting analysis and decisions finally influence the physical transport system either automatically, e.g., via control infrastructure and information systems, or manually through operational or policy interventions.

Digital Twins (DTs) have been identified as a promising approach to address these requirements by coupling a physical system with a continuously updated digital counterpart.

1.1 Digital Twin Definition

DTs were first introduced with respect to product lifecycle management in the manufacturing industry by Grieves in 2003 [1]. The term originally described products in a physical space, their representation in a virtual space, and a connection between both spaces characterized by the exchange of data and information. Later, the concept was adopted by NASA in 2012 to address the limitations of conventional air and spacecraft certification, development, and monitoring practices, which often relied on heuristics and empirical studies before [2]. Based on their definition, a DT integrates physics-based simulation driven by sensor data and historic observations to forecast system health.

Other definitions support this general DT structure consisting of a physical space coupled with a digital counterpart [3]. A representation of this basic refer-

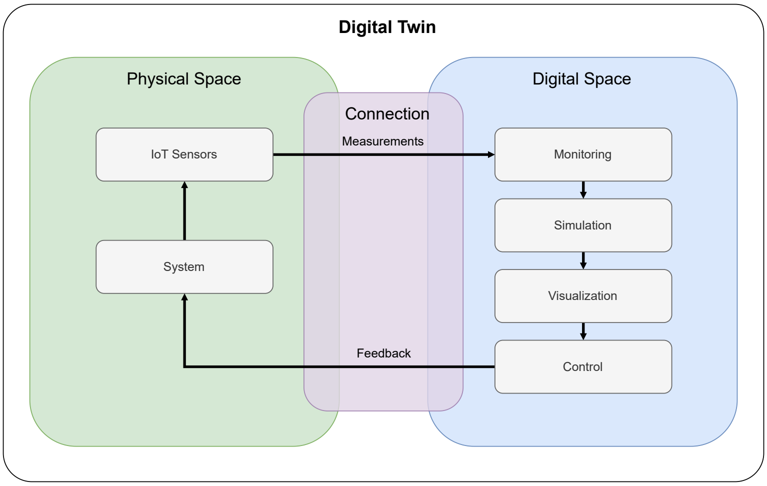


Figure 1.1: High-level Digital Twin Architecture.

ence architecture is shown in Figure 1.1. The coupling is typically realized using IoT sensors continuously observing the physical system's state. The digital copy offers centralized monitoring of various components which enables applications such as preventive or predictive maintenance. Assuming the digital counterpart and the real system behave in the same way, the digital copy can be influenced to explore effects of changes in a safe and efficient way. Through simulation alternative conditions and settings can be explored to optimize production processes or the product itself [4].

The gained insights can then either be used by decision-makers or applied in an automated way to control the real system. These two ways describing the connection from the digital part back to the physical space, distinguish the maturity levels of a DT. While the former, having a manual feedback loop, describes a Digital Shadow, the latter, offering automated control, describes a DT [5]. Even though this definition refers to different degrees of maturity, the term DT is mainly used in literature as a collective term that often also includes Digital Shadows.

As the DT concept gained more attention in the manufacturing industry, it was also adopted in other areas such as healthcare and agriculture. In healthcare, DTs are often defined as virtual counterparts of patients or organs that are kept synchronized with the physical person via heterogeneous data streams such as wearables and electronic medical records, enabling analysis and simulation for personalized and precision medicine [6]. In agriculture, DTs are described as virtual representations of farm products, processes, or environments, typically built on sensing and simulation [7]. Building DTs in these fields is often complex because both require the modeling of living subjects. Especially when dealing with humans, further challenges such as privacy and ethical concerns emerge.

1.2 Digital Twins in Transportation

Beyond the aforementioned application areas, DTs are increasingly used in urban settings, particularly for transportation systems. Similarly to manufacturing, a city and its transport network can be represented by a continuously updated virtual counterpart, enabling operators and planners to monitor current conditions and test alternative policies or operational strategies through simulation before implementing interventions in the real world [8]. In transportation, the DT's physical space comprises elements such as the transport network, vehicles, pedestrians, control infrastructure (e.g., traffic signals), and the built environment, which are monitored via IoT sensors (e.g., loop detectors, traffic cameras, or GPS) [9]. In the digital space, data is centralized and processed to either collect information and analyze past events to develop a better understanding of the underlying system or to monitor existing real-time data directly. Furthermore, the data can be used as input for simulation models that allow exploration of how the transport system evolves, e.g., under different policy interventions or network changes. Even though the simulation component in a DT is vital for exploring alternative scenarios, through real-time data integration, transport DTs go beyond pure simulation. By that, transportation DTs are able to connect the past (retrospective analysis through data analysis), present (nowcasting through real-time data monitoring), and future (forecasting through simulation) within one integrated workflow.

Beyond the underlying data and modeling components, many DTs rely on interactive 3D visual interfaces to communicate system states and model outputs to stakeholders. Visualizing sensor observations and simulation results within a shared spatial context can improve interpretability and contextualization, especially when multiple data sources and levels of aggregation must be considered simultaneously. In addition, interactive visualization can facilitate communication across stakeholder groups, e.g., between analysts, operators, and planners, by making assumptions, scenarios, and results more transparent and easier to discuss [10].

Existing research identified several functional dimensions in which DTs innovate current monitoring, planning and control practices. These dimensions can be broadly classified into 1) safety-, 2) environmental- and 3) mobility-related, hence directly contributing to the United Nations Sustainable Development Goals (SDG), particularly SDG 3 (Good Health and Well-Being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action) [11].

DTs for transportation can increase safety in urban areas by improving the situational awareness of drivers, cyclists, and pedestrians. For instance, by predicting the behavior and intentions of other traffic participants [12] and providing direct feedback to drivers through Advanced Driver Assistance Systems (ADAS) [13]. Furthermore, increasing connectivity and low latency communication enabled through 5G allows for fast incident detection and increases the robustness of the transport system [14].

Existing research on the mobility dimension of transportation DTs focuses on three main areas: Intelligent Transportation Systems (ITS), intelligent vehicles, and individual mobility. DT-enabled ITSs allow real-time monitoring and adaptive control strategies, such as traffic signal optimization, to reduce congestion and improve traffic flow [15, 16]. At the vehicle level, DTs support cooperative and intelligent driving by enabling trajectory prediction, coordination, and optimized energy management, particularly for electric vehicles [17, 18]. Moreover, DTs contribute to individual mobility planning by generating synthetic travel demand from large-scale sensor data [19].

DTs further provide a valuable tool to reduce the impact of traffic emissions in urban areas. For instance, by using their real-time monitoring and simulation capabilities, real-world control such as traffic signals can be optimized towards lower fuel consumption and hence lower emissions [15].

Existing research on DTs in transportation highlights its potential, but it also reveals limitations in how DT capabilities are utilized for concrete planning and operational tasks. In particular, environmental DT studies often do not provide an integrated architecture that couples real-time sensing with network-wide emission monitoring and scenario-based forecasting within an integrated framework. At the same time, research on DTs for Public Transportation (PT) remains comparatively underdeveloped, and many existing implementations address individual components (e.g., data integration, simulation, or visualization) rather than offering a unified, extensible platform that supports monitoring, retrospective analysis, and forecasting in an integrated way.

Motivated by these limitations, this thesis focuses on emission- and PT-oriented DTs. By that it contributes to advancing both the system architecture and the practical integration of data, simulation, and visualization required for decision support. The following two sections, therefore, review the state of the art and the remaining limitations for both focus areas to identify the need for the solutions proposed in the included papers.

1.3 Digital Twins for Road Traffic Emissions

To enable efficient emission monitoring and control, a tool is required that allows for the observation of current emission levels and provides high-resolution network-wide estimates. Furthermore, to explore the effects of alternative scenarios and policy interventions, simulation-enabled forecasting capabilities are necessary.

Existing studies developing DTs with an environmental focus lack a unified integration of all these components. Many solely focus on the collection and visualization of sensor data, limiting the observation of emissions to areas equipped with sensors. The authors of [20], for instance, connect to existing traffic and emission-related sensing infrastructure in Sydney, Australia, and automatically detect correlations in the observed data for decision support. Their approach is, however, limited to reactive monitoring, disregarding alternative conditions and network-level estimates. The authors of [21] approach the problem of missing network-level emission estimates by first calibrating more cost-effective sensors to increase sensor coverage and subsequently use dispersion models to generate emission estimates for regions without sensor coverage. While their approach is able to effectively estimate network-level emissions, it lacks the ability to evaluate alternative scenarios.

Other research takes alternative scenario evaluation into account, but only from a built-environment or land-use perspective. In [22], the authors present a DT that allows estimating changes in emission dispersion based on building construction and land-use changes. In their work, however, they assume constant traffic patterns. Similarly, the authors of [23] propose a DT that serves as a decision support tool for assessing land-use plans based on macroscopic emission maps, completely disregarding detailed traffic and emission measurements.

Few studies take changing traffic patterns and alternative scenarios into account. These studies often focus only on limited study areas and simplified traffic scenarios. In [24], the authors, for instance, develop a DT that allows for the analysis of changing traffic patterns and the resulting emission changes of connected vehicles on highway sections. [15] optimizes, with the help of a DT, signal control plans to reduce CO_2 emissions. The DT is, however, limited to individual intersections.

The investigated literature remains fragmented and does not provide a traffic emission oriented DT that jointly supports real-time emission monitoring, network-level emission estimation, and the evaluation of alternative scenarios within a single framework. Concretely, the following four research gaps have been identified:

- **Gap 1.1:** Lack of an integrated DT architecture for urban traffic emission combining monitoring, estimation, and forecasting.
- **Gap 1.2:** Lack of methods for high-resolution, network-wide urban emission analysis in areas without sensor coverage.
- **Gap 1.3:** Lack of network-level scenario-based emission analysis under changing urban traffic conditions.

1.4 Digital Twins for Public Transportation

DTs for PT have the potential to support both planning and operations by integrating heterogeneous static and real-time data sources with simulation in a unified environment. Despite this potential, research on DTs in the PT domain remains in an early stage, and existing implementations often realize only subsets of the aforementioned capabilities in isolation. The prevailing literature can be broadly categorized by the capabilities it focuses on: 1) IoT- and technical DT architecture-, 2) simulation-, and 3) visualization-focused approaches.

Existing work focusing on the DT's technical architecture and the integration of IoT sensors with other data sources often emphasizes interoperability, scalability, and the integration of real-time data [25, 26]. The developed DTs, however, lack scenario exploration or visualization capabilities, which are essential when applying the DT for planning purposes. Similar work adds predictive capabilities to pure real-time monitoring, allowing joint supervision and control using PT real-time data (e.g., GTFS and incident data) [27, 28]. While these approaches demonstrate the value of real-time monitoring and short-term prediction, they are often designed for short-term operational time-horizons and are not suited for planning tasks that often require mid- and long-term, simulation-driven what-if analyzes. Existing studies on simulation-oriented DTs, on the other hand, enable the exploration of alternative scenarios, but offer limited integration of real-time data and often present results via aggregated KPIs rather than in interactive visualizations [29–31].

Previous research has already identified interactive spatial 3D visualization as beneficial for improving contextualization when analyzing complex spatial dynamics and supporting communication with stakeholders [32]. However, in DTs for PT 3D visualization shows low utilization, with many PT DT implementations relying on 2D KPI dashboards. However, implementations that use 3D visualizations employ highly detailed 3D modeling approaches and are often based on custom LIDAR scans, which limit transferability between cities and require a high initial development effort [31, 33].

The existing literature indicates a need for a PT DT that combines real-time data integration, simulation for retrospective and scenario-based analysis, and interactive visualization within an extensible framework. From that the following three concrete research gaps can be identified:

- **Gap 2.1:** Lack of an integrated PT DT framework combining real-time data, historical data, simulation, and visualization.

- **Gap 2.2:** Lack of a transferable interactive spatial visualization approach for PT DTs.

- **Gap 2.3:** Lack of an open, automated, and extensible development pipeline for PT DTs.

1.5 Transport Digital Twin Modeling - Path Flow Estimation

Both DTs for PT and urban road-traffic emissions rely on traffic simulation to evaluate policy interventions and alternative future scenarios. In order to achieve meaningful results, such simulations require well-calibrated traffic demand data as input. This demand is commonly represented as Origin Destination (OD) demand matrices, which specify the number of vehicles traveling from each origin to each destination. Depending on the application, OD demand can be represented as static (e.g., a daily matrix) or dynamic, i.e., disaggregated into time-of-day intervals (e.g., 15-minute slices).

In practice, the true OD demand matrix is unknown and must be estimated from partial observations such as roadside detectors or Floating Car Data (FCD). Model-based methods for OD estimation typically formulate the problem as a bi-level optimization task [34]. Here, the upper-level problem minimizes the difference between modeled and observed traffic data (e.g., link flows). The lower-level assigns travel demand from the given OD matrix to the network by solving a traffic assignment problem, most commonly formulated as an User Equilibrium (UE). Especially, the lower-level is time consuming as it requires iterative execution of traffic assignment and network loading.

An alternative to estimating aggregated OD demand is estimating path flows. Since path flows already incorporate implicit route choice information, the costly assignment step can be avoided. In the literature, path flow estimation is commonly formulated as an optimization problem under Stochastic User Equilibrium (SUE), with multinomial logit route choice and link cost functions defining the mapping between path and link flows [35, 36]. For these methods, the estimation performance depends strongly on the validity of the assumed behavioral model and cost structure. Data-driven alternatives do not rely on such assumptions. Instead, these methods estimate path flows by fitting models that map observed measurements (e.g., link counts or floating-car data) to path flow vectors. Models typically use machine learning methods, e.g., Convolutional Neural Networks (CNN)-based designs [37]. Once trained, they replace iterative optimization with a single forward pass, making their execution computationally efficient. This aligns well with urban DTs applications requiring online estimations.

Both families of methods ultimately depend on link flow measurements. In real networks, however, observations are typically available only for a subset of links. Since the number of paths is often larger than the number of observed links, the resulting estimation problem is commonly under-determined [38]. Moreover, link observations may be strongly collinear, which can make the path flow reconstruction problem ill-posed even in cases where the number of link observations equals the number of OD variables [39]. Regularization is often performed in order to obtain reliable estimates nonetheless. This is done, for example, through historic priors or by dimensionality reduction methods such as Principal Component Analysis (PCA) [40]. However, many of the currently employed techniques have important shortcomings. Prior-based approaches can be highly sensitive to the reliability of the historic priors [41], and unsupervised reduction may remove structure that links predictors and responses [42].

An alternative to the unsupervised dimensionality reduction techniques is Partial Least Squares Regression (PLSR). PLSR fits a low-dimensional latent representation that maximizes the covariance between predictors and responses and establishes a linear mapping between both in this latent space [42]. Through this projection into the latent space, the method implicitly realizes regularization. To the best of the authors' knowledge, PLSR has not yet been used for the path flow estimation problem. Its suitability for this problem setting, therefore, remains a research gap.

In summary, there exist the following two research gaps:

- **Gap 3.1:** Limited robustness of existing path flow estimation methods under partial observability and ill-posedness.

- **Gap 3.2:** Lack of evidence on the suitability of PLSR for path flow estimation problem in transportation applications.

1.6 Aim of the Thesis

To address the research gaps identified in the literature, this thesis aims to develop and demonstrate DT architectures for urban transport applications, with a focus on 1) road-traffic emission and 2) PT monitoring and analysis, as well as path flow estimation for transport DT simulation models.

Specifically, the thesis targets the development of DT architectures that integrate heterogeneous data sources (e.g., real-time IoT data, traffic demand, and data representing the built environment), simulation of current and alternative scenarios (Gaps 1.1-1.3, 2.1), as well as interactive 3D visualizations to explore results and facilitate decision making (Gaps 2.2). For both architectures we target transferable designs that allow other cities to set-up DT systems with limited effort (Gap 2.3).

Simulation represents a core DT component, which relies on calibrated demand. By investigating PLSR as a method for path flow reconstruction from link flow observations this thesis addresses Gap 3.2. The method is particularly relevant in DT settings because it enables efficient, data-driven estimation while addressing the ill-posed nature of the inverse problem (Gap 3.1). Concretely, the thesis aims to evaluate the method's stability and transferability under varying correlation structures as a step towards robust OD calibration within DT workflows.

1.7 Thesis outline

The remainder of this thesis is structured as follows: Chapter 2 summarizes the identified research objectives. Next, Chapter 3 develops the conceptual and methodological framework underlying the thesis contributions. Chapter 4 then provides a summary of the three appended papers, clarifying the specific research objectives addressed by each paper and presenting their main findings. Finally, Chapter 5 gives a summary of the thesis, discusses its limitations, and outlines directions for future research.

Chapter 2

Research Objectives

Based on the research gaps identified in the literature, this thesis is guided by the following overarching research question: How can urban transportation DTs be designed to integrate real-time monitoring, retrospective analysis, and simulation-based forecasting of what-if scenarios within a unified interactive platform? To address this question, the thesis defines the following four research objectives.

RO1 - Identify the common building blocks of urban transportation DTs and formulate a transferable architecture that integrates real-time data, simulation, and interactive visualization for monitoring, retrospective analysis, and forecasting.

The first objective refers to the architectural gaps identified in the literature (Gaps 1.1 and 2.1). It primarily aims to identify the key building blocks required for urban transportation DTs, including real-world data acquisition, harmonization, simulation, demand estimation, and interactive visualization, and to clarify how these components can be integrated to form a transferable DT architecture for urban transportation. This objective serves as an overarching basis for the two application-oriented objectives on road-traffic emissions and PT specified in the following.

RO2 - Implement and demonstrate a DT architecture for network-level road traffic emission nowcasting, scenario-based forecasting, and retrospective analysis.

The second objective targets the development of a DT platform for road-traffic emission nowcasting and forecasting (Gap 1.1). Concretely, the objective is to establish an integrated architecture that combines heterogeneous IoT observations with simulation to generate network-level emission estimates at high spatial and temporal resolution (Gap 1.2). Further, the objective is to demonstrate how traffic camera data can be used as the primary input source for jointly supporting emission monitoring and OD Matrix Estimation (ODME). The study aims to demonstrate through a case study the DT's technical capabilities, including the ability to provide near-real-time emission estimates and simulation-based analyzes for alternative planning scenarios (Gap 1.3).

RO3 - Implement and demonstrate an open, automated, extensible DT development pipeline for PT DTs.

The third objective relates to the gaps identified for PT-oriented DTs, particularly the lack of transferability and reproducibility (Gaps 2.2 and 2.3). It targets the implementation and demonstration of a DT development pipeline for PT that combines real-world data processing, simulation, and 3D visualization within an open and extensible architecture (Gap 2.1). This objective is to reduce implementation barriers and improve transferability to other urban mobility applications. The objective, further, is to demonstrate its technical capabilities through a prototypical implementation to showcase functionalities for real-time monitoring, scenario analysis, and retrospective exploration.

RO4 - Implement PLSR as a data-driven path flow estimator and evaluate its robustness and transferability under varying data-generating processes.

This study targets the formulation of demand estimation as an ill-posed path flow reconstruction problem under partially observed link counts and investigates PLSR as a data-driven estimator (Gap 3.1). The objective is to evaluate whether and under which conditions PLSR can serve as a reliable path flow estimator for DT simulation workflows. Specifically, the study aims to evaluate the method's stability under stationary conditions and its transferability when the underlying demand and path choice structures differ between model fitting and deployment (Gap 3.2). To this end, PLSR is assessed on a synthetic network under multiple data-generating processes designed to represent practically relevant demand and route choice patterns.

Chapter 3

Research Methodology

This chapter outlines the methodological approach used to address the thesis' research objectives. The work is guided by the Design Science Research (DSR) methodological framework [43] for RO1 and RO2, as these objectives target the development of technical artifacts that address concrete needs in society and practice. Concretely, a DT for emission monitoring and control (Paper I) and a DT for PT planning and operations (Paper II). DSR involves three cycles depicted in Figure 3.1. The first cycle, the relevance cycle, involves motivating the need for the artifact to be developed. In this thesis, the relevance cycle was already addressed as part of Chapters 1 and 2, which establish the problem relevance and specify the aforementioned objectives. This chapter continues with the remaining rigor and design cycles. The rigor cycle includes motivating the selection of established methods and tools tailored to the previously identified needs to integrate them and finally build the targeted artifact in the design cycle.

The design cycle involves evaluating the developed artifact to demonstrate

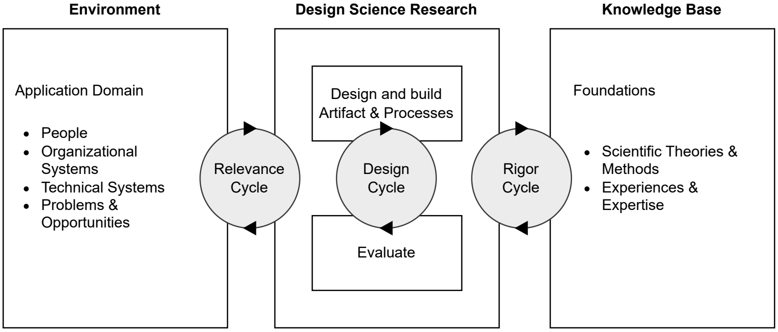


Figure 3.1: Design Science Research Methodological Framework (Figure adopted from [43]).

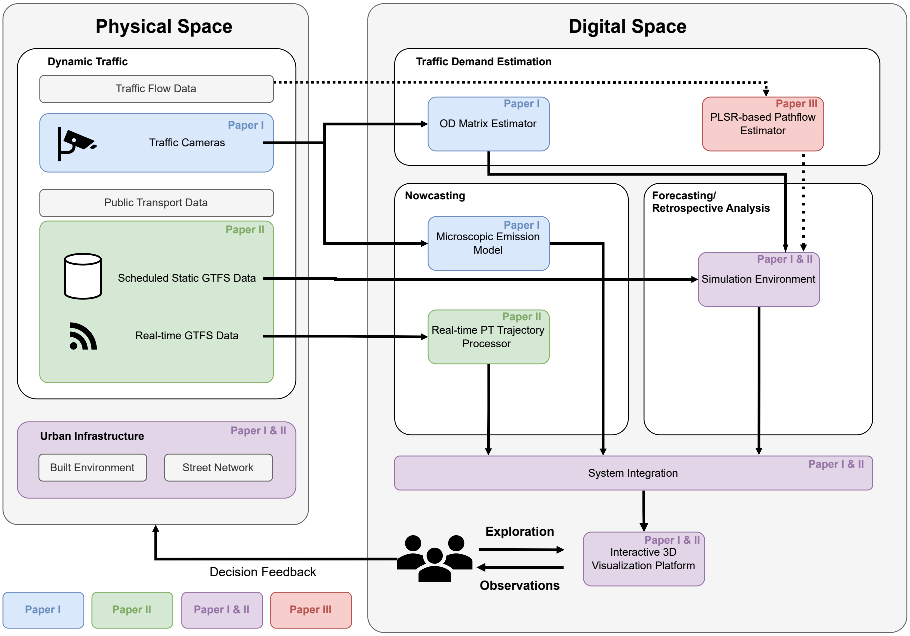


Figure 3.2: High-level DT architecture connecting the individual components of Paper I, II and III.

that it fulfills the specified requirements. In this thesis, this evaluation of the emission and PT DT is performed through a demonstration in a case study setting. The purpose of this demonstration is not to provide a fully validated real-world deployment, but to show the technical feasibility and interoperability of the proposed end-to-end pipelines.

Papers I and II primarily follow the DSR design cycle by integrating established methods into a complete DT artifact. Paper III conducts a quantitative evaluation to extend the knowledge base for a key sub-component for RO3. Specifically, it assesses the stability and transferability of PLSR for path flow estimation, which provides demand inputs to simulation-driven DT workflows.

This chapter is organized according to the overall DT architecture. We present the methodology for each DT component sequentially: 1) data and sensing, 2) demand estimation, 3) simulation, 4) visualization and system integration, and 5) case study set-up. Figure 3.2 provides an overview of the overall architecture and indicates which components are addressed in each paper.

3.1 Data and Sensing

A DT relies on real-world data to represent relevant parts of the physical transport system in its digital space. In this thesis, we focus on dynamic traffic and PT operations (vehicle movements, locations, and flows) and the infrastructure (street network and built environment). In the following, we present the individual data sources used in the included studies and motivate their selection. In general, the data sources were selected to ensure interoperability and transferability across cities with a preference for open standards and open data, and compatibility with downstream processing steps such as demand estimation, simulation, and 3D visualization.

Traffic Flow Data

Traffic observations are required both to quantify current conditions (nowcasting) and to calibrate demand inputs for simulation-based forecasting. Such observations are commonly obtained from stationary detectors (e.g., loop detectors), FCD, or camera-based sensing. Loop detectors are widely used and provide robust point measurements of flows and speeds. However, they typically do not provide more detailed data, such as trajectories, acceleration, or fleet composition. FCD, on the other hand, can provide trajectories, but only for a subset of vehicles and with potential sampling bias depending on penetration rates and user groups.

For the emission-oriented DT (Paper I), traffic cameras are therefore selected for traffic sensing, as they allow for the joint extraction of link flows, speed, acceleration, and vehicle class information, which are required inputs for microscopic emission estimation. Specifically, the camera stream is processed with a computer-vision pipeline that detects vehicles and classifies their type. The resulting per-vehicle driving parameters and vehicle classes are used to nowcast emissions at camera locations by using the data as input for microscopic emission models. Despite delivering more detailed data compared to loop detectors, camera-based sensing still fails to provide vehicle trajectories and demand, which limits the sensing set-up to local emission monitoring and control. To achieve network-level emission monitoring a separate OD estimation and simulation step is required, for which the camera data is used as input. The demand estimation and simulation setup is described in the subsequent sections.

Public Transport Data

For the PT-focused DT development pipeline (Paper II), the DT must represent both the planned PT supply (timetables, routes, stops) and the actual operations (real-time vehicle positions and delays) to support retrospective analyses, real-time monitoring, and alternative scenario evaluation. To improve interoperability and transferability between cities, the PT data is represented using the General

Transit Feed Specification (GTFS) [44]. Originally developed by Google and TriMet in 2006, it is widely adopted by PT operators and provides a standardized data format that allows for easier transfer compared to city-specific formats.

GTFS generally consists of two components. The static component describes the scheduled PT supply (e.g., routes, trips, stop locations, and planned departure/arrival times). This component provides the basis for simulation and a baseline for real-world performance comparisons. The real-time component provides live operational data (e.g., service alerts, trip updates, and vehicle positions), enabling nowcasting and the visualization of current vehicle trajectories. In the proposed DT pipeline in Paper II, real-time GTFS feeds are used for nowcasting through direct visualization and are also saved to a database to enable later replay for retrospective analysis of past operations. This allows for diagnosing recurrent issues and comparing observed conditions to simulated scenarios.

Street Network

A uniform network representation is essential for geographically locating dynamic real-world observations, running traffic and PT simulations, and combining their outcomes in a coherent way. While many cities maintain internal GIS road networks, these datasets are often not directly compatible with open-source simulators, can be difficult to access, and may be seldom updated.

This thesis therefore uses OpenStreetMap (OSM) [45] as the network source, which is motivated by its world-wide availability and open licensing. This supports the thesis' objective of transferable DT architectures because the shared network source allows for the reuse of the developed processing and simulation procedures in other cities with limited manual effort. Further, the selection is based on OSM's compatibility with the remaining DT components, especially its interoperability with the simulation stack used in Papers I and II. Most simulation software provides tools that allow for transforming the OSM network into the data format required for the simulation itself.

Built Environment

Beside the street network, the built environment is included in the transport DT architecture to support interpretability and communication of the derived outputs. Integrating 3D buildings in the visualization component of the DT helps stakeholders to relate traffic and PT dynamics with the surroundings and thereby facilitates informed decision making. Besides the visualization and communication perspective, the integration of physical structures provides the basis for future extensions towards applications such as traffic emission dispersion modeling.

Due to slightly different goals targeted in Papers I and II two different approaches to represent the 3D built environment are used. In the emission DT (Paper I), external 3D data is sourced from Google Earth API to provide a highly detailed 3D representation in a Unity-based visualization. In contrast, the

PT DT development pipeline (Paper II) prioritizes automation and open data to facilitate transferability, which motivates extracting building data and their attributes from OSM and converting them into the open CityGML format [46] for standardized representation. This further allows for the easy addition of new buildings or the removal of existing ones to visualize alternative land-use plans.

3.2 Traffic Demand Estimation

Traffic sensors, such as the camera-based detectors used in Paper I, provide only partial observations of network states, as monitoring all links is infeasible in real urban networks. To effectively monitor and control PT and emissions in the DT framework, network-level traffic modeling is required. Therefore, OD demand is commonly estimated based on the available point observations, which are subsequently used as input for simulation to obtain network-level traffic.

This thesis considers two approaches for demand estimation. Paper I follows a classical ODME procedure in which OD demand is calibrated as part of a bi-level optimization problem to reproduce observed link flows. Paper III, in contrast, investigates the applicability of a data-driven approach that directly reconstructs path flows from observed link flows, thereby avoiding the repeated solution of the costly traffic assignment problem that forms the lower-level component of the bi-level optimization framework.

OD Matrix Estimation (ODME)

In Paper I, camera-derived link flows serve as partial observations and are used to calibrate a dynamic OD matrix for the study area. The ODME is formulated as a bi-level optimization problem, where the upper level minimizes the mismatch between simulated and observed link flows by adjusting the OD matrix, while the lower level corresponds to a dynamic traffic assignment problem that involves route switching of individual travelers until converging towards a UE for a given OD demand. The ODME procedure is implemented using DTALite, a mesoscopic simulator designed for fast dynamic traffic assignment and model calibration [47], which makes it suitable for iterative OD estimation compared to embedding a microscopic simulator in the calibration loop. The ODME is under-determined since the number of observed links is much smaller than the number of OD pairs to estimate. In this study we therefore use regularization using OD priors to obtain reasonable solutions. Because the reliability of OD priors are uncertain, Paper I also uses weights to express the confidence into the priors and the the observed link flows, and iteratively adjusts the weights to prioritize fitting the link flow observations.

Path Flow Estimation using PLSR

Especially for the described nowcasting DT use cases, i.e., network-wide monitoring of emissions and real-time PT operation control, demand calibration needs to be fast to explore alternative strategies repeatedly or even online. Conventional ODME, as currently used in the emission-oriented DT, can be computationally demanding because it embeds a dynamic traffic assignment in each iteration and depends on prior OD matrices to regularize the ill-posed inverse problem. This motivates the exploration of alternative, data-driven estimation methods that can reconstruct demand at a lower computational cost.

Instead of estimating OD flows and repeatedly assigning demand to the network to eventually receive path flows, path flow estimators estimate path flows directly to avoid the costly assignment step. The traffic assignment step is hereby implicitly modeled. Path flows $\mathbf{f}$ and link flows $\mathbf{y}$ are related through the link–path incidence matrix A via

$$\mathbf{y} = A \times \mathbf{f}. \tag{3.1}$$

The entries of A are defined as

$$A_{lp} = \begin{cases} 1, & \text{if link } l \text{ is contained in path } p, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

In realistic networks only a subset of links is observed and the matrix A is typically rank-deficient due to collinearity (e.g., caused by overlapping paths). Thus, the inverse path flow reconstruction problem is typically ill-posed, i.e., many path flow vectors correspond to the same link flow observations [39].

To address this, Paper III evaluates PLSR as a data-driven estimator that learns a low-dimensional latent representation maximizing covariance between predictors (observed link flows) and responses (path flows) [42], thereby providing implicit regularization for the inverse mapping. The method requires samples of paired link and path flow vectors, obtained for example from simulation outputs or historical datasets where the true path flows are available.

PLSR is generally suitable for underdetermined problems, where there are more variables to estimate than available observations [42]. It has been successfully applied in related settings such as link capacity estimation [48] and internet traffic estimation [49], but has not yet been explored for traffic-related path flow estimation. To assess its applicability in this context, we design an experimental set-up with synthetic data-generating processes that mimic different practically relevant OD and path flow dependence patterns. These include independent OD demand, correlated OD demand, and cases where dependencies arise jointly from OD variability and path choice behavior via mixture models for OD flows and stochastic path choice. This set-up enables a systematic assessment of the real-world conditions under which PLSR is expected to provide reliable path flow estimates.

3.3 Simulation

Traffic simulation forms the core analytical component of the DT platform, enabling monitoring beyond observed locations through simulation-based estimates (nowcasting) and allowing exploration of alternative demand and infrastructure scenarios (forecasting). While camera-based sensing, explored in Paper I, provides localized measurements at selected sites, simulation allows the DT to approximate traffic states and associated emissions across the entire study area, including links without detectors, and to quantify how demand changes through network adjustments or policy interventions affect emissions network-wide.

Traffic simulation models can be grouped based on their level of detail into macroscopic, mesoscopic, and microscopic approaches [50]. Macroscopic models represent traffic as continuous flows and do not model individual vehicles and their interactions, which limits their ability to simulate phenomena such as queue spillbacks, detailed interactions at intersections, or stop-and-go waves that affect emissions. Mesoscopic models simulate individual vehicles but model their movements in aggregated forms such as queues, offering high computational efficiency but still lack the modeling of detailed vehicle interactions. Microscopic models, on the other hand, explicitly simulate the trajectories and interactions of individual vehicles in the network, producing high resolution speed and acceleration data that are suitable for microscopic emission modeling, as well as modeling PT–private traffic interactions.

Given this thesis' focus on emission-oriented and PT-related use cases that require high spatial and temporal resolution, microscopic-level is selected as the primary level of detail for simulations in both DT implementations. This enables the platform to represent fine-grained traffic dynamics at the street and intersection level, capture the impact of local control measures (e.g., parking availability, signal plans, speed limits, priority rules), and generate the vehicle-level data required for instantaneous microscopic emission models.

A major drawback of microscopic simulations is their computational cost associated with simulating detailed vehicle movements and their interactions. Processes requiring multiple simulation runs, such as demand or route flow calibration, can therefore require long execution times. Since the computational complexity scales with network size, employing microscopic simulation quickly becomes infeasible for larger areas. This creates a challenge, for instance, when demand is projected for a broader region but detailed emission estimates are required for a smaller subarea within it. To still obtain the targeted high resolution emission estimates for this study area, we employ a hybrid simulation approach in Paper I. This approach combines two simulation models with different resolutions, e.g., a microscopic simulation model and a mesoscopic simulation model. Using the example of the calibration of route flows, the mesoscopic simulation model can be used to repeatably simulate a larger area e.g., as part of Dynamic Traffic Assignment. The resulting route flows obtained from this simulation step are subsequently filtered to the area of interest and used as input for the microscopic simulation model. It is important to note, that since the included simulation models use different modeling assumptions their results are not be perfectly compatible. While this still requires additional calibration in the microscopic model for the smaller selected area, the results from the mesoscopic simulation model can serve as a starting point, reducing the computational burden for the microscopic model.

In both the emission-oriented DT (Paper I) and the PT DT pipeline (Paper II), the microscopic simulator SUMO is used as the main traffic simulation engine. SUMO is open source, widely adopted in research and practice, and supports multiple modes (e.g., private vehicles and buses) as well as integration with standardized PT formats (GTFS), which aligns with the DT's extensibility and open data and software requirements [51]. It models individual vehicles using car-following, lane-changing, and intersection-control models and produces both disaggregated FCD and aggregated link- and lane-level statistics.

As part of the hybrid simulation framework, used in Paper I we couple SUMO with the mesoscopic simulator MATSim [52]. MATSim is a widely used open source, queue-based simulator. The coupling of SUMO and MATSim and their implications has been already explored in the literature [53–56]. In Paper I, MATSim is used to calibrate route flows for OD demand for a larger area first to produce the OD matrix for the study area for which emission estimates are required. Subsequently, this OD matrix is used for calibrating route flow for the study area with SUMO. The resulting microscopic FCD can then be used for emission estimation.

SUMO's integrated microscopic emission models provide both network-wide, emission estimates using the simulation obtained FCD, as well as emission nowcasting by replacing the simulated vehicle driving parameters with real-world measurements collected at sensor locations. SUMO contains two emission models: an HBEFA-based model [57] and the PHEMlight model. The PHEMlight model is derived from the detailed PHEM [58] vehicle emission model. In the emission-oriented DT, the HBEFA-based formulation is selected because it provides emission factors for a broad set of vehicle categories and traffic situations and is designed for scenarios where detailed per-vehicle technology attributes are not fully known. SUMO reformulates HBEFA into a continuous microscopic function of instantaneous speed and acceleration, enabling second-by-second emission estimation for pollutants such as CO_2, NO_x, CO, and PM. This yields vehicle-level emission trajectories that can be aggregated to link-level emission rates (e.g., kg/km/h) for visualization and analysis.

3.4 System Integration and Visualization

In the DT's digital space, the previously described components are integrated to achieve interoperability within a single coherent workflow that supports nowcasting, retrospective analysis, and scenario-based forecasting. This integration

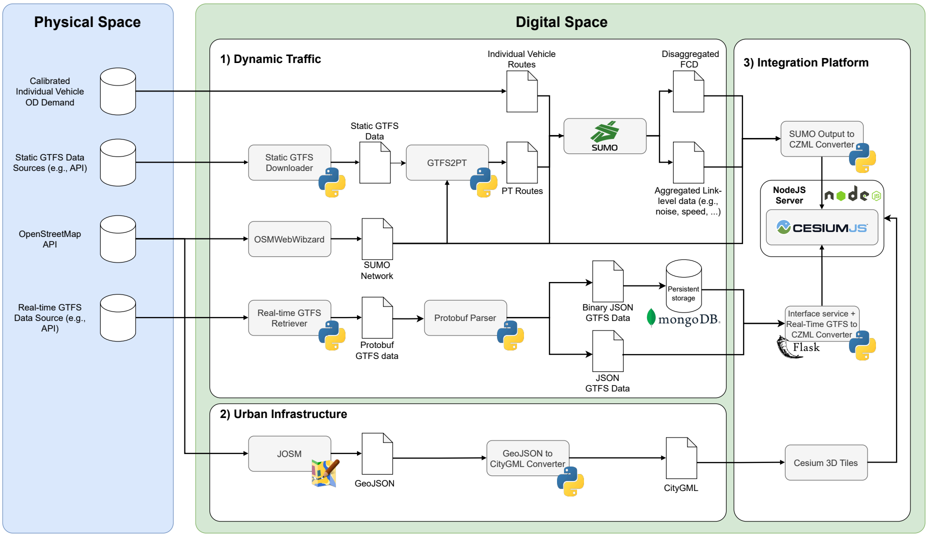


Figure 3.3: PT DT pipeline overview (source: Paper II).

includes processing, and storing of the real-world data sources for immediate visualization or using them as input for the simulation and comparing its results with real-world data. In the following the integration methodology is exemplarily presented for the PT DT of Paper II. Thereafter, a brief summary of additional steps for the emission monitoring and control application (Paper I) is provided. Figure 3.3 depicts the steps required to collect and process the individual data sources for simulation and visualization.

Paper II, groups the aforementioned components in a 1) dynamic traffic part comprising traffic data, network, and simulation, 2) an urban infrastructure part containing mainly the retrieval and processing of built environment data, and 3) their combined visualization. The key challenge is that the input for the individual components originate from heterogeneous sources with different data formats (e.g., XML-based simulator outputs, and binary real-time feeds), time semantics (e.g., static, and dynamic), and intended use (e.g., real-time monitoring or replay of historic data), which requires harmonization to ultimately enable joint analysis and visualization.

The PT DT's visualization is realized through the web-based open source software CesiumJS [59], which aligns with the platform's demands for openness and extendability. It allows to visualize dynamic geospatial data via CZML [60] (a JSON-based format for time-varying objects) and static 3D objects via 3D Tiles [61]. To be visualized and analyzed jointly in the Cesium-based platform, all the introduced components from simulation, over real-time data to the built environment need to be transformed ultimately in these formats.

Beginning with the simulation component, for effective PT planning, the pipeline relies on the existence of calibrated OD demand to jointly simulate PT with private vehicles. Paper II assumes this demand and resulting routes to be given. The static GTFS data, used to simulate PT, is usually provided by transport operators via an API and follows a standard schema, which allows the following processing steps to be easily transferred to other cities. After retrieving the data, it is fused with a corresponding network obtained from OpenStreetMap. The network itself is downloaded via SUMO's OSMWebWizard, which allows to specify the region of interest and directly download the network and data (e.g., PT stop locations). Via another SUMO built-in Python script GTFS2PT, the previously downloaded static GTFS data is prepared for simulation by correcting routes, and matching them to the topography of the prepared SUMO network. Following that SUMO is able to jointly simulate the individual vehicle traffic and the PT traffic and allows to create detailed disaggregated FCD and aggregated link-level data which allows to gain insights, e.g., when comparing a baseline simulation vs. an alternative scenario by comparing delays, emissions etc.

For visualization, these outputs in SUMO-specific XML-format are parsed and converted into CZML with custom Python-based scripts to support animation and interactive exploration through replay along a timeline. Aggregated edge-level measures (e.g., average speeds, emissions, delays) are assigned to fixed network segments, which are modeled as adjustable GeoJSON objects. These segments are then dynamically colored based on the simulation-derived edge-level measurements. This allows the exploration and comparison of baseline and alternative simulation scenarios within the same environment.

For real-time PT operations, the integration must support both low-latency visualization of current operations (e.g., real-time bus locations) and persistent storage for later replay. Real-time GTFS feeds are retrieved from the corresponding API and deserialized. Deserialization is necessary as real-time GTFS feeds are usually stored in binary Protobuf format to ensure efficient streaming [62]. Following that, the data is on the one hand converted to CZML to visualize the latest update for nowcasting, while the decoded messages are also stored in a MongoDB for time-window queries and retrospective analysis.

Finally, for contextualization and informed decision-making, lightweight building models are added from open data and converted into 3D Tiles. Thereafter, the Cesium-based platform can jointly visualize the static 3D environment and the time-varying CZML layers from simulation output and real-time operations. Together, the common CZML/3D Tiles representations and the web-based client enable switching between monitoring the current PT operations (nowcasting), replaying and analyzing historic operations (retrospective analyzes), and simulating the future (forecasting) within one interactive 3D environment.

The emission-centered DT (Paper I) follows a similar integration architecture, but extends the processing and simulation components to support emission nowcasting and network-level forecasting.

On the nowcasting side, traffic cameras feed video footage into a computer-vision pipeline that detects and tracks vehicles, identifies their types, and extracts driving variables relevant to emissions (e.g., speed and acceleration). These variables are fed into an HBEFA-based emission model to compute near-real-time, local emission estimates. At the same time, the cameras provide traffic count observations to perform ODME. This provides the input for the SUMO-based simulation environment to compute emission estimates for the area surrounding the cameras. Since the camera coverage may not allow OD estimation for larger areas, the emission-oriented DT further adopts a hybrid simulation pipeline that connects the mesoscopic simulator MATSim [52] with the microscopic simulator SUMO to generate vehicle trajectories from large-scale OD demand for a selected study area to obtain emission estimates for this area under both current and alternative scenarios.

For interactive exploration of these results, Paper I implements a Unity-based visualization and couples the backend processing, simulation, and data storage via a lightweight communication layer built with MQTT. Within the Unity-based platform, the built environment is derived from Google Earth, providing more detailed and realistic visualizations than the PT DT, but at the cost of higher resource requirements and reduced openness and transferability.

3.5 Case Study

The design-cycle evaluation in this thesis is conducted as a case-study demonstration in Kista, Stockholm, Sweden (see Figure 3.4, which serves as a prototypical implementation environment for both DT artifacts (Paper I and Paper II). The purpose of the case study is to show the technical feasibility and interoperability of the individual components, rather than to provide a fully validated, policy-ready deployment. Kista was selected because it serves as a technological test-bed in Stockholm and simultaneously contains a sufficiently heterogeneous transport setting for DT prototyping, including urban road traffic and PT services. In both DT demonstrations, the Kista network is based on an OpenStreetMap (OSM) street network, ensuring compatibility between the simulation and visualization layers used for analysis.

For the emission-oriented DT (Paper I), the case study demonstrates the emission nowcasting and forecasting abilities, as well as their visualization. An experimental camera set-up allows to demonstrate the computer vision model used for the localized emission nowcasting. The number of cameras in the area, however, is insufficient for OD estimation. Therefore, calibrated OD demand for the Stockholm area is used as part of the hybrid simulation framework, which allows extraction of the OD demand for the study area used for emission forecasting under current and alternative conditions (what-if analysis). As an example for such a what-if analysis, the case study includes a parking scenario, in which parking availability in a neighborhood is reduced. The demand is adjusted according to a literature-based elasticity parameter to evaluate the change in emissions in

Figure 3.4: Selected case study area Kista in Stockholm for emission (Paper I) and PT DT (Paper II).

the affected area. Emissions for the current and alternative parking scenarios are computed using a microscopic emission model of the SUMO-based simulation environment, enabling second-by-second pollutant estimation (e.g., CO_2, NO_x, CO, PM) from speed and acceleration trajectories, visualized in the DT platform.

For the PT-oriented DT development pipeline (Paper II), the case study illustrates how a lightweight DT can be implemented with low entry barriers by relying on open data and open-source software within an actual real-world example. It shows the interoperability of the pipeline components by integrating real-time monitoring of operations, retrospective analysis, and joint PT-traffic simulation outputs for what-if scenario exploration within a single interactive environment. The PT system is represented using GTFS, including both static schedules and real-time feeds retrieved via Stockholm's publicly accessible Trafiklab API [63]. The simulation component is demonstrated using synthetic car traffic together with the obtained static GTFS schedules for the Kista area on a representative weekday. To demonstrate the platforms visualization capabilities, the results of the simulation and real-time feeds are translated to CZML to be jointly visualized with the processed 3D built environment.

In contrast to Papers I and II, which concern the development of a technical artifact (i.e., the DT frameworks), Paper III evaluates a key methodological building block relevant to the DT's simulation component. Specifically, it studies data-driven reconstruction of path flows from partially observed link counts

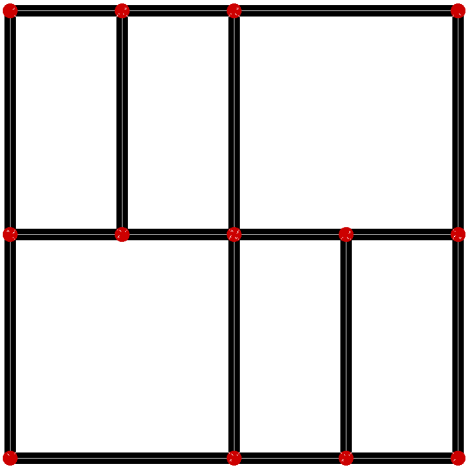


Figure 3.5: Selected synthetic network for PLSR experiments (Paper III).

using PLSR. To assess its stability and transferability under controlled, real-world reflecting OD and path choice structures, Paper III uses a synthetic network (see Figure 3.5) and multiple data-generating processes. The generators include independent OD demand with fixed path shares, fixed OD demand with multinomial path choices, and correlated OD demand modeled via a Poisson log-normal distribution, combined with stochastic path-choice mechanisms based on either independent traveler-specific path-cost perturbations or shared link-level perturbations that induce cross-OD path flow correlations. Using these data-generating processes, the method is assessed under both in-distribution and out-of-distribution scenarios to quantify stability and transferability. This experimental design allows for assessing under which conditions the PLSR method can be expected to provide reliable path flow estimates and thereby evaluates its suitability as an input component for the DT's simulation component.

Chapter 4

Thesis Contributions

This chapter summarizes the contributions of the papers included in this thesis and maps them to the overarching research objectives (see Table 4.1).

	Research Objective	**Paper**		
		I	II	III
RO1	Identify the common building blocks of urban transportation DTs and formulate a transferable architecture integrating real-time data, simulation, and interactive visualization.	×	×	
RO2	Implement and demonstrate a DT architecture for network-level road traffic emission nowcasting, scenario-based forecasting, and retrospective analysis.	×		
RO3	Implement and demonstrate an open, automated, and extensible DT development pipeline for PT DTs.		×	
RO4	Implement PLSR for path flow reconstruction from link flow observations and evaluate its robustness and transferability.			×

Table 4.1: Research objectives mapped to included papers.

4.1 Paper I

Paper I addresses both RO1 and RO2 by developing and prototypically demonstrating a transferable DT architecture for network-level road traffic emission monitoring and scenario-based analysis. The paper focuses on the method selection and technical integration to form an end-to-end emission-focused DT framework, that is transferable to other cities. Consistent with the DSR methodological framework, the primary result of this study is the technical artifact that consists of a modular DT framework coupling sensing, traffic and emission modeling, and interactive 3D visualization. With the development of these components, the paper makes the following three contributions:

1. It develops an emission-oriented DT architecture that replicates the physical space through IoT sensing, camera-based near-real-time emission nowcasting, and simulation-based network-level forecasting, all connected to an interactive 3D visualization platform through a data streaming set-up.
2. It presents a computer-vision-based emission nowcasting component using vehicle classification and tracking, along with microscopic emission modeling.
3. It realizes a network-wide emission estimation and scenario forecasting pipeline by combining camera-based OD estimation and hybrid traffic simulation.

The developed architecture is demonstrated through a case study in Kista, Stockholm, showcasing the interoperability of the components. The results of this case study highlight the value of coupling both camera-based emission nowcasting and network-level simulation. While localized emissions and flow parameters can be estimated well through the cameras, network-wide emission estimation through simulation remains expensive. However, the simulation component was found to be well suited for offline planning and comparative analysis, as demonstrated by an example alternative scenario for reducing parking space availability.

4.2 Paper II

Paper II develops an open, automated, and extendable DT development pipeline for PT that lowers technical and resource-related barriers by relying on open data, open standards, and open-source components. By developing this technical artifact, Paper II addresses RO1 and RO3 and makes the following four contributions:

1. Paper II develops an automated, extendable development pipeline for PT DTs based on open data and open-source software, enabling the creation of DTs with low entry barriers and allowing for future extensions.
2. It enables the integration of standardized PT data, including GTFS-based real-time operations, schedule-based simulation results that support PT nowcasting, retrospective analysis, and scenario exploration in one platform.
3. It provides an automated workflow to transform and jointly visualize OSM-derived 3D building data with dynamic vehicle movement.
4. Finally, it maps potential use cases and extensions of the proposed pipeline for academics and practitioners.

Like Paper I, this work demonstrates the technical feasibility and the process of setting-up the DT as part of a real-world case study for Kista, Stockholm. The case study showcases the DT's ability to represent both real-time PT operational data and simulation results in a 3D interactive environment for improved contextualization.

In line with the open-source nature of the development pipeline, the code and documentation are accessible via GitHub[1].

4.3 Paper III

In contrast to Papers I and II, the third paper is a quantitative study that evaluates PLSR as a data-driven method for path flow estimation, motivated by the need for reliable online path flow estimation in urban DTs. The method's ability to address the ill-posed nature of the problem, to rely on limited training data, and its execution speed presents it as a promising solution.

Since the PLSR method has not yet been employed in the path flow estimation context, Paper III evaluates its suitability for this problem setting and, by that, addresses RO4. The paper makes the following three contributions:

1. It formulates a PLSR-based online path flow estimator using link flow observations.
2. It creates synthetic data-generating processes that replicate realistic OD patterns and path flow correlation structures.
3. Finally, it performs an in-distribution and out-of-distribution assessment of PLSR and derives real-world implications from the obtained results.

The experimental results suggest that PLSR is best suited for scenarios in which the main variability in path flows originates from the OD demand uncertainties, as they are most recoverable from the resulting link flow observations. With rising path choice variability, e.g., through traveler-specific cost perception uncertainty or link-level day-to-day perturbations, the recoverability of path flows is reduced. However, the results indicate that once the system enters a dominant path choice competition regime, the reproduction error stabilizes. These regimes are characterized by a shift from OD-driven positive path flow correlations within OD pairs to negative correlations due to path choice competition. The correlation strength, and with that the regime, is driven by the level of cost uncertainties. In particular, the out-of-distribution experiments show that, in practical applications, these regimes and the intensity of path choice competition must be monitored carefully, since performance can deteriorate substantially in regions with transitions between regimes. PLSR, however, shows stable performance when fitted and deployed with varying levels of cost uncertainties as long as no regime shifts are present.

---

[1] `https://github.com/MobilityInformaticsLab/opdtwin`

Chapter 5

Summary, Discussion and Future Directions

This chapter summarizes the results of the thesis, discusses its contributions as well as its limitations, and outlines directions for future research.

5.1 Summary

By replicating real-world components in a digital space, DTs enable efficient and safe monitoring, planning, and control for urban transportation systems. This thesis approached the architecture development of DTs using emissions and PT as use cases.

DTs have already been developed to observe and reduce urban traffic induced emissions. Currently, however, their capabilities focus mainly on connecting real-world sensors to KPI dashboards or performing simulation-based emission estimation in isolation. Motivated by this gap, Paper I developed and prototypically demonstrated a transferable DT framework that targets high-resolution, network-level emission monitoring and scenario-based analysis within one integrated architecture. Concretely, the DT couples camera-based traffic sensing for localized flow and emission nowcasting with an OD estimation and hybrid simulation pipeline for network-wide emission forecasting, thereby enabling both near-real-time monitoring at sensor locations and what-if analysis beyond sensor coverage. A central outcome is the end-to-end technical integration of heterogeneous components (sensing, data management, modeling, and visualization) into a modular architecture, demonstrated in a case study in Kista, Stockholm.

To overcome the technological and resource barriers driving low adoption of DTs in PT, this thesis proposed, as part of Paper II, an open, extendable PT DT development pipeline. The pipeline uses standardized GTFS data for representing PT schedules and real-time operations, integrates simulation outputs for scenario exploration, and combines these data sources in a 3D interactive platform to improve contextualization, diagnosis of spatial dynamics, and communication of results. A prototypical implementation for Kista, Stockholm demonstrates how real-time feeds and simulated trajectories can be explored jointly within one system, thereby illustrating the feasibility and transferability of the pipeline concept.

To drive the simulation-based analysis modules of both DT architectures, Paper III evaluated PLSR as a data-driven solution for online path flow estimation. The work studied PLSR's stability and transferability under multiple synthetic data-generating processes representing different OD and path flow correlation regimes. The results for a synthetic network show that PLSR performs best when the main variability in path flows originates from OD demand uncertainty, while increasing path choice variability generally reduces path flow recoverability. The out-of-distribution experiments indicate that PLSR can transfer reliably when fitting and deployment conditions remain within the same correlation regime, but performance can deteriorate substantially under regime shifts.

This shows that effective transport DTs require not only the technical integration of sensing, simulation, and visualization, but also robust methods for estimating the traffic demand that drives simulation-based analysis.

5.2 Discussion

Even though Papers I and II established individual architecture designs accommodating two different use cases, they share fundamental components. This includes the connection to real-world data, calibration and execution of simulations for what-if analyzes, as well as result visualization in an interactive 3D platform. In this way, the thesis demonstrates how DT platforms can be built, points out challenges, and provides directions for cities and transport authorities to address them. Besides that, the studies present urban transportation use cases for which a DT is a valuable tool, some of which have been implemented as part of the included case studies.

Both architectures show that a DT is not a monolithic architecture but consists of multiple components, forming a system of systems. The first study uncovered the significant development effort required to connect these subsystems and harmonize the data shared among them. Especially, Paper II provides guidelines and implements concrete software that automates large parts of the development steps using standardized data sources available in many cities.

The thesis successfully demonstrated the DTs technical capabilities in the included case studies. Even though these case studies do not evaluate their potential quantitatively, the thesis identified many advantages DTs have over purely simulation-based or sensor dashboard based approaches.

One advantage is the possibility to explore the real-time and simulation data in a 3D interactive environment for better contextualization and more informed, transparent decision making. While this advantage has not been quantitatively explored in the included studies, the results of the study by Grant et al. [32] imply this advantage. There, however is a trade-off between the 3D visualization's level of detail and the required effort to create it. Many DT platforms, including the one presented in Paper I, use highly realistic visualizations, which often require substantial manual effort to set up. While visually appealing, it remains unknown which use cases actually benefit from this high level of detail. For contextualization, the visualization of dynamic data at the building level or using the 3D models as input for dispersion models could also be realized using a lower level of detail. Thus, Paper II questions the necessity of highly detailed 3D visualizations and implements a DT development pipeline that automatically retrieves and processes building models in a lower level of detail from open data, reducing manual labor. Further, for some use cases 3D visualizations may even be an obstacle, e.g., buildings blocking the view for street level analysis.

At the current stage, the proposed process consists of exploring real-time monitoring results (nowcasting) and simulation-based what-if analyses (forecasting), using these outputs to identify suitable interventions, and then implementing the selected actions manually in the physical system, for example by reducing parking availability or adjusting traffic control schedules. The DT architectures presented in this thesis therefore rely on a human-in-the-loop feedback process, meaning that the connection from the digital system back to the physical system is not yet automated.

Although the presented case studies demonstrate the feasibility of urban transportation DT architectures, important deployment challenges remain, particularly with respect to privacy and data security. These issues are especially relevant when DTs rely on detailed sensing such as camera images. Recent public debate about a regional DT in Germany shows that such systems may be valued for public benefits such as crisis response, while simultaneously raising security concerns when detailed geospatial information is made openly available [64].

The first two studies show that simulation-based DTs require fast and reliable demand estimation in order to support what-if analysis. While execution speed may be less important for long-term planning applications, it is particularly relevant for short-term operational tasks. In this context, the PLSR-based path flow estimator investigated in Paper III presents a promising approach because it enables computationally efficient path flow reconstruction and is therefore suitable for simulation-based planning and real-time DT workflows.

The results indicate that the usefulness of PLSR depends on the stability of the underlying traffic regime. The method transfers well when fitting and deployment conditions share similar correlation structures, but its performance may degrade when those structures shift. This suggests that when PLSR-based estimation is implemented, continuous monitoring is required to detect regime shifts and trigger model updates when necessary.

A limitation of the study is that the analysis was performed on a synthetic, uncongested network with static demand conditions. Although this makes it possible to isolate demand and path flow correlations in a controlled way, it does not evaluate the estimator's performance under congested, dynamic traffic conditions.

Apart from the implementation of the path flow estimator, Paper III also contributes a structured way of generating synthetic scenarios for controlled scenario evaluation. This is relevant for DT workflows, as these are not only used to reproduce the observed reality but also to support the analysis of alternative and previously unseen conditions. In that sense, synthetic scenario generation provides a useful and computationally efficient way to test robustness and better prepare DT-related methods for sim-to-real challenges.

5.3 Future Directions

While both studies demonstrated the technical feasibility of DT development for emission monitoring and PT, future work should focus on increasing the maturity level of the presented DT architectures. Concretely, the manual human-decision-based connection between physical and digital space currently used in both DTs, should be replaced with an end-to-end cyber-physical feedback loop.

A central question for DT research is the practical value of the DT for transport practices compared with existing solutions. Is it worth the effort to build such a complex digital platform? While there are arguments and some results demonstrating the value of similar platforms, a dedicated quantitative comparison against current alternatives, such as dashboards, standalone simulations, or existing planning tools is missing and thus should be investigated. Similarly, the 3D visualization component and the described trade-off concerning the level of detail require quantitative evaluation, comparing it against 2D and alternatives offering a lower level of detail.

Apart from researching architectural design questions to enable certain functionalities for urban DTs, it is equally important to identify ways to provide these functionalities in a safe and privacy preserving way. While many DT implementations remain experimental, this aspect is important to investigate further in the future to make DTs for urban transportation deployable.

Moreover, to further assess PLSR as a data-driven online path flow estimator for DT workflows, the method needs to be evaluated on larger and more realistic networks. Future research should therefore investigate its performance under congested and dynamic traffic conditions, as well as under more realistic sensing set-ups, such as noisy or incomplete measurements. In addition, since the results of this thesis indicate that the performance of PLSR depends on the stability of the underlying traffic regime, future work should develop monitoring systems capable of detecting these regime shifts and triggering model recalibration as soon as changing traffic patterns are detected.

Beyond the estimator itself, future research should also further investigate the generation of synthetic data for DT workflows. In Paper III, synthetic data was primarily introduced to generate demand and route choice structures. Future studies should extend this to generate plausible data representing systematic sensing errors, missing data, and other real-world imperfections. Such controlled scenario generation would support more realistic robustness testing and model training, which, in turn, helps to reduce the gap between simulation and real-world deployment.

References

[1] M. Grieves, “Digital twin: Manufacturing excellence through virtual factory replication,” 03 2015.

[2] E. Glaessgen and D. Stargel, “The digital twin paradigm for future NASA and U.S. air force vehicles,” in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference.*

[3] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, “Digital twin in industry: State-of-the-art,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019.

[4] L. Li, B. Lei, and C. Mao, “Digital twin in smart manufacturing,” *Journal of Industrial Information Integration*, vol. 26, p. 100289, 2022.

[5] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital twin in manufacturing: A categorical literature review and classification,” *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

[6] M. Turab and S. Jamil, “A comprehensive survey of digital twins in healthcare in the era of metaverse,” *BioMedInformatics*, vol. 3, no. 3, pp. 563–584, 2023.

[7] W. Purcell and T. Neubauer, “Digital twins in agriculture: A state-of-the-art review,” *Smart Agricultural Technology*, vol. 3, p. 100094, 2023.

[8] G. Mylonas, A. Kalogeras, G. Kalogeras, C. Anagnostopoulos, C. Alexakos, and L. Muñoz, “Digital twins from smart manufacturing to smart cities: A survey,” *IEEE Access*, vol. 9, pp. 143222–143249, Oct 2021.

[9] M. S. Irfan, S. Dasgupta, and M. Rahman, “Toward Transportation Digital Twin Systems for Traffic Safety and Mobility: A Review,” *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 24581–24603, 2024.

[10] Y. Zhang and H. Zhang, “Urban digital twins: Decision-making models for transportation network simulation,” in *Proceedings of the 2022 International Conference on Computational Infrastructure and Urban Planning*, CIUP ’22, (New York, NY, USA), p. 18–21, Association for Computing Machinery, 2022.

[11] United Nations Department of Economic and Social Affairs, *The sustainable development goals report 2025.* United Nations, July 2025.

[12] S. A. P. Kumar, R. Madhumathi, P. R. Chelliah, L. Tao, and S. Wang, “A novel digital twin-centric approach for driver intention prediction and traffic congestion avoidance,” *J. Reliab. Intell. Environ.*, vol. 4, pp. 199–209, Dec. 2018.

[13] K. M. Alam and A. El Saddik, “C2ps: A digital twin architecture reference model for the cloud-based cyber-physical systems,” *IEEE Access*, vol. 5, pp. 2050–2062, 2017.

[14] M. Buechel, M. Schellmann, H. Rosier, T. Kessler, and A. Knoll, “Fortuna: Presenting the 5g-connected automated vehicle prototype of the project providentia,” *Researchgate*, 11 2019.

[15] H. Kamal, W. Yánez, S. Hassan, and D. Sobhy, “Digital-twin-based deep reinforcement learning approach for adaptive traffic signal control,” *IEEE Internet of Things Journal*, vol. 11, no. 12, pp. 21946–21953, 2024.

[16] A. Rudskoy, I. Ilin, and A. Prokhorov, “Digital twins in the intelligent transport systems,” *Transportation Research Procedia*, vol. 54, pp. 927–935, 2021. International Scientific Siberian Transport Forum - TransSiberia 2020.

[17] Z. Wang, K. Han, and P. Tiwari, “Digital twin-assisted cooperative driving at non-signalized intersections,” *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 198–209, 2022.

[18] T. Zhang, X. Liu, Z. Luo, F. Dong, and Y. Jiang, “Time series behavior modeling with digital twin for internet of vehicles,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, Dec. 2019.

[19] C. Anda, S. A. Ordonez Medina, and K. W. Axhausen, “Synthesising digital twin travellers: Individual travel demand from aggregated mobile phone data,” *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103118, 2021.

[20] A. Sohail, B. Shen, M. A. Cheema, M. E. Ali, A. Ulhaq, M. A. Babar, and A. Qureshi, “Beyond data, towards sustainability: A sydney case study on urban digital twins,” *PFG-J. Photogramm. REMOTE Sens. GEOINFORMATION Sci.*, 2025 APR 25 2025.

[21] P. Zhivkov, S. Fidanova, and I. Dimov, “Digital twin-based framework for real-time monitoring and analysis of urban mobile-source emissions,” *ATMOSPHERE*, vol. 16, June 2025.

[22] D. Teutscher, F. Bukreev, A. Kummerlaender, S. Simonis, P. Baechler, A. Rezaee, M. Hermansdorfer, and M. J. Krause, “A digital urban twin enabling interactive pollution predictions and enhanced planning,” *Build. Environ.*, vol. 281, Aug. 2025.

[23] F. Jiang, L. Ma, T. Broyd, W. Chen, and H. Luo, “Digital twin enabled sustainable urban road planning,” *Sustain. CITIES Soc.*, vol. 78, Mar. 2022.

[24] E. Llagostera-Brugarola, E. Corpas-Marco, C. Victorio-Vergel, E. Lopez-Aguilera, F. Vazquez-Gallego, and J. Alonso-Zarate, “A digital twin for intelligent transportation systems in interurban scenarios,” *Appl. Sci.-BASEL*, vol. 15, July 2025.

[25] A. De Benedictis, F. Rocco di Torrepadula, and A. Somma, “A digital twin architecture for intelligent public transportation systems: A FIWARE-based solution,” in *Web Wirel. Geogr. Inf. Syst.* (M. Lotfian and L. L. L. Starace, eds.), (Cham), pp. 165–182, Springer Nature Switzerland, 2024.

[26] S. Laso, L. Tore-Gálvez, J. Berrocal, C. Canal, and J. M. Murillo, “Deploying digital twins over the cloud-to-thing continuum,” in *2023 IEEE Symposium on Computers and Communications (ISCC)*, vol. 1, p. 1–6, IEEE, July 2023.

[27] A. Amrani, H. Arezki, D. Lellouche, V. Gazeau, C. Fillol, O. Allali, and T. Lacroix, “Architecture of a public transport supervision system using hybridization models based on real and predictive data,” in *2020 23rd Euromicro Conference on Digital System Design (DSD)*, pp. 440–446, 2020.

[28] S. Van Den Berghe, “A processing architecture for real-time predictive smart city digital twins,” in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 2867–2874, 2021.

[29] P. Ruiz, M. Seredynski, Á. Torné, and B. Dorronsoro, “A digital twin for bus operation in public urban transportation systems,” in *Big Data Intelligence and Computing* (C.-H. Hsu, M. Xu, H. Cao, H. Baghban, and A. B. M. Shawkat Ali, eds.), (Singapore), pp. 40–52, Springer Nature Singapore, 2023.

[30] W. Li, B. Wang, R. Sun, L. Ai, and Z. Lin, “Energy-efficient multimodal mobility networks in transportation digital twins: Strategies and optimization,” *Energy*, vol. 318, p. 134587, 2025.

[31] M. Shulajkovska, M. Smerkol, G. Noveski, and M. Gams, “Enhancing urban sustainability: Developing an open-source AI framework for smart cities,” *Smart Cities*, vol. 7, no. 5, pp. 2670–2701, 2024.

[32] G. Herbert and X. Chen, “A comparison of usefulness of 2d and 3d representations of urban planning,” *Cartography and Geographic Information Science*, vol. 42, no. 1, pp. 22–32, 2015.

[33] N. C. C. Tiglao, E. Sanciangco, and N. J. Tolentino, "Digitally enabled collaborative governance for sustaining bus reforms on the EDSA Busway in Metro Manila," *Res. Transp. Econ.*, vol. 113, p. 101633, 2025.

[34] H. Yang, T. Sasaki, Y. Iida, and Y. Asakura, "Estimation of origin-destination matrices from link traffic counts on congested networks," *Transportation Research Part B: Methodological*, vol. 26, no. 6, pp. 417–434, 1992.

[35] M. G. Bell, C. M. Shield, F. Busch, and G. Kruse, "A stochastic user equilibrium path flow estimator," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 3, pp. 197–210, 1997.

[36] Y. Nie and D.-H. Lee, "Uncoupled method for equilibrium-based linear path flow estimator for origin-destination trip matrices," *Transportation Research Record*, vol. 1783, no. 1, pp. 72–79, 2002.

[37] C. Chen, Y. Cao, K. Tang, and K. Li, "Dynamic path flow estimation using automatic vehicle identification and probe vehicle trajectory data: A 3d convolutional neural network model," *Journal of Advanced Transportation*, vol. 2021, no. 1, p. 8877138, 2021.

[38] H. Lam, H. Shao, S. Cao, and H. Yang, "Origin-destination demand estimation models," in *Encyclopedia of Transportation*, vol. 4, pp. 515–518, Elsevier Ltd., May 2021.

[39] H. Yang and J. Zhou, "Optimal traffic counting locations for origin–destination matrix estimation," *Transportation Research Part B: Methodological*, vol. 32, no. 2, pp. 109–126, 1998.

[40] K. Pearson, "LIII. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[41] M. G. Bell, "The estimation of origin-destination matrices by constrained generalised least squares," *Transportation Research Part B: Methodological*, vol. 25, no. 1, pp. 13–22, 1991.

[42] H. Abdi, "Partial least square regression (pls regression)," *Encyclopedia for research methods for the social sciences*, vol. 6, no. 4, pp. 792–795, 2003.

[43] A. R. Hevner, "The three cycle view of design science research," *Scandinavian Journal of Information Systems*, vol. 19, no. 2, p. 87, 2007.

[44] GTFS, "Gtfs overview documentation." `https://gtfs.org/documentation/overview/`, 2024. [Accessed: 2024-08-14].

[45] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ." `https://www.openstreetmap.org`, 2017.

[46] G. Gröger and L. Plümer, “Citygml – interoperable semantic 3d city models,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 71, pp. 12–33, 2012.

[47] X. Zhou and J. Taylor, “DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and calibration,” *Cogent Eng.*, vol. 1, p. 961345, Oct 2014.

[48] G. Wei, J. Ekström, and G. Flötteröd, “Calibration of urban road network capacities,” in *HEART 2022 10th Symp. Eur. Assoc. Res. Transp.*, Swedish National Road and Transport Research Institute, Traffic analysis and logistics, 2022.

[49] F. J. Cuberos, I. Herrera, K. Wasielewska, and J. Camacho, “Network tomography and partial least squares for traffic matrix estimation,” in *2021 17th Int. Conf. Netw. Serv. Manag. CNSM*, pp. 259–263, 2021.

[50] J. d. D. Ortúzar and L. G. Willumsen, *Modelling Transport.* Chichester, West Sussex, England: John Wiley & Sons, 5th ed., 2024.

[51] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wiessner, “Microscopic traffic simulation using sumo,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, 2018.

[52] A. Horni, K. Nagel, and K. Axhausen, eds., *Multi-Agent Transport Simulation Matsim.* London: Ubiquity Press, Aug. 2016.

[53] M. Gütlein, R. German, and A. Djanatliev, “Towards a hybrid co-simulation framework: HLA-based coupling of MATSim and SUMO,” in *IEEE/ACM 22nd Int. Symp. Distrib. Simul. Real Time Appl. -Rt*, pp. 1–9, 2018.

[54] C. Rakow, I. Kaddoura, R. Nippold, and P. Wagner, “Investigation of the system-wide effects of intelligent infrastructure concepts with microscopic and mesoscopic traffic simulation,” in *27th ITS World Congr. Hambg. Ger.*, Oct. 2021.

[55] K. Schrab, R. Protzmann, and I. Radusch, “A large-scale traffic scenario of berlin for evaluating smart mobility applications,” in *Smart Energy Smart Transp.* (E. G. Nathanail, N. Gavanas, and G. Adamos, eds.), pp. 276–287, 2023.

[56] H. Triebke, M. Kromer, and P. Vortisch, “Bridging the gap between mesoscopic transport planning and microscopic traffic simulation: An analytical and numerical analysis of traffic dynamics,” *Transp. Res. Rec.*, vol. 2677, pp. 62–76, Oct 2023.

[57] “The handbook of emission factors for road transport.” https://www.hbefa.net/, 2024.

[58] “Passenger car and heavy duty emission model.” https://www.itna.tugraz.at/en/research/areas/em/simulation/phem.html/, 2024.

[59] Cesium, “CesiumJS — cesium.com.” `https://cesium.com/platform/cesiumjs/`, 2024. [Accessed 2024-11-27].

[60] L. Zhu, Z. Li, and Z. Wang, “Representing time-dynamic geospatial objects on virtual globes using czml—part ii: Impact, comparison, and future developments,” *ISPRS International Journal of Geo-Information*, vol. 7, no. 3, 2018.

[61] Open Geospatial Consortium, “OGC API - Features: Part 1: Core - Standard.” `https://docs.ogc.org/cs/22-025r4/22-025r4.html`, 2023. [Accessed 2024-11-27].

[62] Trafiklab, “The protobuf file format.” `https://www.trafiklab.se/docs/using-trafiklab-data/the-protobuf-file-format/`, 2024. [Accessed 2025-10-21].

[63] Trafiklab, “GTFS at Trafiklab.” `https://www.trafiklab.se/api/gtfs-datasets/overview/`, 2024. [Accessed 2024-11-27].

[64] F. Flade and R. Pinkert, “Präzise Daten und Koordinaten: Sicherheitsrisiko Digitaler Zwilling?.” https://www.tagesschau.de/investigativ/ndr-wdr/geodaten-digitaler-zwilling-nrw-100.html, May 2025. [Accessed: 2026-03-25].