NEXUS

D7.1 - Validation and exploitation of metro system simulation models and AI-Applications

PROJECT ACRONYM	NEXUS
PROJECT START DATE	01/10/2024
PROJECT DURATION (MONTHS)	24
GRANT	10177985
CALL	HORIZON-JU-ER-2023-01
TOPIC	HORIZON-ER-JU-2023-EXPLR-02
CONSORTIUM COORDINATOR	STAM
TITLE OF THE DELIVERABLE	D7.1 – Validation and exploitation of metro system simulation models and AI Applications
WORK PACKAGE	WP7 — Validation, Exploitation and Improvement campaign
TYPE OF DELIVERABLE	R — Document, report
DISSEMINATION LEVEL	PU - Public
STATUS – VERSION, DATE	Final, 18/06/2026
SUBMISSION DATE	25/06/2026

Co-funded by the European Union

Co-funded by the European Union. Views and opinion expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Europe's Rail Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them. The project 'NEXUS' is supported by the Europe's Rail Joint Undertaking and its members. UK participants in this project are funded by the UKRI.

NEXUS

AUTHORS/CONTRIBUTORS

Name	Organisation	Contribution
Hing Yan Tong	AU	Drafted Chapters 1, 2, 4 and Executive Summary; Compiled the whole Draft
Marin Marinov	AU	Review of the Drafts
Lydia Egbo	AU	Passenger Demand Forecasting Use Case
Patrick Bannon	AU	Timetable Creation Use Case
Prachiti Schinde	AU	Contributed to Simul8 Network Models
Summair Anis	UNIGE	Drafted Chapters 2 and Review of Draft
Edoardo Bozza Costa	UNIGE	Drafted Chapters 2
Nicola Sacco	UNIGE	Review of the Draft
Nikos Skandamis	PT	Drafted Chapter 3.4
Nikos Avgerinos	PT	Review of Chapter 3.4
Pietro De Vito	STAM	Drafted Chapter 2 and Review of Draft
Anna Bolognesi	STAM	Drafted Chapters 2 and 3
Lissa de Souza Campos	STAM	Contributed to AnyLogic models, and overall to inputs and outputs for all network and station models
Michael, Alexander	Vif	Drafted and coordinated inputs for Chapter 3
Mirena Todorova	VTU	Coordinated data collection and passenger OD estimation for the MetroS case

QUALITY CONTROL

	Name	Organisation	Date
Internal Review	Markus Haferl Peter Schindler	TUW	12.03.2026
Internal Review	Zlatin Trendafilov Violina Velyova	VTU	18.03.2026

NEXUS

VERSION HISTORY

Version	Date	Author	Summary of changes
01	03/03/26	Hing Yan Tong	Final Version for Internal Review
02	31/03/26	Hing Yan Tong	Revised for Internal Review
03	18/06/26	Hing Yan Tong	Revised for External Review (with support from relevant partners)

APPROVED FOR SUBMISSION BY

Name	Organisation	Approval date
Pietro De Vito	STAM	03/07/2026

LEGAL DISCLAIMER

This project has been funded by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101177985. UK participants in this project are funded by the UKRI. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The information in this document is provided "as is", and no guarantee or warranty is given that it is fit for any specific purpose. The Nexus project Consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

NEXUS

Table of contents

List of figures

List of tables

Executive summary

Deliverable D7.1 reports on the validation and exploitation of metro simulation models and AI applications for the management and planning of future metro operations, including the exploration activities on the characteristics of future metro train control system as well as strategies and measures for improving data handling and cybersecurity within Task 7.1 and Task 7.2. The outcome of WP7 is to enable that the metro system models, concepts and applications will be properly tested and validated to ensure that they reflect the real-world system and its specific requirements. They can be used for decision-making purposes without reservation. This deliverable focuses mainly on documenting the key activities undertaken to validate and exploit the metro simulation models and AI applications to evaluate current (as-is) system performances and the possible impacts of future (as-is) fully automated operations.

To achieve the intended outcomes of WP7, various activities have been conducted within Task 7.1 and Task 7.2 including 1) validation and cross-validation of a multi-perspective suite of metro simulation models, 2) carried out a multi-purpose and multi-level evaluation of current metro system performances; 3) examined the possible impacts of GoA4 based on current passenger demand conditions; 4) validation of AI applications; 5) exemplary investigation of metro system energy consumption and CO_2 emissions; and 6) cybersecurity strategy assessment.

Within Task 7.1, a **multi-perspective** (using a suite of validated metro simulation models), **multi-purpose** (according to 23 scenarios organised into 5 themes), **multi-level** (looking at macro-, meso-, and microscopic level impacts) **approach was implemented to evaluate metro system performances and the possible impacts of future GoA4 fully automated operations**. Each metro simulation model was calibrated individually for accuracy and cross-validated against their counterparts to ensure consistency. Scenarios were developed in close coordination with metro operators, ensuring accurate reflection of practical metro operational concerns and relevance. Impacts were revealed at network-wide, metro line, metro station, as well as for crowd management levels. It was found that **GoA4 (in the form of reduced train headways) is most effective as a disruption-mitigation tool when the disruption affects a single node rather than the entire line.** When considering switching to GoA4 operations, **a suitable trade-off between reducing passenger wait times and improving train capacity utilisation should be made** when determining the optimal train headways.

As for AI applications, three use cases on passenger demand forecasting, timetable creation using GTFS feed, and vehicle uncleanliness detection were validated. An exemplary analysis linking AI applications to **energy and CO_2 optimisation**, particularly in combination with the Total Cost Ownership (TCO) tool and future TCS interactions was also implemented. This analysis established a connection between AI-enabled optimisation and operational cost reduction, energy efficiency improvements, sustainability targets, and future automation scenarios. Although still preliminary, these findings demonstrate a **direct link between AI exploitation and measurable environmental impact**, reinforcing the strategic relevance of AI integration in metro systems.

The Cybersecurity landscape is changing rapidly with the continuous expansion of AI use in the technologies used by the metro operators but also in the tactics and tools used by the attackers

The related section extends the completed cybersecurity baseline assessment by focusing on the future-oriented validation and safe exploitation of AI-enabled systems for metro operations and planning. Using a threat-driven, model-agnostic approach, it examines how AI/ML lifecycles (data ingestion, training, deployment, inference) expand the attack surface and introduce new trust boundaries across IT and operational environments. MITRE-based AI threat patterns (MITRE ATLAS, complemented by ATT&CK/ICS where relevant) are used to define “must-test” adversarial scenarios and to compare current vs AI-enabled exposure. The outcome is a set of practical strategies and measures for secure data handling, trustworthy AI operation, and cybersecurity safeguards. Clear gating criteria are provided to support evidence-based progression from experimentation to pilots and production-like exploitation for a large metro operator.

Social Media link:

 @nexus-project

For further information please visit nexus-project.eu

LIST OF ABBREVIATIONS AND ACRONYMS

Acronym	Meaning
ABM	Agent-Based Modelling
WP	Work Package
KPI	Key Performance Indicator
AI	Artificial Intelligence
API	Application Programming Interface
ATT&CK	Adversarial Tactics, Techniques, and Common Knowledge (MITRE knowledge base)
ATLAS	Adversarial Threat Landscape for AI Systems (MITRE knowledge base)
CI/CD	Continuous Integration / Continuous Delivery
CMMS	Computerized Maintenance Management System
EAM	Enterprise Asset Management
ETL	Extract, Transform, Load
GoA	Grade of Automation
IAM	Identity and Access Management
ICS	Industrial Control Systems
JIT	Just-In-Time (privilege / access)
KPI	Key Performance Indicator
LLM	Large Language Model
MAE	Mean Absolute Error
MFA	Multi-Factor Authentication
ML	Machine Learning
MLOps	Machine Learning Operations (operationalisation of ML lifecycle)

Acronym	Meaning
NDR	Network Detection & Response
OCC	Operations Control Centre
OT	Operational Technology
PAM	Privileged Access Management
RACI	Responsible, Accountable, Consulted, Informed
RBAC	Role-Based Access Control
RMSE	Root Mean Square Error
SBOM	Software Bill of Materials
SDLC	Software Development Life Cycle
SIEM	Security Information & Event Management
sMAPE	symmetric Mean Absolute Percentage Error
SOC	Security Operations Centre
SLA	Service Level Agreement
TCO	Total Cost of Ownership
TTPs	Tactics, Techniques and Procedures

1 INTRODUCTION

This section will provide background information on the model development activities and AI application work carried out for the NEXUS project to this point. It will also describe the purpose of this deliverable, and set out its detailed objectives, contents to be included, as well as the structure of the report.

1.1 OBJECTIVES

Work Package 7 (WP7) employs scenario-based simulation methods to validate and exploit the tools, concepts and applications developed in the WPs completed in the first year of the NEXUS project. This addresses the requirements of WS1, WS2 and WS3. A comprehensive set of simulation modelling scenarios will be carried out to ensure that all the metro system models, concepts and applications are accurate, consistent, and complete. All the metro system models, concepts and applications will be properly tested and validated to ensure that they reflect the real-world system operations and can be used for decision-making purposes. The specific objectives of this work package are:

- To validate and exploit simulation modelling tools for future metro systems improvement, with a focus on improving crowd management in metro stations and passenger experience in a network.
- To validate and exploit AI models and applications for future metro operations management and planning including exploration activities on the characteristics of future metro train control system as well as strategies and measures for improving data handling and cybersecurity.
- To validate and exploit Optimised Vehicle Concepts using Operational Scenarios, with a focus on improving the interior design of metro vehicles using user-centric design principles.
- To validate and exploit future Metro service improvement (including energy use) in complex networks using scenario analysis, modelling and optimisation techniques, and as a result recommend tangible and workable solutions for a better mass rapid transit system.

These objectives would be achieved by completing four specifically design tasks as follows:

Task 7.1 – Validation and exploitation of Metro network simulation models (led by STAM and supported by AMT, AU, ERTICO, MetroS, TIS, TUW, UNIGE, and VTU).

Task 7.2 – Validation and exploitation of AI applications for future metro operations management and planning (led by VIF and supported by AU, ERTICO, PT, TIS, and UNIGE).

Task 7.3 – Optimised Vehicle Concepts (led by TUW and supported by AMT, AU, ERTICO, MetroS, SIEM, and UNIGE).

Task 7.4 – Metro Service improvement using Optimisation ("To-Be") Scenario Experimentation (led by UNIGE, and supported by AMT, AU, ERTICO, MetroS, STAM, TIS, and VTU).

1.2 SCOPE AND STRUCTURE OF THE REPORT

This section provides a summary of the overall scope and structure of this report.

This technical report is deliverable D7.1 for the NEXUS project, titled “Validation and exploitation of metro system simulation models and AI- Applications”. It describes the modelling activities developed based on Tasks 7.1 and 7.2. Hence, it documents activities undertaken to validate and exploit the functionalities and applicability of metro system simulation models and newly developed AI-applications for a better mass rapid transit system. The primary focus will be evaluating the strengths and weaknesses of the current metro systems (i.e. AMT and MetroS) and the potential impacts of future fully automated operations in dealing with the current passenger demand and some notable disruption scenarios. Further improvements achievable in terms of optimised train vehicle concepts and service optimisation will be undertaken in Tasks 7.3 and 7.4 to be documented in deliverable D7.2.

Following this introductory chapter, the Report will be arranged as follows.

Chapter 2 first provides a clear explanation of the overall evaluation framework employed for Task 7.1. This includes a brief description of the simulation models previously developed in WP4, detailed descriptions of the scenarios to be used for model evaluation, as well as the model validation and cross-validation approach for the metro network and station models. These were developed using the three simulation platforms – AnyLogic, Extend Sim and Simul8. The model validation and cross-validation results would then be presented demonstrating model validities, reliabilities and consistencies. exploitation results, using the validated and cross-validated models, are then presented to highlight the strengths and weaknesses of the current metro systems and the potential impacts of future fully automated operations on AMT in Genoa and MetroS in Sofia.

Chapter 3 provides details on the validation and exploitation of the AI applications developed previously in WP6. It begins with an overview of the overall validation and exploitation approach. Validations of the selected uses cases developed in WP6 are then described. Exemplary energy and CO_2 optimisation results using the TCO tool are then presented and discussed together with their relationships with subsequent Task 8.3. Eventually, the overall metro system cybersecurity strategies would be examined and elaborated.

Chapter 4 summarises the validation and exploitation activities undertaken in Task 7.1 and Task 7.2 and highlights the key study findings and outcomes achieved. The report is concluded by describing the next step forward for optimising the vehicle concept in Task 7.3 and further service optimisations in Task 7.4.

1.3 PARTNERS’ CONTRIBUTIONS

As the WP leader, AU has been coordinating the activities within this WP and has worked collaboratively with other partners to complete the works required for Task 7.1 and Task 7.2, as well as ensuring relevant results will be appropriately articulated to the subsequent Task 7.3, Task 7.4 and WP8. This includes the development of the overall study methodology and metro system evaluation framework using the developed network and station simulation models.

As for the validation and exploitation of all aspects related to simulation modelling of metro stations and networks, STAM, AU, and UNIGE were the key partners in developing, validating, cross-validating and exploiting the metro simulation models. AMT and MetroS (together with VTU) provided all the metro passenger demand and operational data required for the simulation modelling works. The two operators’ also contributed their inputs for model exploitation scenario modelling to ensure that (current

and GoA4) metro system performance evaluations were conducted realistically and consistent with real-world metro operations. TUW and TIS closely followed the works throughout the entire Task 7.1 process, making sure that relevant results and findings will be properly incorporated into their subsequent works for Task 7.3 and WP8 (with support from ERTICO), respectively. VIF, being the leader of Task 7.2, coordinated inputs for the works required in Task 7.2. They specifically conducted an exemplary energy and CO_2 optimisation study enabled by fully automated metro operations (i.e. GoA4), as well as an AI-enabled train uncleanliness detection use case. AU validated a long-term passenger demand forecasting model during network expansion using the MetroS system as a case study. AU also developed a use case for GTFS timetable creation, reviewing the potential benefits achievable focusing on optimising train service timetables. PT provided a dedicated study on the overall metro system cybersecurity strategies. TIS closely followed the works throughout the entire Task 7.2 process, making sure that relevant results and findings will be properly incorporated into their subsequent works for WP8 (with support from ERTICO).

2 VALIDATION AND EXPLOITATION OF METRO SYSTEM SIMULATION MODELS

2.1 THE EVALUATION FRAMEWORK

2.1.1 BRIEF MODEL DESCRIPTIONS

This section provides a brief summary of the network and station models to be validated within Task 7.1, including the purposes and focuses of the models.

2.1.1.1 ANYLOGIC NETWORK MODELS

The AnyLogic network model represents the metro system at a macroscopic level and is designed to simulate train movements and service operations across the entire metro network under real-world conditions. The model adopts an agent-based and discrete-event hybrid logic, where trains are represented as dynamic entities moving along predefined network links connecting successive stations.

The network model explicitly captures key operational characteristics, including train headways, station dwell times, travel times between stations, service frequencies, and rolling stock capacity constraints. These parameters are directly derived from historical operational datasets provided by the operators, ensuring that the simulated behaviour accurately reflects the current ("as-is") state of the metro systems under investigation.

Within WP7, the network model is primarily exploited to validate train movement dynamics and service regularity by comparing simulated outputs against observed data. This includes the assessment of KPIs such as average travel times, headway stability, train saturation levels, and the number of passengers left behind during peak demand periods. The model also enables scenario-based experimentation, allowing controlled variations in service parameters and demand levels to evaluate network performance under peak, off-peak, degraded, or disrupted operating conditions.

The AnyLogic network model provides a robust digital twin component for analysing the strengths and limitations of current metro operations and for supporting exploratory comparisons with future operational concepts. This is possible by reproducing the complex interactions between train scheduling, capacity, and passenger demand at network scale.

2.1.1.2 ANYLOGIC STATION MODELS

The AnyLogic station models focus on a microscopic representation of passenger behaviour and infrastructure usage within individual metro stations. Selected case study stations, including De Ferrari (Genoa) and Serdica (Sofia), are modelled in detail using an agent-based approach, where passengers are represented as autonomous agents interacting with station facilities and train services. The station models explicitly simulate passenger arrival processes, boarding and alighting dynamics, queue formation, platform occupancy, vertical circulation (stairs, escalators, elevators), and inter-line transfer movements. Passenger demand profiles are time-dependent and derived from real-world historical

data, allowing the reproduction of daily demand variability, peak and off-peak conditions, and asymmetric inbound and outbound flows. In the context of WP7, the station models are exploited to validate the ability of the digital twin to replicate observed station-level phenomena, such as congestion hotspots, waiting times, and platform saturation. Simulated KPIs are systematically compared with measured data to assess model accuracy and robustness. In addition, the station models support scenario-based analyses exploring how variations in passenger demand, service patterns, or operational constraints influence station performance and user experience. The AnyLogic station models provide critical insights into the current level of service in “as-is” conditions and establish a validated baseline for comparative assessments of alternative operational scenarios, including future “to-be” concepts with increased levels of automation.

2.1.1.3 SIMUL8 NETWORK MODELS

A set of agent-based discrete event simulation models has been developed, using the Simul8 modelling software package by Aston University (AU), for the metro systems of AMT in Genoa (in Italy) and Metro Sofia (MetroS) in Sofia (in Bulgaria). It includes the following metro network models:

- Two AMT Metro Network Models, one for each direction.
- Two MetroS Network Models, one for each direction.

The aim of the metro network models is to simulate operations of the train entities along the metro network. In this context, the models are developed with a specific focus on improving train entities scheduling to better respond to passenger demands, and thus optimising train entities scheduling and overall travel time of passengers. The ultimate goal is to build network models that are capable of optimising train resources whilst maintaining a reasonably high level of services. This could be achieved by simulating metro network performances based on user-customisable parameters such as train dwell times at stations, departure intervals and timetables.

2.1.1.4 SIMUL8 STATION MODELS

A set of agent-based discrete event simulation models has been developed, using the Simul8 modelling software package by AU, for the two uses cases. It includes the following Metro Station Models:

- One Metro Station Model for the “De Ferrari” Station of AMT.
- One Metro Station Model for the “Serdica” Station of MetroS.

The aim of AU’s Metro station models is to simulate passenger flow dynamics within the two representative stations, De Ferrari Station of AMT and Serdica Station of MetroS. In this context, the models are developed with a specific focus on improving station crowd management through identifying critical bottlenecks (e.g. in-station passenger flow congestion locations) for any possible passenger circulation improvement and optimisations. The key station model processes modelled within the Simul8 environment will be simulating how (inbound / outbound / transfer between platforms) passengers navigate through the stations, including passenger vertical and horizontal movements, their interactions with in-station facilities, as well as train loading / unloading passenger activities.

2.1.1.5 EXTEND SIM NETWORK MODELS

A discrete event simulation model has been developed using the Extend Sim simulation tool and evaluated by UNIGE under several different operational conditions, as already discussed in deliverable D4.1 from NEXUS project. These operational conditions support the preliminary evaluation of the modelling framework developed, justifying the effectiveness of the proposed approach as provided by the outcomes of deliverable D4.2 from NEXUS project. To this end, passenger dynamic demands can be well served while the impacts on the key performance indicators of passenger travel times, their queue lengths, train travel times, and many other parameters, can be observed.

Thereby, the network models have been developed for two metro lines i.e., AMT (Italy) and MetroS (Bulgaria). These two metro lines serve as the accurate representation of the metro network from small and simple to large and complex networks, thus establishing a sound basis for the initial validation of the developed models. Therefore, the network models developed across the Extend Sim platform not only accounts for the modelling of small metro networks (i.e., metro Genova), but also flexible to be extended to larger and more complex metro networks (i.e., metro Sofia).

In summary, the goal of the network model is to provide a close to reality simulation environment to evaluate the current operations of metro networks and identify the key performance indicators to improve the metro operations. To this effect, the goal of the modelling effort is to develop and validate the network models for two use cases (i.e., Genova and Sofia) to provide the ready-to-use modelling environment to be applied for any case study around the world by suitably adjusting different run parameters. A detailed description of different scenarios for the validation of models developed across Extend Sim platform and their cross-validation with the other models developed across different modelling platforms by the other partners, will be discussed in detail in the following sections.

2.1.2 BASELINE SIMULATION ENVIRONMENT CORE PARAMETERS

Table 1: Metro Network Model Core Baseline Parameters

Parameter Category	Specific Parameters	AMT	MetroS
Network Topology	Number of Stations	8	47
	Number of Lines	1	4
Operational Parameters	Train Capacity	Old: 420 pax/train New: 290 pax/train	1226 pax/train (for L1, L2 &L4) 785 pax/train (for L3)
	Service Timetable	Derived from GTFS data (with minimum headway of 6 mins)	Official daily timetable (with minimum headway of 3 mins)
	Platform Capacity	200 passengers	7200 passengers
	Dwell Time	22 seconds (Fixed)	10 to 35 seconds (Fixed)
Passenger Flow	Baseline OD Demand	40,350 passenger trips / day	484,640 passenger trips / day

Parameter Category	Specific Parameters	AMT	MetroS
	Peak Hour Percentage	19%-22% of daily total	30% of daily total

The specific parameters of the constructed high-fidelity simulation environment are summarised in **Table 1** and **Table 2**. These baseline parameter values were collected from and confirmed by the respective metro operators AMT and MetroS. Apart from these core parameters, some additional parameters such as passenger boarding / alighting time per train door and passenger vertical movement facility operational parameters (i.e. capacity, travel time, etc.) were determined according to relevant literature (details refer to D4.2). These core sets of input parameters are served as the common baseline conditions and are incorporated into the models in the format appropriate to the three adopted simulation platforms AnyLogic, Extend Sim and Simul8 to carry out model validations (for individual simulation platform) and cross-validations across different simulation platforms.

Table 2: Metro Station Model Core Baseline Parameters

Parameter Category	Specific Parameters	AMT	MetroS
Location	Station Name	De Ferrari Station	Serdica Station
Layout	Number of Levels	4 (G/F to -3/F)	5 (G/F to -4/F)
	Number of Entrances / Exits	2	12
	Number of Platforms	2	4
	Platform Capacity	500 passengers	1500 passengers
Operational Parameters	Train Departure Schedules	Derived from GTFS data (with minimum headway of 6 mins)	Official daily timetable (with minimum headway of 3 mins)
	Dwell Time	22 seconds (Fixed)	35 seconds (Fixed)
	Number of Lines Served	1	3 (L1, L2 and L4)
	Train Capacity	Old: 420 pax/train New: 290 pax/train	1226 pax/train
	Number of Train Doors	Old: 3 per car New: 3 per car	5 per car for 1st and last carriage; 6 per car for intermediate carriage
Passenger Flow	Baseline OD Demand	Approximately 11,000 passenger visits / day	Approximately 80,000 passenger visits / day
	Peak Hour Percentage	15%-16% of daily total	18% of daily total

2.1.3 SCENARIO MODELLING

This section describes the set of operational scenarios evaluated in Task 7.1, using the metro network and station simulation models developed during the first year of the project and described in **Section 2.1.1**. The scenarios are designed to systematically exploit the WP4 digital twin framework to assess metro system performance under a range of demand levels, operational conditions, and levels of automation.

All scenarios are analysed under two complementary perspectives:

- **"As-is" scenarios**, representing current metro operations with automation levels below GoA4, reflecting existing operational constraints and human-driven control strategies.
- **"To-be" scenarios**, assuming GoA4 (fully automated) metro systems, enabling the exploration of enhanced operational flexibility, responsiveness, and resilience.

Across all scenarios, a harmonised set of model parameters and performance indicators is applied, allowing consistent comparison between scenarios, automation levels, and use cases.

Table 3: Brief summary of scenarios

	Focus area	Scenario	Key Interventions & Model Parameters
1	Service optimisation - Peak	Peak-hour demand management	**Timetable:** headway **Fleet:** Number of trains in rotation **Station control:** dwell time, max. platform capacity, boarding / alighting time **Service pattern** all stops
2	Service optimisation - off Peak	off peak-hour demand management	**Timetable:** headway **Fleet:** Number of trains in rotation **Station control:** dwell time, max. platform capacity, boarding / alighting time **Service pattern:** all stops
3	System resilience	Extraordinary events and service disruption	**Service recovery:** contingency headway, short turning location. **Resource allocation:** rolling stock redeployment time **Passenger info:** alternative route adoption/passenger compliance rate.
4	Service optimisation – Peak (GoA4)	Peak-hour demand management with shorter train headways (120-90 sec)	**Timetable:** 120 and 90 seconds headway **Fleet:** Number of trains in rotation **Station control:** dwell time, max. platform capacity, boarding / alighting time **Service pattern:** all stops
5	System resilience (GoA4)	Extraordinary events and service disruption with shorter train	**Service recovery:** 120 and 90 seconds headway, short turning location.

Focus area	Scenario	Key Interventions & Model Parameters
	headways (120-90 sec)	**Resource allocation:** rolling stock redeployment time **Passenger info:** alternative route adoption/passenger compliance rate.

The analysis in the GoA4 scenarios presented here focuses on variations in train headways to evaluate system performance. Dynamic dwell times, which adapt station dwell times based on passenger boarding and alighting, are recognised as a promising strategy for improving service regularity and capacity utilisation further. However, their quantitative assessment is reserved for the subsequent optimisation phase, enabling future analyses to explore their full impact under fully automated operations.

2.1.3.1 SCENARIO 1 – SERVICE OPTIMISATION UNDER PEAK DEMAND

Scenario 1 focuses on the assessment of metro system performance under peak-demand conditions, representing ***current operations (as-is)*** and serving as a reference scenario for subsequent analyses. The scenario is fully based on the Baseline Scenario defined in **Section 2.1.2**, which simulates a typical day of operation using harmonised model parameters and validated demand inputs.

This analysis concentrates on peak-hour periods, during which passenger demand reaches its highest levels and operational stress is maximised. For both use cases, the simulation covers the morning peak between 07:00 and 10:00. In the afternoon, the analysis considers a peak period between 16:00 and 20:00 for the AMT use case, while for the MetroS use case it explicitly accounts for the observed PM peak occurring between 17:30 and 19:30.

No changes are applied to model settings with respect to the baseline configuration. Network topology, rolling stock characteristics, service timetables, and passenger behaviour assumptions remain unchanged, ensuring that the scenario reflects realistic current operations. Each model run generates performance outputs according to the harmonised set of KPIs, including indicators related to passenger waiting times, train saturation, congestion levels, passengers left behind, and overall service performance during peak conditions.

Scenario 1 provides the benchmark against which the impacts of alternative operational strategies and higher levels of automation are evaluated in later scenarios.

2.1.3.2 SCENARIO 2 – SERVICE OPTIMISATION UNDER OFF-PEAK DEMAND

Scenario 2 evaluates metro system performance during off-peak periods under ***current operating conditions (as-is),*** with the objective of assessing service efficiency and quality when passenger demand is lower and more variable. As in Scenario 1, the scenario is based entirely on the Baseline Scenario defined in **Section 2.1.2**, with no modifications to model parameters or operational assumptions.

The analysis focuses on three distinct off-peak time windows: early morning (05:00–07:00), inter-peak (10:00–16:00), and late afternoon/evening (20:00–24:00). These periods capture different demand

profiles and operational contexts, allowing the evaluation of system behaviour across a broad range of low- and medium-demand conditions.

All simulations use the same baseline configuration in terms of infrastructure, rolling stock, service timetables, and passenger behaviour. Model outputs are generated in accordance with the same set of KPIs as described for Scenario 1, with particular attention to passenger waiting times, train utilisation, congestion levels, passengers left behind, and energy consumption.

Scenario 2 complements the peak-demand analysis by providing insight into system performance during off-peak operations and serves as a reference for assessing the potential benefits of more flexible and automated service management strategies in later scenarios.

2.1.3.3 SCENARIO 3 – SYSTEM RESILIENCE UNDER DISRUPTED CONDITIONS

Scenario 3 is designed to assess the resilience of the metro systems under disrupted operating conditions, focusing on their ability to absorb shocks, manage abnormal demand patterns, and recover acceptable levels of service within a reasonable timeframe. The scenario is analysed under the ***as-is perspective***, representing current operations with automation levels below GoA4 and existing operational constraints.

The scenario configuration is based on the Baseline Scenario defined in **Section 2.1.2**, with identical network topology, rolling stock characteristics, service timetables, and passenger behavioural assumptions. Disruptions are introduced by modifying demand profiles and/or operational availability, depending on the specific use case, while keeping all other parameters unchanged. This approach allows the direct attribution of performance degradation and recovery dynamics to the disruption itself.

AMT Use Case – Demand-Driven Service Disruptions

For the AMT metro system, Scenario 3 focuses on demand-driven disruptions caused by extraordinary events that generate sudden, spatially concentrated, and temporally limited passenger flows. Two disrupted sub-scenarios are analysed:

- **Match-day scenario**, representing major football events in Genoa, with abnormal passenger arrivals concentrated around the Brignole station, serving the stadium area and within specific time windows before and after the event.
- **Public event / protest scenario**, representing unplanned or semi-planned social events that cause irregular demand surges, uneven origin–destination patterns, and temporary platform overcrowding.

In both cases, disruptions are modelled by increasing passenger arrival rates during the affected periods and altering OD matrices to reflect event-related travel behaviour. The OD information is to be provided by AMT, according to their experiences and understanding of similar events previously observed, which is believed to be reasonably representative of the actual conditions. No proactive timetable changes are assumed, reflecting realistic as-is operational conditions where response capabilities are limited by human-driven control and predefined schedules. ***The scenario evaluates how quickly congestion propagates through the network and how effectively the system can return to normal operating conditions once demand normalises.***

MetroS Use Case – Infrastructure and Operational Disruptions

For the MetroS system, Scenario 3 focuses on operational and infrastructure-related disruptions. As informed by MetroS, according to their experiences and local understanding, the disrupted scenario assumes an overloaded OD demand combined with partial station unavailability, specifically:

- Closure of Serdica station for three hours during the PM peak period (16:00–19:00).

To properly capture recovery dynamics, the simulation time horizon is extended beyond the disruption window, covering approximately 15:00 to 24:00. This enables the analysis of residual congestion effects, passenger redistribution across alternative stations, and the time required for the system to stabilise after service restoration.

The station closure is represented in the model through the suspension of boarding and alighting activities at Serdica station, rerouting of passenger flows, increased dwell times and congestion at neighbouring stations, and temporary capacity reductions at system level.

Performance Assessment

Across both use cases, performance is evaluated using the set harmonised KPI set as described for Scenarios 1 and 2, including average waiting times, number of passengers left behind, station overcrowding events, congestion levels, and recovery time indicators. ***The results provide a consistent basis for cross-validation between AMT and MetroS and establish a reference point for comparison with the corresponding GoA4 disruption scenarios analysed in Scenario 5.***

2.1.3.4 SCENARIO 4 – SYSTEM ELASTICITY AND CAPACITY SATURATION (GOA4)

Scenario 4 focuses on the evaluation of metro system performance under peak-demand conditions assuming fully automated operations (GoA4). The scenario represents the ***to-be operational perspective*** and aims to quantify the potential benefits of automation in managing high passenger demand, improving service regularity, and enhancing overall system efficiency.

The scenario is directly derived from Scenario 1, adopting the same peak-demand profiles, time windows, and network configurations for both AMT and MetroS. This ensures full comparability between current operations (as-is) and fully automated operations (to-be), isolating the effects of automation from those related to demand or infrastructure changes. All infrastructure-related parameters, rolling stock characteristics, and passenger behavioural assumptions remain unchanged with respect to the baseline.

Under GoA4 assumptions, operational constraints associated with human-driven control are removed, allowing the system to operate with reduced train headways and more responsive station management. Train headways are progressively reduced to 120 and, 90 seconds to explore different levels of service intensity under peak conditions.

This analysis allows the assessment of how fully automated systems can better absorb peak demand, reduce platform congestion, and minimise the number of passengers left behind. Performance is evaluated using the same harmonised KPI set as described for Scenarios 1, with particular emphasis

on passenger waiting times, train saturation levels, congestion at stations, service regularity, and energy consumption.

The results of Scenario 4 provide a structured basis for identifying optimal GoA4 operational configurations under peak demand and support evidence-based discussions on the capacity gains and operational flexibility enabled by full automation when compared to current metro operations.

2.1.3.5 SCENARIO 5 – SERVICE DISRUPTIONS UNDER FULLY AUTOATED OPERATIONS (GOA4)

Scenario 5 extends the analysis of disrupted operating conditions introduced in Scenario 3 by assuming fully automated metro operations (GoA4). The scenario represents the ***to-be perspective*** and aims to assess how automation can improve system resilience, responsiveness, and recovery capabilities when service disruptions occur under high operational stress.

The scenario adopts the same disruption types and demand conditions defined in Scenario 3 for both AMT and MetroS, ensuring direct comparability between current (as-is) and fully automated (to-be) operations. Network topology, infrastructure characteristics, rolling stock parameters, and passenger behavioural assumptions remain unchanged with respect to the baseline. Differences in system performance are therefore attributable exclusively to the introduction of GoA4 automation and associated control strategies.

Under GoA4 assumptions, the system benefits from enhanced operational flexibility, particularly in terms of headway management and station control. Train headways are reduced to 120 and 90 seconds, enabling a more aggressive and adaptive response to demand surges and service disruptions.

For the AMT use case, the scenario evaluates how fully automated operations can mitigate the impacts of extraordinary, demand-driven disruptions such as match days and public events, reducing platform congestion and accelerating the recovery of regular service levels once abnormal demand subsides. For the MetroS use case, the scenario focuses on operational and infrastructure-related disruptions, including temporary station unavailability, with particular attention to how automation supports faster passenger redistribution, more stable headway regulation, and reduced recovery times following service restoration.

Performance is assessed using the same harmonised KPI set as described for Scenario 3, with emphasis on recovery-related indicators such as average waiting times during and after the disruption, number of passengers left behind, station overcrowding events, congestion propagation, and overall system stabilisation time. ***By comparing Scenario 5 with Scenario 3, the analysis provides quantitative evidence of the potential benefits of GoA4 automation in enhancing metro system robustness and resilience under disrupted conditions.***

2.1.4 MODEL VALIDATION AND EXPLOITATION APPROACH

The general approach for validating, cross-validating, and exploiting the developed metro network and station models are outlined in **Figure 1**. The overall process starts by compiling a common set of model input parameters and passenger demand information, according to the scenario defined in **Section**

2.1.2. In particular, the model validation and cross-validation exercises would be based on the conditions defined in the Baseline Scenario, simulating a typical day of metro operation with standard passenger demand information. It is important to emphasise that the baseline conditions were derived in close coordination with the two metro operators, making sure that the input operational parameters and passenger demand information are representative of the real-world metro systems to the extent possible.

Model validations

Both the metro network and station models developed using AnyLogic, Extend Sim and Simul8 will be validated on the basis of the baseline conditions described in **Section 2.1.2** where modelled outputs from each individual modelling platform will be compared with real-world observable metrics from the two use cases. The aim is to demonstrate that each (network and station) model will be able to replicate the real-world metro systems in terms of providing the best possible fit between the observed and modelled metrics. As outputs generated from different platforms vary, different platforms will adopt a slightly different set of output metrics for individual model validation, as summarised in **Table 4**.

In line with common model validation practices, these validation output metrics are normally key standard high-level aggregated metro system operational statistics. Also, during the model calibration and validation stage, different modelling platforms might offer different additional functionalities to aid the fine-tuning of model parameters so as to provide the best fit possible. More details on the validation process for individual simulation platform will be provided later in **Section 2.2**.

Table 4: Summary of Adopted Output Metrics for Model Validations

Simulation Platform	Model Type	Output Metrics for Model Validations
AnyLogic	Network Models	Total Number of Trains Operated, Train Overall Journey Time, Passenger Inflows and Outflows.
	Station Models	Passenger Flow at Key Check Points (e.g. entrances/exits, platforms, concourse areas, inbounds, outbounds, transfers, etc.)
Simul8	Network Models	Total Number of Trains Operated, Train Overall Journey Time
	Station Models	Passenger Flow at Key Check Points (e.g. entrances/exits, platforms, concourse areas, inbounds, outbounds, transfers, etc.)
Extend Sim	Network Models	Total Number of Trains Operated, Train Overall Journey Time, Passenger Inflows and Outflows.

Model cross-validations

Model cross-validations will seek to achieve consistent and comparable results across the three modelling platforms. Again, based upon the baseline conditions described in **Section 2.1.2**, outputs generated from the individually validated models using each modelling platform would be compared with one another (i.e. among network models developed using AnyLogic, Extend Sim and Simul8;

between station models developed using AnyLogic and Simul8), with an aim to achieve the best possible fit. The comparison will be based on some similar output metrics, which primarily include total number of services operated and train overall journey time for network models, and passenger flow output statistics at key check points for station models.

Model exploitations

At the model exploitation stage, the validated and cross-validated models will be used to evaluate the potential impacts of various scenarios simulating the current metro systems and future fully automated operations, according to the scenarios described in **Section 2.1.3** where the followings will be undertaken for each scenario to be evaluated:

- Apply changes onto the validated and cross-validated baseline models.
- Carry out model runs.
- Compare model results across scenarios to evaluate performance of the current metro systems and potential impacts of future fully automated operations.

The evaluation will be carried out on the basis of an extensive range of KPIs, enabling a multi-perspective, multi-purpose, multi-level model assessments to identify the strengths and weaknesses of the current metro systems, as well as testing the extent of improvements possible should the future metro system is to be operated at full automation.

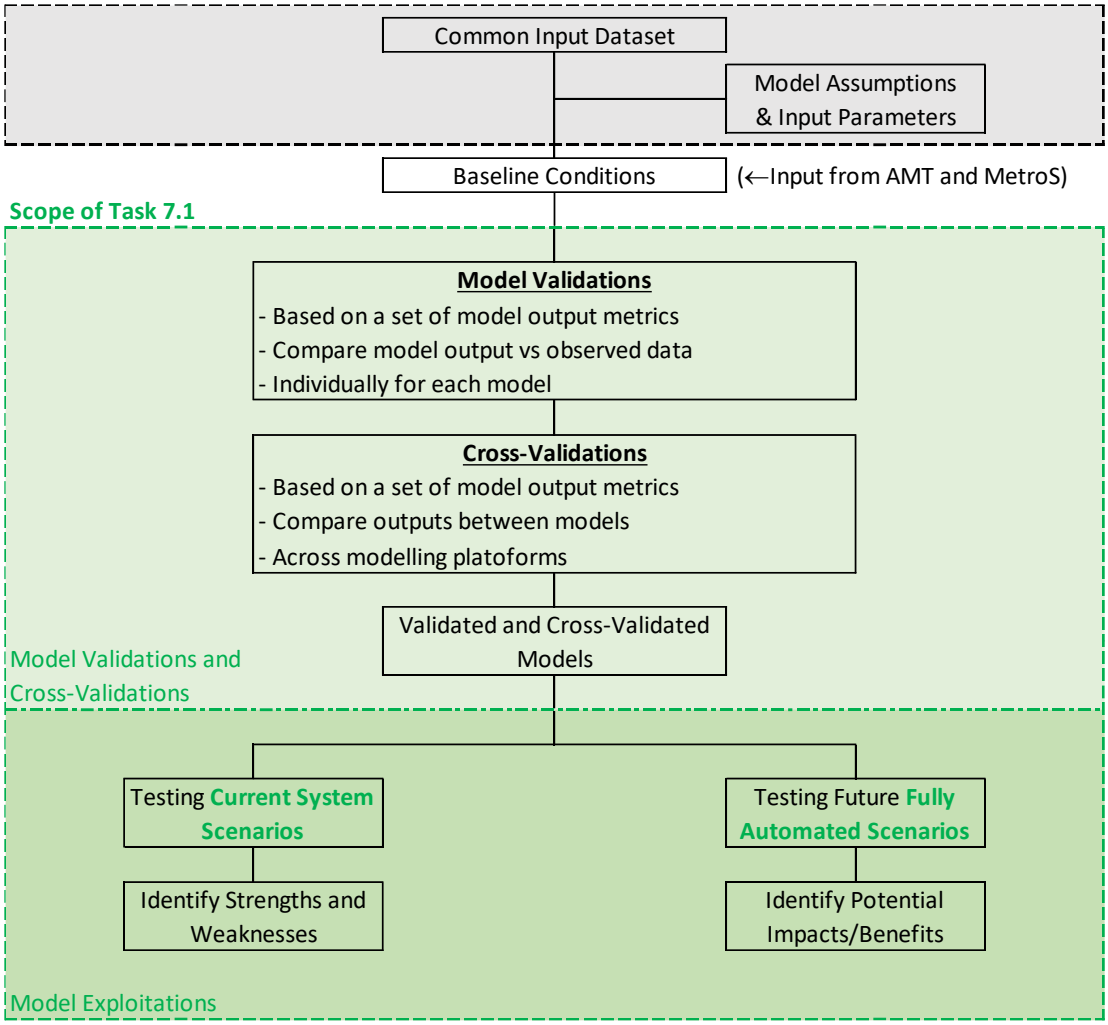


Figure 1: Model Validations, Cross-Validations, and Exploitations Approach for WP7

2.2 MODEL VALIDATIONS AND CROSS-VALIDATIONS RESULTS

2.2.1 NETWORK MODEL VALIDATIONS AND CROSS-VALIDATIONS

This section outlines the validation and cross-validation processes for the network models using AnyLogic, Extend Sim and Simul8, in accordance with the framework in **Section 2.1**. Models run were carried out using standard passenger demand information and operational conditions as defined in **Section 2.1.2** concerning the simulation environment core parameters. The network model validation and cross-validation exercises will focus on the followings:

- Model Validations Results – for each network model, comparing model outputs against real-world observable metrics as defined in **Table 4**.

- Model Cross-Validations Results – comparing compatible network model outputs between AnyLogic, Extend Sim and Simul8.

2.2.1.1 VALIDATION RESULTS OF ANYLOGIC NETWORK MODEL

2.2.1.1.1 STRUCTURAL AND BEHAVIOURAL VALIDATION

Prior to quantitative validation, the AnyLogic network models for the AMT and MetroS metro systems were thoroughly verified to ensure their structures, logic and parameterisation accurately reproduced real-world operations.

The verification process consisted of:

- a conceptual-to-implementation consistency check: cross-checking all input parameters (headways, dwell times, inter-station travel times, rolling stock capacity, and service frequencies) against the historical datasets provided by the operator and used to represent the baseline ("as-is") scenario. - The implemented network topology and directional service patterns were verified against the conceptual network specification.
- Output trace consistency checks: key state variables (train arrival/departure times, passenger boarding/alighting counts and vehicle loads) were monitored during simulation runs to confirm that operational rules were applied consistently across the entire network.

These checks confirmed that the AMT and MetroS models has reproduced realistic train movement dynamics and station operations, allowing to proceed to quantitative validation.

2.2.1.1.2 AMT NETWORK MODEL (ANYLOGIC)

Total Number of Train Services and Journey Times

Table 5 summarises the comparison between the total number of completed train services and end-to-end journey times, as simulated and observed.

Table 5: Observed and simulated AMT total number of services and journey times (AnyLogic)

BRIN TO BRIGNOLE DIRECTION						
	Observed		**Simulated**		**(Observed – Simulated)**	
Journey Time	18.34 min	1101 sec	18.35 min	1101 sec	0 min	0 sec
Number of Trains	145		145		0	
BRIGNOLE TO BRIN DIRECTION						
	Observed		**Simulated**		**(Observed – Simulated)**	
Journey Time	16.35 min	981 sec	18.92 min	981 sec	0 min	0 sec
Number of Trains	146		146		0	

The results show a perfect match between the number of simulated and observed train services in both directions. End-to-end journey times also align perfectly with observed values (zero-second deviation).

These results demonstrate that the AnyLogic model accurately reproduces train circulation dynamics and timetable adherence for the AMT baseline configuration.

Passenger Flow Validation at Station Level

In addition to validating train movements, the AMT model was validated against observed passenger boarding (In) and alighting (Out) counts at station level.

This comparison revealed minimal discrepancies between the simulated and observed passenger flows at all stations (**Table 6**). The percentage differences remain within a narrow range (approximately ±3%), which confirms that:

- passenger demand allocation across stations is correctly represented;
- boarding and alighting processes are consistently reproduced;
- network-wide flow conservation is maintained.

Table 6: Observed and simulated AMT network model total inflows and outflows (AnyLogic).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
BRIN	6779	6866	1,28%	7040	7248	2,95%
DINEGRO	5912	6056	2,44%	5088	5004	-1,65%
PRINCIPE	5265	5172	-1,77%	4303	4227	-1,77%
DARSENA	3237	3149	-2,72%	3395	3375	-0,59%
S. GIORGIO	3850	3837	-0,34%	4364	4298	-1,51%
SARZANO	1911	1932	1,10%	2078	2063	-0,72%
DE FERRARI	4523	4511	-0,27%	6204	6099	-1,69%
BRIGNOLE	8877	8906	0,33%	7882	8110	2,89%
Total	**40354**	**40235**	-0,29%	**40354**	**40235**	-0,29%

2.2.1.1.3 METROS NETWORK MODEL (ANYLOGIC)

Total Number of Train Services and Journey Times

Table 7 summarises the comparison between the observed and simulated total number of train services and end-to-end journey times for each line and direction.

The simulated journey times across all lines show extremely small deviations from observed values, ranging between 1 and 3 seconds. These negligible differences confirm that the model accurately captures:

- inter-station travel times,
- dwell time logic,
- line-specific operational characteristics.

Small discrepancies in the total number of trains (±1–3 units on selected directions) are explained by the simulation time window.

The AnyLogic model simulates operations from 05:00 to 24:00, whereas the observed dataset includes a small number of services occurring before 05:00 (early departures), shortly after midnight.

Table 7: Observed and simulated MetroS total number of services and journey times (AnyLogic)

LINE 1 TO SOFIA BUSINESS PARK DIRECTION						
	Observed		**Simulated**		**(Observed – Simulated)**	
Journey Time	32.42 mins	1945 sec	32,38 mins	1943 sec	0.04 min	2 sec
Number of Trains	132		131		1	
LINE 1 TO SLIVNITSA DIRECTION						
	Observed		**Simulated**		**(Observed – Simulated)**	
Journey Time	32.42 mins	1945 sec	32.38 mins	1943 sec	0.04 min	2 sec
Number of Trains	130		130		0	
LINE 2 TO VITOSHA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.42 min	1465 sec	24.40 min	1464 sec	0.02 min	1 sec
Number of Trains	141		140		1	
LINE 2 TO OBELYA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.42 min	1465 sec	24.40 min	1464 sec	0.02 min	1 sec
Number of Trains	139		139		0	
LINE 3 TO HADZHI DIMITAR DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.17 min	1450 sec	24.147 min	1449 sec	0.023 min	1 sec
Number of Trains	191		191		0	
LINE 3 TO GORNA BANYA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.17 min	1450 sec	24.147 min	1449 sec	0.023 min	1 sec
Number of Trains	192		190		2	
LINE 4 TO SOFIA AIRPORT DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	44.42 min	2665 sec	44.374 min	2663 sec	0.046 min	3 sec
Number of Trains	133		130		3	
LINE 4 TO OBELYA DIRECTION						

	Observed		Simulated		Observed - Simulated	
Journey Time	44.42 min	2665 sec	44.374 min	2663 sec	0.046 min	3 sec
Number of Trains	130		130		0	

Specifically, the observed counts include one early Line 3 (Green line) train, one early Line 4 (Yellow line) train, and one late Line 2 (Blue line) train after midnight, but fall outside the simulated time horizon. Therefore, the minor differences in train totals are not due to modelling inaccuracies but to the deliberate definition of the simulation time frame.

Passenger Flow Validation at Station Level

Inbound Validation

The inbound passenger flows show a strong agreement between observed and simulated values across all four lines. At the aggregate level, the total difference between observed and simulated entries remains well within acceptable margins: 0.61% for the Line 4, 0.48% for the Line 2, 0.10% for the Line 3, and 0.86% for the Line 1. These figures confirm that the boarding demand, as derived from the origin-destination matrix and the temporal delta distribution, is accurately reproduced at the network level.

At the individual station level, the per-station percentage differences are small and distributed symmetrically around zero, with no systematic over- or under-estimation in any consistent direction across stations or lines. This balance indicates that the discrepancies are stochastic in nature rather than the result of a structural miscalibration, and that the boarding logic correctly reflects the spatial and temporal distribution of demand as encoded in the input data.

Outbound Validation

The outbound passenger validation reveals a systematic discrepancy at the individual station level that is not present in the aggregate totals, where the overall difference between observed and simulated exits is 0.86% for the Line 1, 0.48% for the Line 2, and 0.10% for the Line 3, all within acceptable bounds. The Line 4 records a larger aggregate deviation of -1.29%. While these network-level totals remain relatively contained, the per-station distribution shows substantially higher variance than the inbound case, with individual stations recording differences well beyond the aggregate figures in both directions.

This behaviour originates in the interaction between aggregate hourly demand data and train-level flow execution.

Alighting demand is derived from the origin-destination matrix combined with a temporal delta matrix that distributes daily flows across hourly bands. However, while the input data is specified at hourly granularity, the simulation executes boarding and alighting flows at the level of individual train passages. At each station stop, the number of passengers to be discharged is computed as the hourly alighting demand weighted by the time band share and divided by the train frequency in that band.

Because hourly demand is distributed across individual services, the expected alighting quota per train is determined proportionally within the time band. However, the actual number of passengers on board a specific train at a given stop depends on the cumulative boarding and alighting dynamics along the

line. When the calculated alighting quota exceeds the instantaneous onboard occupancy, the full expected discharge cannot be executed at that station. The residual onboard passengers remain in the system and are subsequently discharged at downstream stops, with terminal stations acting as the final balancing nodes.

Table 8: Observed and simulated inflows and outflows for Line 1 of MetroS (AnyLogic).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Slivnitsa	6971	7059	-1.26%	6143	8262	-34,49%
Lyulin	9872	9939	-0,67%	7881	6137	22,13%
Zapaden Park	6710	6605	1,57%	5677	5292	6,79%
Vardar	6881	6700	2,64%	6720	6447	4,06%
Konstantin Velichkov	9432	9310	1,30%	8619	8438	2,10%
Opalchenska	11971	11876	0,79%	11744	11586	1,35%
Serdica	71247	70560	0,96%	74266	73889	0,51%
SU Sv. Kliment Ohridski	16185	16011	1,08%	16185	16016	1,05%
Vasil Levski Stadium	10348	10241	1,03%	11851	11669	1,54%
F. Joliot-Curie	6306	6219	1,38%	8527	8362	1,94%
G.M. Dimitrov	13409	13255	1,15%	10856	10590	2,45%
Musagenitsa	3120	3063	1,84%	3595	3362	6,49%
Mladost 1	6769	6735	0,51%	7751	7350	5,17%
Aleksandar Malinov	5409	5404	0,09%	5031	4603	8,51%
Akad. Al. Teodorov - Balan	8305	8286	0,23%	8493	7412	12,73%
Business Park	11868	11786	0,69%	11465	13634	-18,92%
Total	204808	203049	0,86%	204808	203049	0,86%

Table 9: Observed and simulated inflows and outflows for Line 2 of MetroS (AnyLogic).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Obelya	6969	7061	-1,32%	6369	11665	-83,15%
Lomsko Shose	2339	2230	4,66%	1974	1415	28,32%
Beli Dunav	10839	10562	2,56%	8911	5982	32,87%
Nadezhda	8608	8415	2,24%	7842	6853	12,61%
Khan Kubrat	5824	5778	0,79%	6359	5993	5,76%
Princess Marie Louise	5691	5793	-1,79%	5936	5689	4,16%
Central Railway Station	5434	5437	-0,06%	8979	8766	2,37%
Lavov most	7317	7329	-0,16%	7516	7337	2,38%
Serdika	127839	127462	0,30%	117658	116645	0,86%
National Palace of Culture (NDK)	86432	86096	0,39%	99782	98949	0,83%

European Union	10177	10230	-0,52%	8154	7879	3,37%
James Bourchier	10891	10661	2,11%	10796	8935	17,24%
Vitosha	18547	18384	0,88%	16631	19330	-16,23%
Total	306907	305438	0,48%	306907	305438	0,48%

Table 10: Observed and simulated inflows and outflows for Line 3 of MetroS (AnyLogic).

Station Name	In			Out		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Gorna Banya	4063	4103	-0,98%	4058	7011	-72,77%
Ovcha Kupel	3111	3125	-0,45%	2673	2038	23,76%
Moesia	9989	10024	-0,35%	7577	6740	11,05%
Ovcha Kupel	4279	4408	-3,01%	4166	3836	7,92%
Krasno selo	12130	12180	-0,41%	18099	17600	2,76%
Bulgaria	8580	8658	-0,91%	12324	12071	2,05%
Medical University	7968	7996	-0,35%	9898	9656	2,44%
National Palace of Culture (NDK)	83892	83755	0,16%	70224	68997	1,75%
Sveti Patriarh Evtimiy	5804	5699	1,81%	8656	8492	1,89%
Orlov Most	7616	7509	1,40%	7938	7660	3,50%
Teatralna	6962	6890	1,03%	7834	7451	4,89%
Hadzhi Dimitar	10957	10839	1,08%	11904	13634	-14,53%
Total	165351	165186	0,10%	165351	165186	0,10%

Table 11: Observed and simulated inflows and outflows for Line 4 of MetroS (AnyLogic).

Station Name	In			Out		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Obelya	9448	9375	0,77%	8779	13624	-55,18%
Slivnitsa	7919	7903	0,20%	6430	7903	-22,91%
Lyulin	10566	10615	-0,46%	8626	10615	-23,06%
Zapaden park	6652	6546	1,59%	6076	6546	-7,74%
Vardar	7102	7169	-0,94%	7117	7169	-0,73%
Konstantin Velichkov	10132	9992	1,38%	8622	9992	-15,89%
Opalchenska	11630	11251	3,25%	11501	11251	2,17%
Serdika	72619	71976	0,89%	78216	71976	7,98%
Sofia University St Kliment Ohridski	17451	17262	1,08%	13949	17262	-23,76%
Vasil Levski Stadium	10246	10124	1,19%	11635	10124	12,98%
Joliot-Curie	5770	5728	0,73%	9014	5728	36,45%
G. M. Dimitrov	13779	13750	0,21%	11168	13750	-23,13%
Musagenitsa	3381	3442	-1,82%	3556	3442	3,19%
Mladost 1	7256	7341	-1,17%	7629	7341	3,77%
Mladost 3	5363	5301	1,16%	4888	5301	-8,45%

Inter Expo Center - Tsarigradsko shose	9712	9794	-0,84%	9092	9794	-7,72%
Druzhba	5948	6026	-1,31%	7167	6026	15,92%
Iskarsko shose	4439	4444	-0,11%	4847	4444	8,31%
Sofiyska Sveta gora	908	893	1,65%	801	893	-11,49%
Sofia Airport	3119	3142	-0,74%	4329	3142	27,42%
Total	223438	222074	0,61%	223438	226323	-1,29%

The practical effect is that exits are under-recorded at intermediate stations where demand exceeds instantaneous train occupancy, and over-recorded at terminal stations. Since all passengers are ultimately discharged, the network-level total remains accurate, while the per-station distribution reflects the interaction between hourly demand allocation and train-level operational dynamics.

Overall Assessment

Taken together, the validation results confirm that the MetroS AnyLogic network model reproduces observed passenger flows with good accuracy at the network level. The inbound mechanism is well-calibrated across all lines and stations. The outbound discrepancies, while pronounced at the individual station level, are structurally explained by the granularity mismatch between input data and simulation execution, and do not represent a fundamental miscalibration of demand. The model is considered suitable for the analysis of the scenarios described in this study.

2.2.1.2 VALIDATION RESULTS OF SIMUL8 NETWORK MODEL

The Simul8 network model validation exercise is to compare the observed and simulated number of train services completed and train journey times. To do this, the Simul8 model structure and logics were first checked thoroughly to ensure that they were built as expected to replicate the real-world operation of the two metro systems, using a three-layer checking process. First, a structured walkthrough was undertaken to check that the correct input parameters, data, assumptions, and the logics have been properly and accurately encoded. The model setup was checked against the conceptual model specifications to make sure that all elements were built consistently with the specifications. Secondly, the run-time animations were examined to visually validate model behaviours. Model runs were conducted slowly (with on-screen animations) to visually examine how each element of the model behaved. A screenshot for the AMT network model is presented in **Figure 2** as an example. This was done repeatedly to identify and rectify any inappropriate or inaccurate model setups. A trace analysis was also carried out using the “Monitor Simulation” feature in Simul8 where each single “Work Items” in the model could be traced for their movements, routes and actions taken within the simulation system. This verified every process of the model was run properly as expected.

The above-mentioned three-layer checking process has confirmed that the Simul8 model runs have performed very well and the on-screen animations have also shown logical movements of the simulated trains and passengers. This reflects that, visually, the models could satisfactorily replicate the real-world operations of the two metro systems. Statistical outputs were then extracted from the models to further validate against corresponding observed metrics.

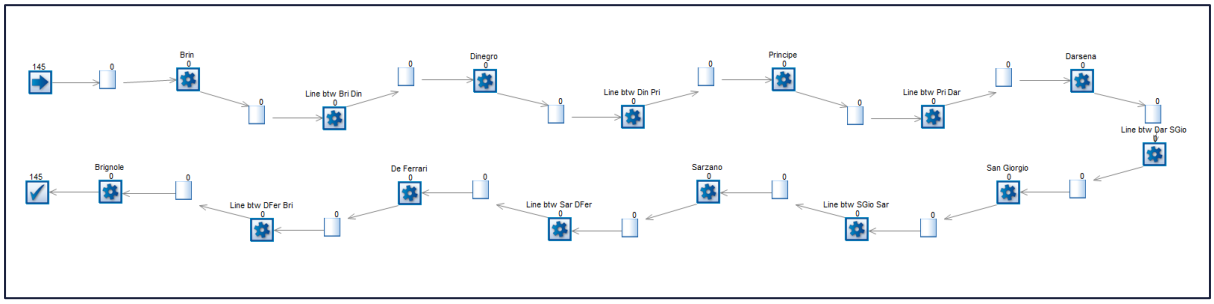


Figure 2 – Example Metro Network Model for AMT (Brin to Brignole direction) using Simul8.

2.2.1.2.1 AMT NETWORK MODEL (SIMUL8)

Total Number of Train Services and Journey Times

The comparison between the simulated and observed total number of train services and their journey times to complete the service for the AMT metro network are summarised in **Table 12**. It shows that the simulated number of train services simulated matched exactly with the observed number of train services. The simulated train journey times were also close to within 1 second of the observed (i.e. actual) train journey times.

These validation results demonstrate that the developed models could reliably replicate the train services for the AMT metro network, in terms of the total number of completed train services, and the train journey time.

Table 12: Observed and simulated AMT total number of services and journey times (Simul8)

BRIN TO BRIGNOLE DIRECTION						
	Observed		**Simulated**		**Observed - Simulated**	
Journey Time	18.34 min	1101 sec	18.35 min	1102 sec	-0.01 min	-1 sec
Number of Trains	145		145		0	
BRIGNOLE TO BRIN DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	16.35 min	981 sec	16.38 min	982 sec	-0.03 min	-1 sec
Number of Trains	146		146		0	

2.2.1.2.2 METROS NETWORK MODEL (SIMUL8)

Total Number of Train Services and Journey Times

The comparison between the simulated and observed total number of train services and their journey times to complete the service for the MetroS metro network are summarised in **Table 13**. It shows that the simulated number of train services matched exactly with the observed number of train services. The simulated train journey times were also very close to the observed train journey times.

These validation results demonstrate that the developed models could reliably replicate the train services for the MetroS metro network, in terms of the total number of completed train services and the train journey time.

Table 13: Observed and simulated MetroS total number of services and journey times (Simul8)

LINE 1 TO SOFIA BUSINESS PARK DIRECTION						
	Observed		**Simulated**		**Observed - Simulated**	
Journey Time	32.42 mins	1945 sec	32.57 mins	1954 sec	-0.15 min	-9 sec
Number of Trains	132		132		0	
LINE 1 TO SLIVNITSA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	32.42 mins	1945 sec	32.79 mins	1967 sec	-0.37 min	-22 sec
Number of Trains	130		130		0	
LINE 2 TO VITOSHA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.42 min	1465 sec	24.42 min	1465 sec	0.00 min	0 sec
Number of Trains	141		141		0	
LINE 2 TO OBELYA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.42 min	1465 sec	24.42 min	1465 sec	0.00 min	0 sec
Number of Trains	139		139		0	
LINE 3 TO HADZHI DIMITAR DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.15 min	1449 sec	24.17 min	1450 sec	-0.02 min	-1 sec
Number of Trains	191		191		0	
LINE 3 TO GORNA BANYA DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	24.15 min	1449 sec	24.17 min	1450 sec	-0.02 min	-1 sec
Number of Trains	192		192		0	
LINE 4 TO SOFIA AIRPORT DIRECTION						
	Observed		**Simulated**		**Observed – Simulated**	
Journey Time	44.42 min	2665 sec	44.42 min	2665 sec	0.00 min	0 sec
Number of Trains	133		133		0	
LINE 4 TO OBELYA DIRECTION						
	Observed		**Simulated**		**Observed - Simulated**	
Journey Time	44.42 min	2665 sec	44.55 min	2673 sec	-0.13 min	-8 sec
Number of Trains	130		130		0	

2.2.1.3 VALIDATION RESULTS OF EXTEND SIM NETWORK MODEL

This subsection describes how the analysis of the metro networks is implemented and investigated using the Extend Sim simulation environment. In doing so, two case studies of AMT and MetroS networks have been evaluated to justify the efficiency and flexibility of the simulation in terms of its application from small to larger and more complex metro networks.

The AMT metro network is modelled using two separate structures, each representing one direction of travel. All eight stations are explicitly represented, accounting for their specific infrastructural characteristics and operational constraints. Based on the available input data sets as described in **Table 1** and **Table 2**, the model accurately simulates passenger flows over the entire operating day, as well as train movements in accordance with the actual service schedules. To this effect, 19 simulation runs have been conducted.

Similarly, the MetroS model accurately represents the four metro lines Each of them is modelled through two distinct directional structures corresponding to the two operating directions. Given that Line 1 and Line 4 share a significant portion of infrastructure; the stations at common section have been modelled as single stations, in both directions. In doing so, passengers travelling between the shared segment of Line 1 and Line 4 can utilise any of them. Thus, this operational setup reflects the real-world operational configuration of the network.

Apart from the input datasets as described under **Table 1** and **Table 2**, the OD matrices for all the lines have been manipulated to better reflect the real-world conditions by developing line-specific OD matrices and utilised in the simulation environment.

The main performance indicators extracted from the simulation include the number of train journeys simulated, average travel time of trains and the passengers from origins to destinations, average passenger queue length at each station and the comparison between simulated and observed passenger outflows, used to assess the accuracy and validity of the model. Some of these indicators are utilised to validate the models where others are useful for the cross validation of the model. Furthermore, a few other indicators representing the flexibility of models to generate number of different outputs to evaluate the metro network and perform numerous future analyses. A brief discussion of these performance indicators and their intended use have been summarised in the following subsection.

To assess the validity and reliability of the proposed models, a comprehensive model validation task has been carried out in this deliverable based on the analysis of key operational and demand-related performance indicators. Specifically, the validation process focuses on the comparison of simulated and observed results in terms of train travel times, number of operations, and passenger flows.

Train travel times and the number of operations is examined to verify the model's ability to accurately reproduce the operational performance of the metro systems. In parallel, passenger demand dynamics are evaluated through a detailed comparison between simulated and observed passenger outflows at a station level. This comparison is illustrated through graphical representations, allowing for a direct assessment of the model's accuracy in capturing spatial and temporal demand patterns.

2.2.1.3.1 AMT NETWORK MODEL (EXTEND SIM)

This subsection focuses on the validation of the model for the AMT system. **Table 14** presents the results of train travel times along a specific direction, together with the total number of operations. Simulated values are compared against the corresponding observed data to assess the accuracy of the model.

To validate passenger flow-related indicators; the outflows of passengers at Principe station have been taken into consideration. This station has been selected as reference location due to its strategic importance within the network, location near to major railway station and maritime terminal.

Table 14: Observed and simulated AMT total number of services and journey times (Extend Sim)

BRIN TO BRIGNOLE DIRECTION						
	Observed		**Simulated***		**Observed - Simulated**	
Journey Time	18.34 min	1100 sec	18.38 min	1103 sec	-0.04 min	-3 sec
Number of Trains	145		145		0	
BRIGNOLE TO BRIN DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	16.35 min	981 sec	16.38 min	983 sec	-0.03 min	-2 sec
Number of Trains	146		146		0	
***Total travel time values are defined to include the train dwell time, passenger boardings the origin station and the complete alighting process at the terminal station.**						

As a first step, a comparison is made between the simulated passenger outflow and the observed outflow. To this extent, the **Figure 3** compares the simulated outflows (light lines) from the 19 different simulation runs with the observed passenger outflows (bold lines), allowing a direct assessment of the model's ability to reproduce demand patterns at station level.

A thorough analysis of the **Figure 3** reveals that the simulated outflows completely overlap the observed outflows, thus proving the accuracy of the simulation model in handling the passenger flow accurately. The other stations of the AMT network provide the same statistics of passenger outflows.

Subsequently, the observed and simulated total inflows and outflows at station levels are analysed, and the differences between them are quantified in terms of percentage variation. Similarly, **Table 15** presents the comparison between the observed and simulated total inflows and outflows along with percentage variations.

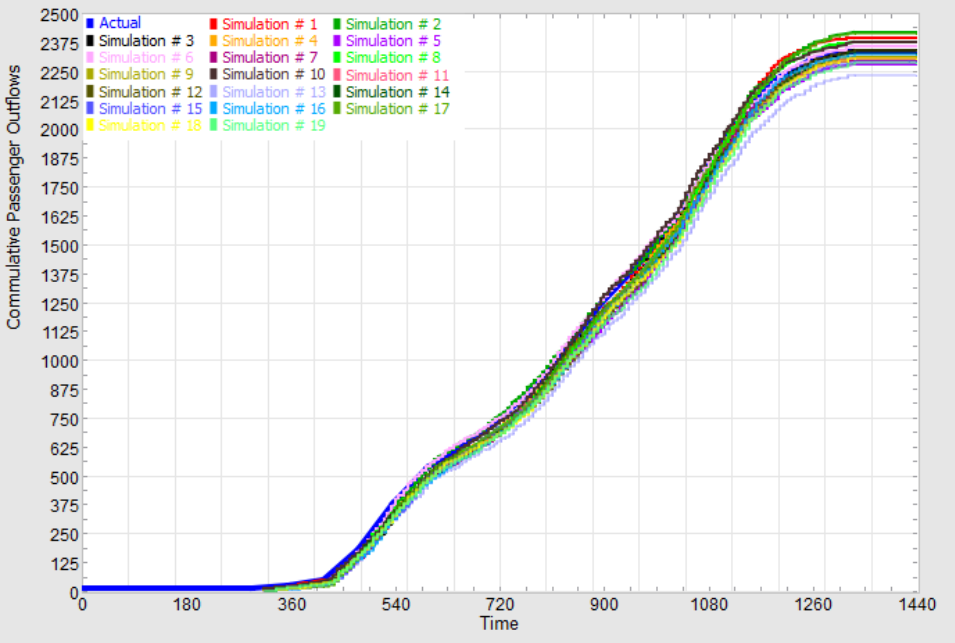


Figure 3: Simulated vs Actual Outflow at Principe Station of the AMT Metro network (Extend Sim)

Table 15: Observed and simulated AMT network model total inflows and outflows (Extend Sim).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
BRIN	6779	6782	0.05%	7040	6949	-1.30%
DINEGRO	5912	5900	-0.20%	5088	5219	2.57%
PRINCIPE	5265	5253	-0.23%	4303	4343	0.92%
DARSENA	3237	3237	-0.01%	3395	3430	1.04%
S. GIORGIO	3850	3829	-0.54%	4364	4352	-0.27%
SARZANO	1911	1908	-0.16%	2078	2084	0.30%
DE FERRARI	4523	4502	-0.47%	6204	6203	-0.01%
BRIGNOLE	8877	8912	0.40%	7882	7743	-1.76%
Total	**40354**	**40323**	**-0.08%**	**40354**	**40323**	**-0.08%**

The **Table 15** highlights divergences of simulated flows from the real flows, both simulated and observed flows are close to each other. In essence, **Table 15** justifies the accuracy of the simulation model where the variations between total simulated and observed inflows and outflows is only ±0.2% at all stations of AMT network. Furthermore, total simulated outflows are also equivalent to the simulated outflows (last row of **Table 15)**, thus providing a quantitative measure of the model's accuracy in reproducing passenger demand patterns accurately.

2.2.1.3.2 METROS NETWORK MODEL (EXTEND SIM)

This subsection addresses the validation of the model developed for the MetroS network. The validation process of metro Sofia network is carried out using the same set of outputs as described for AMT network under **Section 2.1.2**.

Similarly, **Table 16** reports the train travel times in each direction for the four metro lines operating in MetroS network, along with the corresponding total number of services. Simulated results are compared with observed values to evaluate the model's performance.

Table 16: Observed and simulated MetroS total number of services and journey times (Extend Sim)

LINE 1 TO SOFIA BUSINESS PARK DIRECTION						
	Observed		**Simulated***		**Observed - Simulated**	
Journey Time	32.42 mins	1945 sec	32.47 mins	1948sec	-0.05 min	-3 sec
Number of Trains	132		132		0	
LINE 1 TO SLIVNITSA DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	32.42 mins	1945 sec	32.48 mins	1949 sec	-0.06 min	-4 sec
Number of Trains	130		130		0	
LINE 2 TO VITOSHA DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	24.42 min	1465 sec	24.48 min	1469 sec	-0.06 min	-4 sec
Number of Trains	141		141		0	
LINE 2 TO OBELYA DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	24.42 min	1465 sec	24.49 min	1469 sec	-0-07 min	-4 sec
Number of Trains	139		139		0	
LINE 3 TO HADZHI DIMITAR DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	24.17 min	1450 sec	24.18 min	1451 sec	-0.01 min	-1 sec
Number of Trains	191		191		0	
LINE 3 TO GORNA BANYA DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	24.17 min	1450 sec	24.19 min	1451 sec	-0.02 min	-1 sec
Number of Trains	192		192		0	
LINE 4 TO SOFIA AIRPORT DIRECTION						
	Observed		**Simulated***		**Observed – Simulated**	
Journey Time	44.42 min	2665 sec	44.52 min	2671 sec	-0.10 min	-6 sec
Number of Trains	133		133		0	
LINE 4 TO OBELYA DIRECTION						
	Observed		**Simulated***		**Observed - Simulated**	
Journey Time	44.42 min	2665 sec	44.49 min	2669 sec	-0.07 min	-4 sec
Number of Trains	130		130		0	

*Total travel time values are defined to include the train dwell time, passenger boardings the origin station and the complete alighting process at the terminal station.

In case of the Metro Sofia network, the validation of passenger flow-related indicators in terms of the outflows of passengers, have been evaluated at Serdica Station of metro Line 1 and Line 4. This station

has also been selected as reference location due to its strategic importance within the network and its high levels of passenger activity as well as a major transfer hub for the passengers travelling to\from different metro lines.

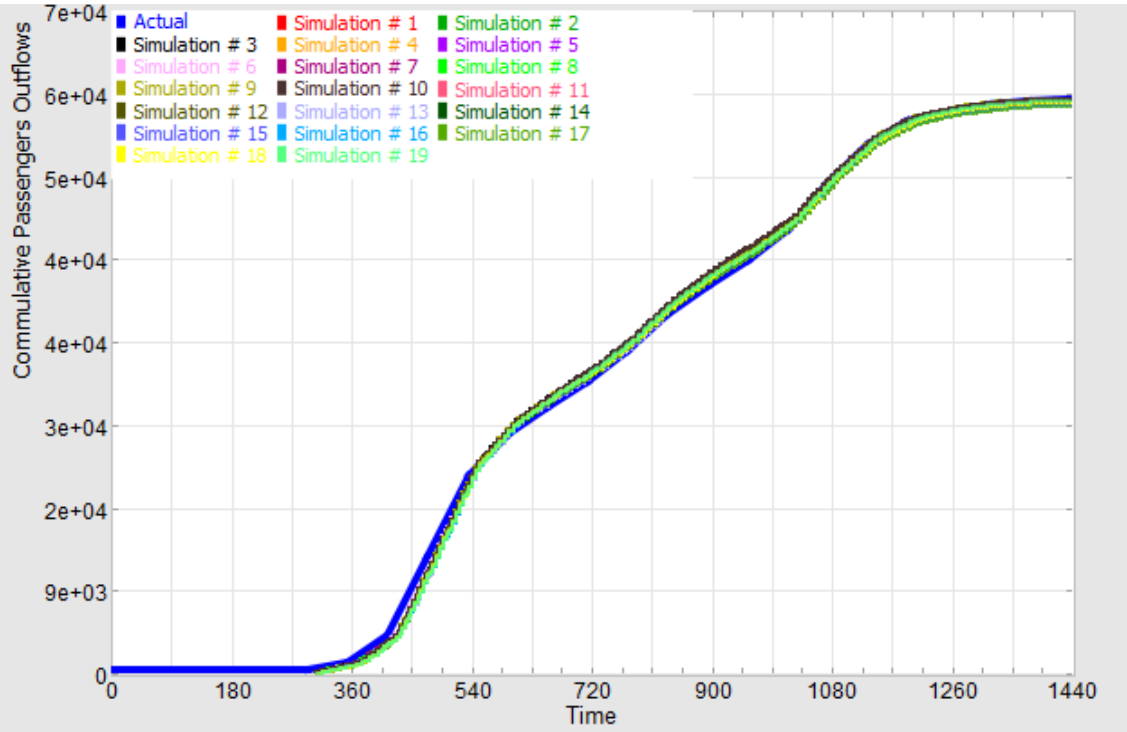


Figure 4: Simulated vs Actual Outflow at Serdica station of MetroS line 1 and 4 (Extend Sim)

Similar to the analysis of AMT network, the **Figure 4** compares the simulated outflows (light lines) from the 19 different simulation runs with the observed passenger outflows (bold lines), allowing a direct assessment of the model's ability to reproduce demand patterns at station level.

A thorough analysis of the **Figure 4** reveals that the simulated outflows completely overlap the observed outflows, thus proving the accuracy of the simulation model in handling the passenger flow accurately. The other stations of all lines of MetroS network provide the same statistics of passenger outflows.

Subsequently, **Table 17, Table 18, Table 19 and Table 20** report the observed and simulated total inflows and outflows at station levels along with percentage variations for all four lines of metro Sofia.

Like the AMT metro network, the tables also justifies the accuracy of the simulation model to regenerate passenger demand patterns while considering larger and more complex metro networks.

Results in Table 20 justify the accuracy of the optimisation model where the variations between total simulated and observed inflows and outflows is also $\pm$0.2% at all stations of the AMT metro network. Furthermore, total simulated outflows are also equivalent to the simulated outflows, thus providing a quantitative measure of the model's accuracy in reproducing passenger demand patterns accurately.

Table 17: Observed and simulated inflows and outflows for Line 1 of MetroS (Extend Sim).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Slivnitsa	6973	6987	0.21	6144	6128	-0.26
Lyulin	9876	9911	0.36	7883	7869	-0.18
Zapaden Park	6712	6713	0.01	5682	5665	-0.30
Vardar	6883	6888	0.07	6722	6716	-0.09
Konstantin Velichkov	9434	9426	-0.08	8620	8643	0.26
Opalchenska	11974	11967	-0.06	11745	11778	0.28
Serdica	71250	71106	-0.20	74269	74060	-0.28
SU Sv. Kliment Ohridski	16190	16138	-0.32	16189	16037	-0.94
Vasil Levski Stadium	10353	10338	-0.15	11853	11775	-0.66
F. Joliot-Curie	6309	6280	-0.47	8529	8514	-0.18
G.M. Dimitrov	13410	13402	-0.06	10858	10819	-0.36
Musagenitsa	3121	3092	-0.93	3597	3590	-0.20
Mladost 1	6772	6762	-0.15	7753	7776	0.30
Aleksandar Malinov	5409	5397	-0.21	5033	5070	0.73
Akad. Al. Teodorov - Balan	8306	8259	-0.57	8497	8472	-0.29
Business Park	11869	11847	-0.18	11467	11601	1.17
Total	**204841**	**204512**	**-0.16**	**204841**	**204512**	**-0.16**

Table 18: Observed and simulated inflows and outflows for Line 2 of MetroS (Extend Sim).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Obelya	6972	6972	0.00	6370	6348.5	-0.34
Lomsko Shose	2340	2338	-0.07	1974	1964	-0.51
Beli Dunav	10839	10837	-0.02	8912	8877	-0.39
Nadezhda	8608	8605	-0.03	7842	7858	0.21
Khan Kubrat	5824	5792	-0.55	6359	6363	0.06
Princess Marie Louise	5691	5668	-0.41	5936	5968	0.54
Central Railway Station	5434	5411	-0.43	8980	8991	0.12
Lavov most	7317	7353	0.49	7516	7503	-0.17
Serdika	127842	126987	-0.67	117659	117472	-0.16
National Palace of Culture (NDK)	86433	86213	-0.25	99783	98993	-0.79
European Union	10177	10171	-0.06	8155	8079	-0.94
James Bourchier	10891	10786	-0.96	10797	10694	-0.95
Vitosha	18547	18483	-0.34	16632	16506	-0.76
Total	**306915**	**305617**	**-0.42**	**306915**	**305617**	**-0.42**

Table 19: Observed and simulated inflows and outflows for Line 3 of MetroS (Extend Sim).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Gorna Banya	4063	4062	-0.02	4058	4580	12.87
Ovcha Kupel	3111	3083	-0.89	2673	2661	-0.45
Moesia	9989	9976	-0.13	7577	7532	-0.59
Ovcha Kupel	4279	4285	0.14	4166	4120	-1.10
Krasno selo	12130	12106	-0.20	18099	17950	-0.82
Bulgaria	8580	8581	0.01	12324	12206	-0.96
Medical University	7968	7991	0.29	9898	9837	-0.62
National Palace of Culture (NDK)	83892	83802	-0.11	70224	69929	-0.42
Sveti Patriarh Evtimiy	5804	5779	-0.42	8656	8650	-0.07
Orlov Most	7616	7615	-0.02	7938	7948	0.13
Teatralna	6962	6948	-0.21	7834	7803	-0.40
Hadzhi Dimitar	10957	10931	-0.24	11904	11941	0.31
Total	**165351**	**165157**	**-0.12**	**165351**	**165157**	**-0.12**

Table 20: Observed and simulated inflows and outflows for Line 4 of MetroS (Extend Sim).

Station Name	Inflows			Outflows		
	Obs.	Sim.	% Diff.	Obs.	Sim.	% Diff.
Obelya	9450	9521	0.75	8781	8976	2.22
Slivnitsa	7921	7928	0.09	6431	6341	-1.40
Lyulin	10570	10543	-0.26	8630	8591	-0.46
Zapaden park	6656	6659	0.04	6081	6055	-0.43
Vardar	7104	7126	0.31	7122	7127	0.07
Konstantin Velichkov	10134	10149	0.15	8623	8650	0.31
Opalchenska	11634	11659	0.22	11503	11394	-0.95
Serdika	72623	72497	-0.17	78218	78058	-0.20
Sofia University St Kliment Ohridski	17455	17456	0.00	13951	13733	-1.56
Vasil Levski Stadium	10251	10239	-0.12	11637	11576	-0.52
Joliot-Curie	5774	5790	0.28	9016	8965	-0.57
G. M. Dimitrov	13781	13765	-0.12	11169	11097	-0.65
Musagenitsa	3384	3368	-0.48	3558	3550	-0.21
Mladost 1	7258	7223	-0.48	7631	7638	0.09
Mladost 3	5366	5321	-0.84	4893	4888	-0.11
Inter Expo Center - Tsarigradsko shose	9715	9720	0.05	9096	9103	0.07
Druzhba	5949	5894	-0.93	7169	7264	1.32
Iskarsko shose	4441	4398	-0.97	4851	4889	0.78
Sofiyska Sveta gora	909	905	-0.50	804	807	0.39

Sofia Airport	3121	3069	-1.67	4332	4528	4.53
Total	**223496**	**223229**	**-0.12**	**223496**	**223229**	**-0.12**

2.2.1.4 CROSS-VALIDATION RESULTS OF NETWORK MODELS

This section aims to cross-validate the metro network model results from each modelling platform and demonstrate if the output generated are consistent and comparable with one another. The cross-validation exercise is to compare output metrics from the three sets of metro network models developed for the two use cases (i.e. AMT in Genoa and MetroS in Sofia) using the three modelling platforms AnyLogic, Simul8 and Extend Sim. All the models adopted the same set of input parameters and assumptions as described in **Section 2.1.2**. Cross-comparison of model outputs includes the total number of services operated, end-to-end train journey times, as well as the station level passenger inflows and outflows. The train journey time was defined as the sum of all the scheduled station-to-station travel times and the station dwell times at all stations (including the start and end stations).

Focusing on train services, the validation results as summarised in **Table 5, Table 12** and **Table 14 (**for AMT in Genoa) and **Table 7, Table 13** and **Table 16** (for MetroS in Sofia) have shown that the simulated total number of train services operated and the train journey times generated from the three modelling platforms are within a very narrow margin, as well as closely matched with the scheduled number of services and train journey times, implying that the network models are not only valid individually but also capable of generating consistent, comparable as well as cross-validating train service operational outputs across different modelling platforms.

As for passenger movements, validation results as summarised in **Table 7** and **Table 15** (for AMT) as well as in **Table 8** to **Table 11** and **Table 20** (for MetroS) have shown that the simulated network level passenger flow information from the metro network models developed across different modelling platforms are consistent with one another as well.

Whilst some discrepancies in the simulated outbound flows were observed for a few individual stations with the AnyLogic network model, it is important to noted that, as explained in **Section 2.2.1.1.3**, these deviations arise from a discrepancy in resolution between the hourly input demand data and the execution of alighting events at the train level in the simulation. As no passengers are lost in the process and the aggregate totals remain accurate across all four lines, these discrepancies are structural and explainable, rather than indicative of model error. The inbound flows, which validate the boarding logic directly, show strong agreement with the observed data at both the aggregate and the individual station levels. The differences between the observed and simulated data are distributed symmetrically around zero, with no systematic bias in any direction. The overall assessment indicates that the network models remain robust and reliable. This justifies the validity and reliability of the metro network models which have broadly replicated the passenger movements across the entire metro networks.

To conclude, the above validation and cross-validation results have demonstrated that outputs generated from this set of metro network models are representative of the real-world metro operations in terms of both train and passenger movements, even though the modelling logics and capabilities vary slightly across different modelling platforms. This provides a multi-perspective powerful simulation tool, to be applied in the model exploitation stage, to evaluate the performance of the current metro

networks as well as to assess the possible impacts of fully automated operations (i.e. GoA4) on the current metro systems.

2.2.2 STATION MODEL VALIDATIONS AND CROSS-VALIDATIONS

This section describes the details of validations and cross-validations for the station models developed using AnyLogic and Simul8, according to the framework in **Section 2.1**. The model runs were carried out according to the standard passenger demand information and operational conditions as defined in **Section 2.1.2** concerning the simulation environment core parameters. The station model validation and cross-validation exercises will focus on the followings:

- Model Validations Results – for each station model, comparing model outputs against real-world observable metrics.
- Model Cross-Validations Results – comparing station model outputs between AnyLogic and Simul8.

2.2.2.1 VALIDATION RESULTS OF ANYLOGIC STATION MODEL

2.2.2.1.1 STRUCTURAL AND BEHAVIOURAL VALIDATION

The AnyLogic station models focus on providing a microscopic, agent-based representation of passenger behaviour and infrastructure usage within individual stations. Selected case study stations include De Ferrari in Genoa and Serdica in Sofia.

The validation process consisted of:

- A conceptual-to-implementation consistency check, in which all station facilities, passenger routing logic, boarding/alighting processes, platform capacities, vertical circulation (stairs, escalators and elevators) and inter-line transfer pathways were verified against station blueprints and operational data.
- Run-time behavioural validation: simulation runs with animation enabled allowed visual inspection of passenger flows, queuing behaviour, boarding/alighting dynamics and transfer movements between lines. These checks confirmed that the implemented logic accurately reproduces observed station operations and congestion patterns.
- Output trace and KPI verification: KPIs, including inbound/outbound station-level flows, ticket type distribution, platform occupancy and transfer volumes, were monitored and compared with observed datasets to ensure consistency and accuracy.

These checks confirmed that the AnyLogic station models accurately reflect the real-world operational behaviour of both stations.

2.2.2.1.2 DE FERRARI STATION MODEL

Inbound Passenger Flows and Ticket Types

The validation of inbound passenger flows at De Ferrari station concentrated on the number of passengers entering the station and reaching the platforms. The model simulated 4,398 inbound passengers, which is slightly below the observed figure of 4,523 (a difference of approximately 2.8%).

This minor discrepancy is consistent with natural fluctuations in passenger arrivals and remains within acceptable limits for validation purposes.

The distribution of ticket types is also accurately reproduced: prepaid tickets account for 97% of both the observed and simulated data, while tickets purchased from machines account for the remaining 3%. These results confirm that the model correctly reproduces passenger behaviour in terms of both volume and ticket usage at entrances.

Table 21: Observed and simulated inbound passenger flows at De Ferrari Station (AnyLogic)

ENTRANCES	OBSERVED	SIMULATED	PLATFORMS	SIMULATED
			To Brin	3430
Entrances	4523	4398	**To Brignole**	968
Total	4523	4398	**Total**	4398

Table 22: Observed and simulated inbound passenger ticket types at De Ferrari Station (AnyLogic)

	OBSERVED		SIMULATED	
Prepaid Tickets	4387	97.0%	4266	97.0%
Ticket Machines	136	3.0%	132	3.0%
Total	4523	100.0%	4398	100.0%

Outbound Passenger Flows

The outbound flows from the platforms are also accurately represented. The number of passengers arriving from Brignole (1,894) and Brin (4,307) matches the observed count exactly. This demonstrates that the dynamics of boarding and alighting, and the logic of platform allocation, are well captured, ensuring realistic passenger movement through the station.

Overall, the De Ferrari station model accurately reproduces the spatial and temporal patterns of passenger movement, providing a reliable basis for further scenario analysis.

Table 23: Observed and simulated outbound passenger flows at De Ferrari Station (AnyLogic)

	OBSERVED	SIMULATED
	Platforms	**Platforms**
From Brignole	1894	1894
From Brin	4307	4307
Total	6291	6291

2.2.2.1.3 SERDICA STATION MODEL

Inbound Passenger Flows

Inbound flows at Serdica station, across its 12 entrances in the north, east and west, are reproduced within ±3% of the observed values. Some east entrances show slight overestimation (+2.7%), while

some north entrances show slight underestimation (-2.1%), reflecting natural variability in passenger arrival patterns. The total simulated inflow of 40,236 is close to the observed figure of 39,984 (a difference of approximately 0.6%).

Table 24: Observed and simulated inbound passenger flows at Serdica Station (AnyLogic)

	NORTH ENTRANCES		EAST ENTRANCES								WEST ENTRANCES		TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	
Observed	6339	6339	2731	2731	2731	2731	2731	2731	2731	2731	2731	2731	39984
Simulated	6206	6504	2786	2723	2805	2772	2728	2734	2745	2797	2703	2709	40212
% Diff	-2.10%	2.60%	2.00%	-0.30%	2.70%	1.50%	-0.10%	0.10%	0.50%	2.40%	-1.0%	-0.80%	0.57%

Similarly, ticket type distribution is well captured, with prepaid tickets accounting for around 97% and tickets purchased from machines accounting for around 3%. This is consistent with the observed behaviour across all entrances.

Table 25: Observed and simulated inbound passenger ticket types at Serdica Station (AnyLogic)

	NORTH ENTRANCES			EAST ENTRANCES			WEST ENTRANCES		
	Observed	**Simulated**	**%**	**Observed**	**Simulated**	**%**	**Observed**	**Simulated**	**%**
Prepaid Tickets	6464	6466	96.8%	19392	26165	97.1%	6464	6411	97.1%
Ticket Machines	200	214	3.2%	600	782	2.9%	200	193	2.9%

Outbound and transfer flows are also accurately simulated. Differences at platform level remain within ±2%. For example, Platform III shows a slight underestimation of -2%, while Platform I matches observed values almost perfectly. The total number of outbound and transfer passengers differs by only -1% from the observed total.

This confirms that the simulation effectively represents passenger movements between platforms, boarding onto trains and transfers between lines, including the timing and distribution of flows across the station.

Table 26: Observed and simulated outbound passenger flows at Serdica Station (AnyLogic)

		Observed			Simulated			Differences		
		Outbound	**Transfer**	**Total**	**Outbound**	**Transfer**	**Total**	**Outbound**	**Transfer**	**Total**
East & West	Platform I	12211	21561	33772	12391	21394	33785	1%	-1%	0%
	Platform II	12211	26074	38285	12159	25961	38120	0%	0%	0%
North	Platform III	7291	27367	34658	7314	26683	33997	0%	-2%	-2%
	Platform IV	7291	24218	31509	7166	23923	31089	-2%	-1%	-1%
Total		**138224**			**138327**			**-1%**		

2.2.2.2 VALIDATION RESULTS OF SIMUL8 STATION MODEL

The Simul8 station models validation exercise is to compare the number of observed and simulated inbound, outbound, and transfer passenger flows at some key check points within the two selected stations The same three-layer checking process as it was done for the Simul8 network models, was undertaken and confirmed that the Simul8 model runs performed very well and the on-screen animations had also shown logical movements of the simulated trains and passengers. This reflects that, visually, the models could satisfactorily replicate the real-world operations of the two selected metro stations.

2.2.2.2.1 DE FERRARI STATION MODEL

Inbound Passenger Flows

De Ferrari station model validation first focused on the accuracy in simulating the passenger flows at the entrances and at the platforms, as well as the ticket types. The comparison between the simulated and observed inbound passenger flows (generated at the entrances and arrived at the platforms) and ticket types are summarised in **Table 27** and **These results** show that the passenger flows are accurately simulated at both the entrances and platforms. The ticket types (of 97% prepaid to 3% purchasing) have also been accurately simulated as well (Table 28).

Table 28.

Table 27: Observed and simulated inbound passenger flows at De Ferrari Station (Simul8)

ENTRANCES	OBSERVED	SIMULATED	PLATFORMS	SIMULATED
Grand Entrance	2242	2242	**To Brin**	3318
Subway Entrance	2242	2242	**To Brignole**	1166
Total	4484	4484	**Total**	4484

These results show that the passenger flows are accurately simulated at both the entrances and platforms. The ticket types (of 97% prepaid to 3% purchasing) have also been accurately simulated as well (**Table 28**).

Table 28: Observed and simulated inbound passenger ticket types at De Ferrari Station (Simul8)

	OBSERVED		SIMULATED	
Prepaid Tickets	4349	97.0%	4346	96.9%
Ticket Machines	135	3.0%	138	3.1%
Total	4484	100.0%	4484	100.0%

Outbound Passenger Flows

As for outbound passenger flows, which are the passenger arriving at the station from the train and then navigating through to the station exits. The simulated and observed outbound passenger flows are summarised in **Table 29**. It shows that the model has accurately simulated the outbound passenger flows at the platforms, concourse and exit levels.

Table 29: Observed and simulated outbound passenger flows at De Ferrari Station (Simul8)

	OBSERVED	SIMULATED		
	Platforms	Platforms	Concourse	Exit
From Brignole	4307	4307	–	–
From Brin	1869	1869	–	–
Total	6176	6176	6176	6176

The above validation results demonstrate that the Simul8 station model could reliably replicate the passenger flows within the De Ferrari station, in terms of inbound and outbound passenger flows at various levels of the station.

2.2.2.2.2 SERDICA STATION MODEL

Inbound Passenger Flows

The Serdica Station Model validation focused on the accuracy in simulating the passenger flows at the entrances, as well as the ticket types.

Table 30: Observed and simulated inbound passenger flows at Serdica Station (Simul8)

	NORTH ENTRANCES		EAST ENTRANCES								WEST ENTRANCES		TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	
Observed	6339	6339	2734	2734	2734	2734	2734	2734	2734	2734	2734	2734	40018
Simulated	6319	6357	2758	2771	2734	2732	2727	2730	2746	2720	2756	2771	40121
% Diff	0.3%	-0.3%	-0.9%	-1.4%	0.0%	0.1%	0.3%	0.1%	-0.4%	0.5%	-0.8%	-1.4%	-0.3%

The comparison between the simulated and observed inbound passenger flows generated at the entrances and the ticket types are summarised in **Table 30** and **Table 31**. These results show that the passenger flows are accurately simulated to within 1.5% of the observed throughputs at both the entrances levels. The ticket types (of 97% prepaid to 3% purchasing) have also been accurately simulated as well.

Table 31: Observed and simulated inbound passenger ticket types at Serdica Station (Simul8)

	NORTH ENTRANCES			EAST ENTRANCES			WEST ENTRANCES		
	Observed	Simulated	%	Observed	Simulated	%	Observed	Simulated	%
Prepaid Tickets	12298	12298	97%	21216	21216	97%	5304	5357	97%
Ticket Machines	380	377	3%	656	656	3%	164	170	3%

Outbound and Transfer Passenger Flows

As for passengers arriving from trains (Lines 1, 2 and 4) at the Serdica station, it included passengers navigating through to the station exits (i.e. “Outbound Passengers”) as well as passengers transferring between platforms I & II (i.e. for Line 2) and platforms III & IV (for Lines 1 and 4) – i.e. “Transfer

Passengers". The simulated and observed outbound and transfer passenger arriving at the Serdica station are summarised in **Table 32**. It shows that the model has accurately simulated the outbound and transfer passenger flows to within 2% of the observed flows.

Table 32: Observed and simulated outbound and transfer passenger flows at Serdica Station (Simul8)

		Observed			Simulated			Differences		
		Outbound	Transfer	Total	Outbound	Transfer	Total	Outbound	Transfer	Total
East & West	Platform I	12211	21561	33772	12183	21645	33828	0.2%	-0.4%	-0.2%
	Platform II	12211	26074	38285	12196	26123	38319	0.1%	-0.2%	-0.1%
North	Platform III	7291	27367	34658	7291	27375	34666	0.0%	0.0%	0.0%
	Platform IV	7291	24218	31509	7272	24242	31514	0.3%	-0.1%	0.0%
Total		**138224**			**138327**			**-0.1%**		

Looking at the combined outbound and transfer passengers arriving at Serdica station by each metro line, the flows are summarised in **Table 33**. Again, the model has accurately simulated the combined outbound and transfer passenger flows arriving from each metro lines to within 2% of the observed flows.

Table 33: Observed and simulated outbound and transfer passenger flows arriving by metro lines at Serdica Station (Simul8)

	Line 2 to Vitosha	Line 2 to Obelya	Line 1 to Slivnitsa	Line 4 to Obelya	Line 4 to Sofia Airport	Line 1 to Business Park	Total
Observed	31509	34658	19142.5	19142.5	16886	16886	138224
Simulated	31514	34666	19164	19155	16949	16879	138327
% Diff	-0.02%	-0.02%	-0.11%	-0.07%	-0.37%	0.04%	-0.07%

Passenger Boarding at Platforms

To check simulated passengers queuing at platforms and getting onboard of trains, observed total number of passengers coming from entrances (i.e. inbound passengers) and transferred between metro lines were compiled and summarised in **Table 34**. Again, it shows that the model could accurately simulate to within 1% of the observed total number of passengers at platforms waiting to board.

Table 34: Observed and simulated passenger boarding at platforms at Serdica Station (Simul8)

	Line 2 to Vitosha	Line 2 to Obelya	Line 4 to Obelya	Line 4 to Sofia Airport	Line 1 to Slivnitsa	Line 1 to Business Park	Total
Simulated	36040	35069	15080	18920	15133	18884	139126
Transferred	24284	23351	11002	14791	11002	14791	99221
Inbound	11807	11807	4101	4101	4101	4101	40018
Observed Total	36091	35158	15103	18892	15103	18892	139239
% Diff	-0.14%	-0.25%	-0.15%	0.15%	0.20%	-0.04%	-0.08%

The above validation results demonstrate that the Simul8 station model could reliably replicate the passenger flows at Serdica station, in terms of inbound, outbound and transfer passenger flows at various levels of the station.

2.2.2.3 CROSS-VALIDATION RESULTS OF STATION MODELS

This section aims to cross-validate the metro station model results generated from the two modelling platforms (AnyLogic and Simul8) and demonstrate if the results generated are consistent and comparable with one another. The cross-validation exercise is to compare outputs from the two sets of metro station models developed for the De Ferrari Station of AMT in Genoa and the Serdica Station MetroS in Sofia using AnyLogic and Simul8. The baseline conditions as described in **Section 2.1.2**. All the models adopted the same set of input parameters and assumptions. Cross-comparison of model outputs mainly concerns the inbound and outbound passenger flow statistics at different key checkpoints within the stations.

Focusing on inbound passenger flows, the validation results as summarised in **Table 21** vs **Table 27 (**at De Ferrari Station of AMT) and **Table 24** vs **Table 30** (at Serdica of MetroS) show that the simulated inbound passenger flows at both stations using both modelling platforms are within a very narrow margin, as well as closely match with the observed inbound passenger statistics. The split between passengers using prepaid and non-prepaid tickets was also consistently simulated (**Table 22** vs **These results** show that the passenger flows are accurately simulated at both the entrances and platforms. The ticket types (of 97% prepaid to 3% purchasing) have also been accurately simulated as well (Table 28).

Table 28 for De Ferrari; and **Table 25** vs **Table 31** for Serdica).

As for outbound passenger flows, validation results as summarised in **Table 23** vs **Table 29** (at De Ferrari of AMT) as well as in **Table 26** vs **Table 32** (at Serdica of MetroS) show that the simulated outbound passenger flows match very well between the two modelling platforms. This further justifies the validity and reliability of the metro station models which accurately replicate the passenger movements within the two selected metro stations.

To conclude, the above validation and cross-validation results demonstrate that outputs generated from this set of metro station models are ***representative of the real-world metro operations*** in terms of passenger movements within the two selection metro stations of AMT and MetroS, even though the modelling logics and capabilities vary slightly across different modelling platforms. This provides a multi-perspective powerful simulation tool, to be applied in the model exploitation stage, to evaluate the performance of the current metro networks as well as to assess the possible impacts of fully automated operations (i.e. GoA4) on the current metro systems.

2.3 MODEL EXPLOITATIONS

In this section, the suite of metro simulation models validated and cross-validated in **Section 2.2** is exploited to simulate the 5 scenarios as described in **Section 2.1.3**. For easier references, these scenarios and their model specifications are recapped in **Table 35**. A ***multi-level, multi-purpose, and multi-perspective evaluation approach*** is adopted to reveal the strengths and weaknesses of the current metro systems, as well as the impacts of fully automated operations.

These scenarios have been compiled to represent different purposes of evaluation. In particular, Scenarios 1 to 3 aim to evaluate the strengths and weaknesses of the current systems, and scenarios 4 and 5 will explore the possible impacts of fully automated metro operations in handling peak demand conditions as well as during unusual situations such as large scale planned events (e.g. football matches, protests, and temporary closure of a critical station).

The evaluation was conducted at three different levels with clear focuses aiming to reveal impacts at

(i) macroscopic level – providing an overall network-wide assessment of the two metro systems using the metro network simulation models;

(ii) mesoscopic level – focusing on individual metro line and/or station level impacts, considering the operation of the whole metro network, using the metro network simulation models; and

(iii) microscopic level – taking a crowd management perspective to reveal impacts on station level passenger movements, using the metro station simulation models.

This multi-level evaluation approach takes advantages of the multi-perspective metro simulation model suite developed and validated in **Section 2.2**. Key findings, impacts and potential benefits will be revealed by comparing simulation results across different scenarios (representing different evaluation purposes, peak, off-peak, resilience to disruptions, and fully automated operations). This provides a comprehensive range of evidence-based support for the evaluation and derivation of final recommendations in terms of the strengths and weaknesses of the current metro systems and the potential impacts of fully automated operations.

Table 35: Summary of Scenario Specifications for Model Evaluations

Description	Scenario ID	Specifications	Metro System
Peak	Scenario 1A	07:00 -10:00 on a typical day	AMT
	Scenario 1B	07:00 - 10:00 on a typical day	MetroS
	Scenario 1C	16:00 – 20:00 on a typical day	AMT
	Scenario 1D	16:00 – 20:00 on a typical day	MetroS
Off-peak	Scenario 2A	05:00 - 07:00 on a typical day	AMT
	Scenario 2B	05:00 - 07:00 on a typical day	MetroS
	Scenario 2C	10:00 -16:00 on a typical day	AMT
	Scenario 2D	10:00 -16:00 on a typical day	MestroS
	Scenario 2E	20:00 – 24:00 on a typical day	AMT
	Scenario 2F	20:00 – 24:00 on a typical day	MetroS
Disruption	Scenario 3A	During Match Day (Focus: 16:00 – 24:00)	AMT
	Scenario 3B	During Planned Events (Focus: 08:00 - 12:00)	AMT
	Scenario 3C	Closing Serdica Station (Focus: 16:00 - 20:00)	MetroS
Peak GoA4	Scenario 4A	120 sec train headway (Focus: 16:00 – 20:00 on a typical day)	AMT
	Scenario 4B	120 sec train headway (Focus: 16:00 – 20:00 on a typical day)	MetroS
	Scenario 4C	90 sec train headway (Focus: 16:00 – 20:00 on a typical day)	AMT

Description	Scenario ID	Specifications	Metro System
	Scenario 4D	90 sec train headway (Focus: 16:00 – 20:00 on a typical day)	MetroS
Disruption GoA4	Scenario 5A	120 sec train headway (Focus: 16:00 - 24:00 on a Match Day)	AMT
	Scenario 5B	120 sec train headway (Focus: 08:00 – 12:00 during Planned Events)	AMT
	Scenario 5C	120 sec train headway (Focus: 16:00 – 20:00 if closing Serdica from 16:00 to 18:00)	MetroS
	Scenario 5D	90 sec train headway (Focus: 16:00 - 24:00 on a Match Day)	AMT
	Scenario 5E	90 sec train headway (Focus: 08:00 – 12:00 during Planned Events)	AMT
	Scenario 5F	90 sec train headway (Focus: 16:00 – 20:00 if closing Serdica from 16:00 to 18:00)	MetroS

2.3.1 OVERALL NETWORK-WIDE EVALUATIONS

This section summarises the key observations and findings from a macroscopic perspective to provide an overall network-wide evaluation of the two metro systems. The evaluation will be primarily based on results from the metro network simulation models, covering all five scenarios for both metro systems considered within the NEXUS project.

2.3.1.1 AMT – GENOA'S METRO

Baseline – Normal Operations (Scenarios 1 and 2)

These scenarios simulate normal operating conditions on the AMT network across five time bands, covering the full-service day from 05:00 to 24:00. Two peak periods are identified: morning (07:00–10:00) and afternoon (16:00–20:00), alongside three off-peak periods. No external demand shocks or infrastructure disruptions are applied; the scenarios reflect the standard scheduled service as currently operated.

The following KPIs are reported for each time band. Waiting time at stations is defined as the time each passenger spends on the platform from the moment they arrive until they are on board a train. This does not include time spent entering the station or walking to the platform, nor the time spent boarding once the train doors open. The reported value is the mean across all passengers who boarded during the timeframe.

Train saturation level is expressed as a percentage of train capacity measuring the ratio of passengers on board to the train's total capacity. At any given moment, this value is computed across all trains currently in service on the network, and an hourly average is derived from these instantaneous readings. This statistic reflects the actual loading conditions of trains during the period in question and does not carry forward loading from previous hours.

Train occupancy is the raw number of passengers on board at a given moment, without normalisation for capacity. Like saturation, it is computed as an hourly mean across all trains in service. While saturation expresses load as a percentage of capacity, occupancy expresses it as an absolute passenger count, making it easier to interpret in terms of comfort and crowding conditions on board.

Passenger throughput per station is the number of passengers who board or alight at each station during the simulated time window. Only passengers who physically board or alight from a train are counted; those who remain on board without stopping are not included. This value is cumulative over the timeframe, allowing direct comparisons to be made between stations and time slots in terms of demand intensity.

Table 36: KPI Summary for Scenarios 1 and 2 (AMT Network model).

(Scenario) Time Band	Average Wait Time (min)	Average Train Saturation Level (%)	Maximum Train Saturation Level (%)	Average Occupancy (# of pax.)	Average Number of Passengers Waiting
(2A) Off-peak 5:00–7:00	2.72	6.4%	11.1%	18.5	30
(1A) Peak 7:00–10:00	**2.91**	**32.7%**	**43.2%**	**94.9**	**190**
(2C) Off-peak 10:00–16:00	2.90	20.0%	33.4%	57.9	151
(1C) Peak 16:00–20:00	**2.96**	**23.3%**	**33.4%**	**67.6**	**195**
(2E) Off-peak 20:00–24:00	3.76	6.8%	21.0%	19.6	44

Note: Train Saturation = % occupancy vs. train capacity. Average number of passengers waiting = number of passengers waiting at stations at any given moment.

- The morning peak (7:00–10:00) is the most stressed band, with saturation level reaching an average of 32.7% and a peak of 43.2%, and occupancy approaching 95 passengers per train on average (momentary maximum of 125.4). This is the only band where train loading approaches comfort-critical levels.
- Waiting time is remarkably stable across all daytime bands (2.72 to 2.96 min), which reflects a consistent and reliable schedule. The clear outlier is the 20:00–24:00 band (Scenario 2F), where the average rises to 3.76 min and the maximum reaches 5.6 min, a direct consequence of reduced service frequency in late-evening hours.

Throughput by Station

Table 37: Number of passengers served per time band per station.

(Scenario) Time Band	Brin	Dinegro	Principe	Darsena	S.Giorgio	S.Agostino	De Ferrari	Brignole
(2A) 5:00–7:00	420	156	147	61	94	37	121	299
(1A) 7:00–10:00	4,449	2,573	2,750	1,190	1,428	646	1,966	4,731
(2C) 10:00–16:00	4,855	4,211	3,372	2,612	3,025	1,331	3,990	6,120
(1C) 16:00–20:00	3,504	3,278	2,676	2,302	2,899	1,642	3,728	4,977
(2E) 20:00–24:00	756	550	468	325	497	168	631	924

- Brignole and Brin are consistently the highest-demand terminals across all time bands, confirming their role as the main entry/exit poles of the line.
- The off-peak midday band (10:00–16:00) records the highest absolute throughput of the day at several stations. Brignole reaches 6,120 passengers over 6 hours, higher than either peak

band. This indicates substantial midday mobility suggests that the distinction between 'peak' and 'off-peak' is less sharp than expected from a pure commuting model.
- Sant'Agostino consistently records the lowest throughput on the line across all bands, indicating structurally lower demand at that node.

Disruption Scenarios (Scenario 3)

Two crowd-driven disruption events were simulated: a protest/demonstration in the 8:00–12:00 window (Scenario 3B) and a football match generating demand in the 16:00–24:00 window (Scenario 3A). Both produce a significant increase in total passenger demand relative to standard operating conditions.

Table 38: KPIs registered in disruption scenarios for the AMT network model.

(Scenario) Time band	Average Waiting Time (min)	Train Saturation Level (%)	Average No. of Passengers Waiting	Total Throughput (# of Pax.)
(1C) Baseline peak 16–20	2.96	23.3%	195	24,726
(3B) Protest 8–12	2.88	28.8%	191	26,925
(3A) Match Day 16–24	3.88	19.6%	143	32,868

- The Protest scenario (8:00–12:00) generates a uniformly higher demand across the entire network, with Brin and Dinegro recording notably elevated throughput relative to their normal morning peak performance. Saturation rises to 28.8% (max 38.5%), yet waiting time remains essentially unchanged at 2.88 min, the system maintains service quality despite the demand surge.
- Match Day (16:00–24:00) produces the largest absolute throughput of any scenario: Brignole alone registers 7,280 passengers over 8 hours, a 46% increase over its standard afternoon-peak equivalent. However, demand is spread across 8 hours rather than concentrated in 4, so average saturation is lower (19.6%) than the standard peak. The critical indicator is the waiting time, which rises sharply to 3.88 min average with a maximum of 6.09 min, driven by late-evening service thinning after the event ends.
- Both disruption scenarios maintain zero overcrowding events and zero passengers left behind, demonstrating solid structural resilience of the AMT network even under significant demand spikes.

GoA4 Scenarios — Full Automation (Scenarios 4 and 5)

GoA4 (Grade of Automation 4) operates enables reduced headways of 120 seconds and 90 seconds. The peak reference for comparison is Scenario 1C (16:00–20:00).

Peak conditions

- Waiting time drops from 2.96 min to 2.01 min at 120s (−32%) and 1.93 min at 90s (−35%). Saturation falls from 23.3% to 10.0% and 5.9% respectively, and average occupancy drops from 67.6 to 29.0 and 17.0 passengers per train. The improvement is substantial and consistent across all KPIs.

- Throughput remains essentially unchanged (23,582–23,842 vs. 24,726 baseline), confirming that GoA4 does not expand total network capacity but redistributes the same demand across more frequent services, improving load distribution and comfort.
- The incremental gain from 120s to 90s headway is real but modest, the step change from baseline to 120s captures most of the benefit.

Table 39: KPIs registered in scenarios 4 and 5 (GoA4) for the AMT network model.

Scenario	Average Waiting Time (min)	Train Saturation Level (%)	Average Train Occupancy (# of Pax.)	Average No. of Passengers Waiting	Total Throughput (# of Pax.)	Ov. events
(1C) Baseline peak 16–20	2.96	23.3%	67.6	195	24,726	0
(4A) GoA4 120s peak 16–20	**2.01**	**10.0%**	**29.0**	**136**	**23,582**	**0**
(4C) GoA4 90s peak 16–20	**1.93**	**5.9%**	**17.0**	**124**	**23,842**	**0**
(3A) Match Day (Existing)	3.88	19.6%	56.9	143	32,868	0
(5A) Match Day 120s GoA4	**1.92**	**7.4%**	**21.6**	**87**	**33,453**	**0**
(5D) Match Day 90s GoA4	**1.84**	**5.6%**	**16.2**	**82**	**30,949**	**0**
(3B) Protest (Existing)	2.88	28.8%	83.4	191	26,925	0
(5B) Protest 120s GoA4	**1.92**	**12.5%**	**36.4**	**102**	**22,912**	**0**
(5E) Protest 90s GoA4	**1.86**	**10.1%**	**29.4**	**95**	**22,210**	**0**

Figure 5 compares train saturation levels during the 16:00–20:00 period for the current baseline operations and the two GoA4 headway configurations. Under current baseline conditions, saturation oscillates between 17% and 33%, which is well within operational comfort thresholds. GoA4 headways introduces a significant reduction in saturation, dropping it to 8–12% at 120-second headway and below 6% at 90-second headway. This is achieved by redistributing demand across a network of more frequent services. Both configurations represent a significant oversupply of capacity relative to actual demand, with the 90-second headway being particularly inefficient. Adopting a dynamic headway management approach, where shorter intervals are deployed only when real-time demand justifies it, would preserve the service quality benefits of GoA4 while avoiding the systematic underloading of trains.

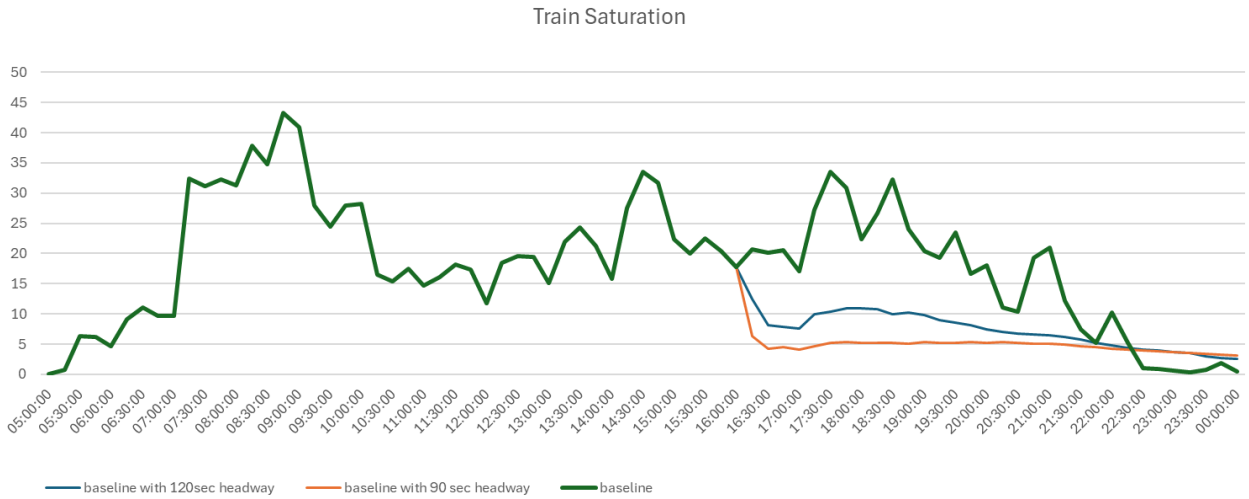


Figure 5: Train saturation levels between current and fully automated operations (Baseline)

Match Day + GoA4

- The impact of GoA4 is most visible under disruption. For the Match Day scenario, waiting time falls from 3.88 min (standard disruption) to 1.92 min at 120s headway and 1.84 min at 90s headway scenarios, representing a reduction of over 50%. Saturation drops from 19.6% to 7.4% / 5.6%, and passengers waiting is nearly halved (143 → 87 → 82).
- GoA4 does not increase total throughput in the Match Day scenario (it remains in the 31,000–33,000 range), but it dramatically reduces the peak-hour concentration of demand by serving passengers faster across more frequent services.

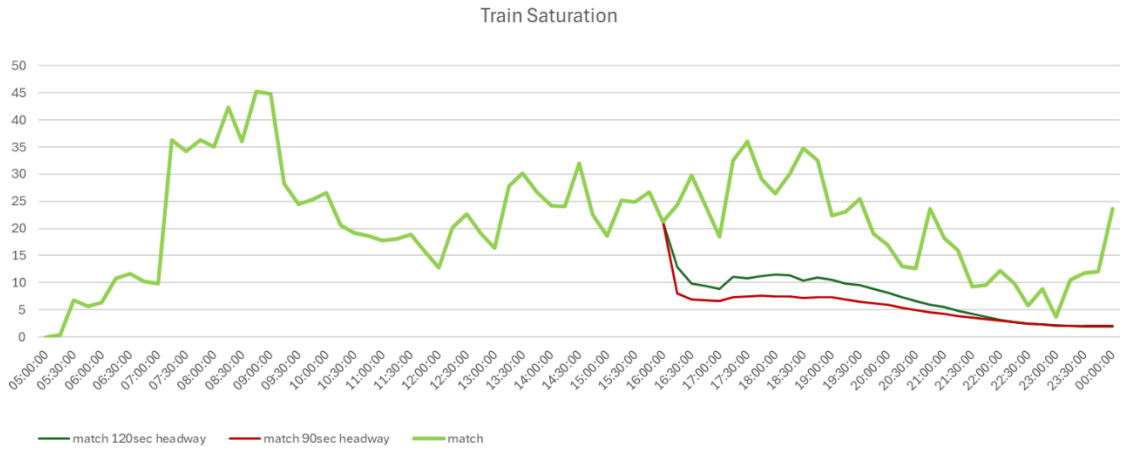


Figure 6: Train saturation levels between current and fully automated operations (Match Day)

The saturation pattern under Match Day conditions mirrors the dynamic observed in the standard peak scenario: GoA4 headways produce a sharp drop in train loading relative to the disruption baseline, with

the same oversupply effect already discussed. The same considerations around dynamic headway management apply.

Protest + GoA4

- For the Protest scenario, waiting time falls from 2.88 min to 1.92 min (120s) and 1.86 min (90s). Saturation goes from 28.8% to 12.5% / 10.1%. These gains are proportionally among the largest across all scenarios, driven by the fact that the protest generates concentrated demand in a short 4-hour window, precisely the condition where higher train frequency is most effective.
- The 90s headway offers a further meaningful reduction in saturation for the Protest case (12.5% → 10.1%), slightly more than in normal peak, suggesting that at this demand level the additional frequency still yields measurable benefit.

Figure 7: Train saturation levels between current and fully automated operations (Protest)

Similarly, the protest scenario exhibits the same pattern: despite the increased demand caused by the event, GoA4 headways reduce saturation well below baseline levels, resulting in trains running significantly underloaded.

2.3.1.2 METROS – SOFIA'S METRO

Baseline — Normal Operations (Scenarios 1 and 2)

Table 40: KPI Summary for Scenarios 1 and 2 (MetroS Network model)

(Scenario) Time Band	Average Wait Time (min)	Average Train Saturation Level (%)	Maximum Train Saturation Level (%)	Average Occupancy (# of pax.)	Average Number of Passengers Waiting
(2B) Off-peak 5:00–7:00	0.54	7.6%	34.1%	1,265	2,522
(1B) Peak 7:00–10:00	0.36	22.3%	59.1%	3,945	6,005
(2D) Off-peak 10:00–16:00	0.43	16.7%	35.8%	2,858	3,869
(1D) Peak 16:00–20:00	0.39	23.9%	44.0%	3,954	5,780

(2F) Off-peak 20:00–24:00	0.46	5.9%	16.2%	955	1,681

- Waiting time is extremely low across all time bands, never exceeding 32 seconds on average. This reflects the high service frequency supported by a multi-line network configuration. The early off-peak band (5:00–7:00) is a notable exception: despite having the lowest saturation (7.6%) and the fewest passengers waiting (avg 1,265), it records the highest average waiting time (0.54 min). This inversion suggests that early-morning service frequency is reduced, so passengers arriving at that time face slightly longer gaps between trains even with light loads.
- The two peak bands, morning (7:00–10:00) and afternoon (16:00–20:00), are near-identical in demand intensity: 3,945 vs. 3,954 average passengers waiting, and saturation of 22.3% vs. 23.9%. This symmetry between AM and PM peaks points to a balanced bidirectional commuting pattern rather than a single dominant flow direction.
- Saturation maximum values (59.1% in the morning peak, 44.0% in the afternoon peak) are notably higher than the averages, indicating that load is unevenly distributed across lines and stations within each band, some segments experience substantially higher strain than others.

Table 41: Network load distribution - Scenarios 1 and 2 (MetroS Network model)

(Scenario) Time Band	Line 1↑	Line 1↓	Line 4↑	Line 4↓	Line 2↑	Line 3↑	Line 3↓
(2B) 5:00–7:00	16.4%	20.3%	19.1%	12.2%	8.0%	10.8%	8.4%
(1B) 7:00–10:00	17.8%	15.1%	16.6%	12.3%	9.2%	12.4%	9.9%
(2D) 10:00–16:00	16.3%	14.4%	13.3%	12.3%	15.1%	13.2%	7.6%
(1D) 16:00–20:00	14.4%	9.8%	12.6%	13.6%	15.7%	15.7%	10.8%
(2F) 20:00–24:00	15.9%	7.3%	12.9%	11.9%	18.9%	16.2%	5.5%

- The Red line consistently carries the largest share of network load across almost all bands, particularly in its outbound direction during off-peak hours. The Yellow line is the second busiest. Green carries the smallest share throughout.
- During the peak 16:00–20:00 band, load distribution is relatively balanced compared to other bands, suggesting that demand is well-spread across lines in the afternoon. Late evening (20:00–24:00) shows a stronger concentration on Red outbound and Blue, pointing to specific residential corridors with heavier return-home flows.

Disruption Scenario — Serdica Station Closed (Scenario 3C)

The disruption simulated for MetroS is the closure of Serdica station during the 16:00–20:00 peak. Trains continue to run normally but do not stop at Serdica. Passengers already on board travelling through Serdica are unaffected. Only passengers who would have boarded or alighted at Serdica are impacted: they are redirected to adjacent stations, Opaltchenska (towards Airport / Business Park direction) or St. Kliment (towards Obelya / Slivnitsa direction).

Table 42: KPI Summary for Scenario 3 (MetroS Network model)

Scenario	Average Waiting Time (min)	Train Saturation Level (%)	Average No. of Passengers Waiting	Maximum No. of Passenger waiting	Time Band
(1D) Baseline peak 16–20	0.39	23.9%	3,954	5,780	16–20
(3C) Serdica closed	0.48	24.7%	2,308	3,427	16–20

- Waiting time increases by 23% (from 0.39 to 0.48 minutes), which is a moderate absolute increase, but a significant one in proportional terms for a system where waits of less than 0.4 minutes are the norm.
- The most counterintuitive result is the sharp drop in the number of passengers waiting (a decrease of 42%, from an average of 3,954 to 2,308). This is a direct consequence of how the disruption is modelled: since trains do not stop at Serdica, passengers who would have alighted there instead travel to the next station. This effectively removes a subset of boarding demand from Serdica's area of influence; passengers who cannot easily change their plans may switch to other modes of transport or delay their journey, thereby reducing the density of passengers at stations across the network. The redistribution to Opaltchenska and St. Kliment partially offsets this, but the net effect is a lower aggregate waiting count.
- Saturation increases only marginally (23.9% to 24.7%), confirming that the remaining lines can absorb the rerouted demand without experiencing significant capacity issues. The network's multi-line redundancy provides effective resilience against the closure of a single station of this type.
- Network load analysis shows that, during the disruption, the red and yellow lines absorb the largest share of the redistributed flows. Their combined share rises to approximately 49–53%, compared to ~30% in normal peak conditions.

GoA4 Scenarios — Full Automation (Scenarios 4 and 5)

GoA4 automation with 120-second and 90-second headways is evaluated for the 16:00–20:00 peak band, both in normal conditions (Scenarios 4B and 4D) and in the Serdica-closed disruption (Scenarios 5C and 5F).

Table 43: KPI Summary for Scenarios 4 and 5 (MetroS Network model)

Scenario	Average Waiting Time (min)	Train Saturation Level (%)	Average # Pax. Waiting	Maximum # Pax. Waiting	vs previous step
(1D) Baseline peak 16–20	0.39	23.9%	3,954	5,780	—
(4B) GoA4 120s peak 16–20	0.22	12.1%	2,125	3,030	−43% wait | −49% sat. | −46% p.wait
(4D) GoA4 90s peak 16–20	0.21	10.5%	2,070	2,772	−5% wait | −13% sat. | −3% p.wait vs 120s
(3C) Serdica closed (no GoA4)	0.48	24.7%	2,308	3,427	—

(5C) GoA4 120s + Serdica closed	0.30	13.7%	1,284	2,008	−38% wait | −45% sat. | −44% p.wait vs Sc.3c
(5F) GoA4 90s + Serdica closed	0.27	13.5%	1,181	1,896	−10% wait | −1% sat. | −8% p.wait vs 120s

Note: ^incremental improvement (i.e. 1D to 4B; 4B to 4D; 3C to 5C; 5C to 5F)

Peak conditions:

- The improvement from standard operation to GoA4 120-second headway is substantial across all KPIs: waiting time falls from 0.39 to 0.22 minutes (−43%), saturation falls from 23.9% to 12.1% (−49%), and the number of passengers waiting falls from 3,954 to 2,125 (−46%). At an average wait time of 0.22 minutes, passengers perceive train availability as essentially continuous.
- The incremental step from 120-second to 90-second headway delivers modest additional gains: -5% in waiting time, -13% in saturation and -3% in passengers waiting. The main benefits are already achieved with 120-second headways; the 90-second configuration provides a narrower margin of improvement, suggesting diminishing returns at that frequency level under normal demand.

GoA4 under disruption — Serdica closed

- The most operationally significant result is that GoA4 at 120s headway during the Serdica closure (Scenario 5C) outperforms the standard baseline on every KPI: waiting time 0.30 min vs. 0.39 min (normal), saturation 13.7% vs. 23.9%, passengers waiting 1,284 vs. 3,954. Higher train frequency more than compensates for the demand redistribution caused by the station closure.
- The step from GoA4 120s to 90s under disruption yields minimal additional improvement (−10% wait, −1% saturation, −8% passengers waiting). Unlike in the Protest scenario for AMT where 90s still adds meaningful value, here the bottleneck has shifted from train frequency to the passenger redistribution mechanism itself — once frequency is high enough, further headway reduction has diminishing returns.
- This finding highlights the interaction between network topology and automation benefits: in a multi-line hub network, GoA4 is most effective as a disruption-mitigation tool when the disruption affects a single node rather than the entire line, because the surrounding network can absorb redistributed demand if served at sufficient frequency.

2.3.2 METRO LINES AND STATION LEVEL EVALUATIONS

This section describes the key observations and findings at mesoscopic level, focusing on *individual metro lines and/or station level impacts, considering the operation of the whole metro network* of AMT and MetroS. The analysis of passenger and metro line specific performance indicators are carried out by running the simulation model for the specified time periods across all the scenarios as described under **Table 35**.

2.3.2.1 STATION LEVEL METRO SYSTEM PERFORMANCES

This section includes the evaluation of the key KPIs of average wait times, average queue lengths and average travel times of passengers at station level across the whole metro network. These KPIs are measured based on the following conceptualization.

- ***Average waiting time***: it is expressed in minutes and is estimated from the instance the passengers enter the platform till they are boarded into the trains. Thus, it is calculated as an average value at a specific station, based on all operations of trains during the analysis time period and from different simulation runs.
- ***Average queue length***: it is expressed in numbers, in terms of the number of passengers gathering at the platform till the arrival of train and getting boarded. Thus, it is calculated as an average value at a specific station, based on all operations of trains during the analysis time period and from different simulation runs.
- ***Average travel time***: it is expressed in minutes and is estimated as the time taken for all passengers travelling from all origin stations to the subject destination station. The travel time includes the wait time, actual travel time with train as well as boarding and alighting times. Thus, it calculated as an average value at a specific destination station, based on the travel times of all the passengers alighting at the same destination station from all operations of trains during the analysis time period and from different simulation runs.

For brevity, only KPIs for two stations are discussed (i.e., Principe station of AMT and Serdica station with Lines 1 and 4 of MetroS). The key findings are reported under **Table 44** and **Table 45** at the Principe and Serdica stations for the operation of AMT and MetroS, respectively in both directions of train travel. All other stations of respective metro lines for both the AMT and MetroS represent similar KPI behaviours as highlighted in **Table 44** and **Table 45**. It is important to note that, Serdica station is inside the shared section of Line 1 and Line 4, therefore the KPIs represented by **Table 45** reflects the combined effects over the passenger performances through the operation of both lines.

Table 44: Principe Station results considering the AMT network operations

KPIS	1A	1C	2A	2C	2E	3A	3B	4A	4B	5A	5B	5D	5E
Brignole To Brin Direction													
Avg. Waiting Time	3.52	3.16	7.95	3.39	5.00	3.41	3.37	0.93	0.68	0.76	0.83	0.59	0.64
Avg. Queue Length	9	7	1	5	2	4	7	2	1	1	2	1	1
Avg. Travel Time	11.4	9.46	14.82	9.86	11.28	10.08	10.58	7.04	6.88	7.24	8.12	7.04	7.85
Brin to Brignole Direction													
Avg. Waiting Time	3.29	3.14	7.11	3.01	6.01	3.66	3.23	0.78	0.53	0.75	0.76	0.53	0.57
Avg. Queue Length	22	13	1	10	5	9	20	3	2	2	5	1	4
Avg. Travel Time	7.61	6.58	9.85	6.69	9.11	7.02	6.96	4.27	4.01	4.37	4.51	4.11	4.25

Table 45: Serdica Station results considering the MetroS Line 1 and Line 4 network operations

KPIs	1B	1D	2B	2D	2F	3C	4C	4D	5C	5F
Obelya/Slivnitsa to Sofia Airport/Sofia Business Park Direction										
Avg. Waiting Time	3.29	3.35	4.18	3.58	4.91	0	1.59	1.24	0	0
Avg. Queue Length	82	95	4	55	27	0	61	49	0	0
Avg. Travel Time	10.08	9.65	11.15	10.14	10.5	0	8.44	8.23	0	0
Sofia Airport/Sofia Business Park to Obelya/Slivnotsa Direction										
Avg. Waiting Time	2.83	2.52	4.12	2.67	4.29	0	1.36	1.24	0	0

Avg. Queue Length	61	73	3	38	24	0	40	45	0	0
Avg. Travel Time	17.75	17.98	19.61	18.39	19.59	0	16.48	16.17	0	0

The average wait times of the passengers at both considered stations are highly dependent on the hours of the day and service frequencies of the metro operations as well as the type of disruption considered. During the peak hours (i.e., Scenarios 1A and 1C in **Table 44** for AMT as well as 1B and 1D in **Table 45** for MetroS), the average wait time values are lower than the off peak hour periods (i.e., Scenarios 2A , 2C and 2E in **Table 44** for AMT as well as 2B, 2D and 2F in **Table 45** for MetroS). This is mainly due to lower service frequency of the metro operations during the off-peak hour periods as compared to the peak hour operations. Furthermore, no significant impacts of disruptions over the average waiting time of the passengers (i.e., Scenarios 3A and 3B in in **Table 44** for AMT as well as 3C in **Table 45** for MetroS) are observed while comparing them with peak hour operations.

Furthermore, average queue lengths are longer during peak hours when comparing with their off-peak counterparts since the OD passenger demand is higher during peak hours. Finally, passenger travel times are shown to be dependent upon the average waiting times as highlighted by the entries in both **Table 44** and **Table 45** under Scenarios 1 and 2. To this effect, both the average queue lengths and passenger travel times decrease while considering peak periods under Scenario 1, 2 and 3 as evident by both **Table 44** and **Table 45**.

According to Tables 45 and 45, switching the operations from the existing mode to GoA4 operations under Scenarios 4 and 5 results in a huge reduction in average waiting times and a slight reduction in average travel times for passengers, either during disruptions (i.e. Scenario 3) or normal operations (i.e. Scenarios 1 and 2). This effect is due to the increased service frequency of metro operations under Go4 operations as compared to the current operations. Furthermore, the reduction in both the average waiting times and passenger travel times is more pronounced when the service frequency is further enhanced by shortening train headways from 120 to 90 seconds. This impact can be better analysed by comparing the KPIs of Scenarios 4A vs 4B, 5A vs 5D and 5B vs 5E as reported under **Table 44** for AMT. The same trend can also be observed by comparing the KPIs from cenarios 4C vs 4D and 5C vs 5E for MetroS.

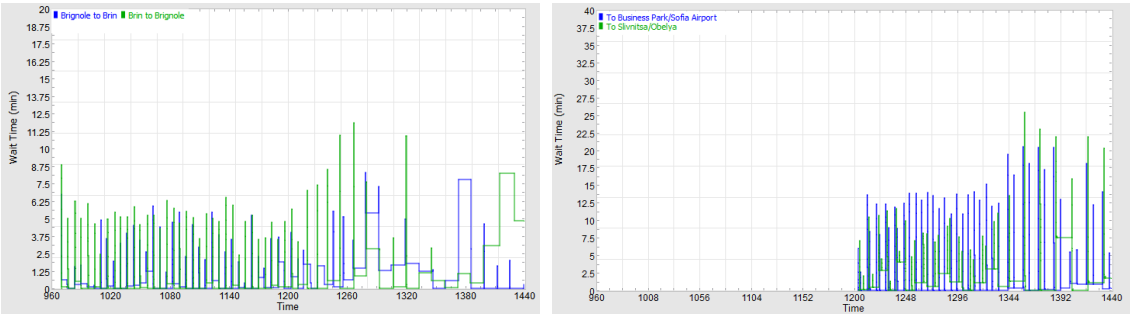


Figure 8: Wait time (3A: Principe (Left); and 3C: Serdica, MetroS Lines 1&4 (Right))

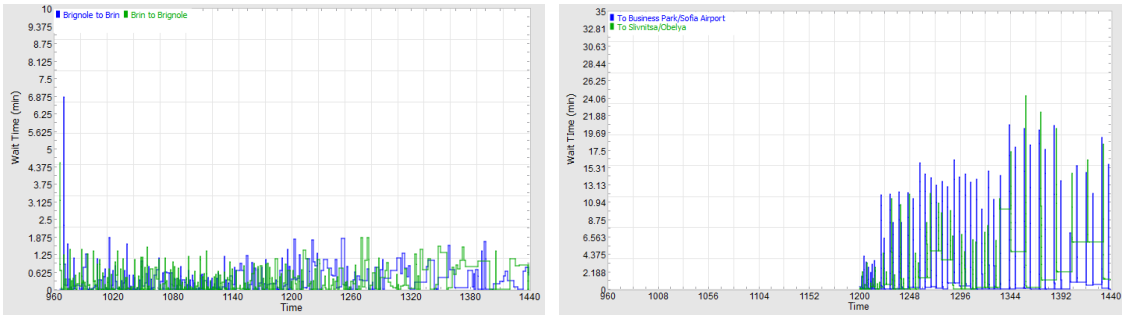


Figure 9: Wait time (5A: Principe (Left); and 5C: Serdica, MetroS Lines 1&4 (Right))

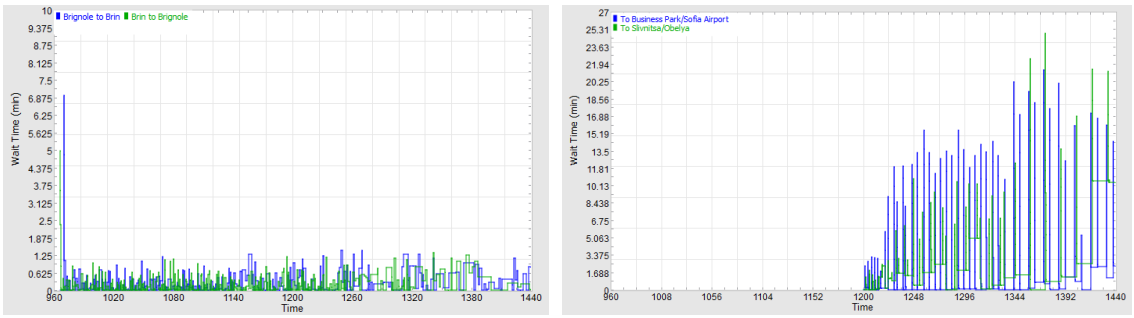


Figure 10: Wait time (5D: Principe (Left); and 5F: Serdica, MetroS Lines 1&4 (Right))

A further comparison of all disruption scenarios of AMT and MetroS networks are drawn from **Figure 8** to **Figure 10**, highlighting the impacts of GoA4 operations over the current operations while considering waiting times of passenger from the start of disruption till the end of daily operations.

A deeper analysis of **Figure 8** to **Figure 10** confirms the same findings as reported in **Table 44** and **Table 45**, during different time periods of daily operations for both networks in both directions. Similarly, the passenger waiting times in **Figure 8** are consistently longer between 16:00-24:00 under Scenario 3A and between 20:00-24:00 under Scenario 3C at Principe and Serdica stations, respectively. It is worth noting that from **Figure 8** to **Figure 10** that passenger waiting times from 16:00-20:00 at Serdica, MetroS is zero. This is due to the closure of Serdica stations during disruption periods from 16:00-20:00. Thus, to understand impacts of the disruption, the simulation time period is extended from 20:00 till the end of the daily operations. To this effect, the waiting time in **Figure 8** to **Figure 10** starts to rise after the system is restored to normal operations at Serdica stations starting 20:00.

Furthermore, the extremely low waiting times are observed while switching from the existing to GoA4 operations as highlighted by **Figure 9** and **Figure 10** during 16:00-20:00. Similarly, the reduction in the wait times is more prominent during the same time period when comparing the **Figure 9** with **Figure 10**. These trends can only be visualised at Principe of AMT as Serdica of MetroS is closed for all the passengers between 16:00-20:00. It is important to note that in both **Figure 9** and **Figure 7** at Principe station, passenger waiting times are longer at the beginning phase of the simulation but then drops immediately. These longer wait times represent the transition period from the ending of existing operations (i.e., Scenario 1, 2, and 3) to the beginning of GoA4 operations (i.e., Scenarios 4 and 5).

It is interesting to note from **Table 45** and **Figure 8** to **Figure 10** that the KPI values are all zero at Serdica station while considering disruption scenarios under the existing and GoA4 operations. This is due to the closure of Serdica during the disruption period 16:00-20:00. To this effect, the potential impacts of the disruptions spreading over to nearby stations are also examined in terms of the queue lengths when extending the simulation from 16:00 till the end of daily operations.

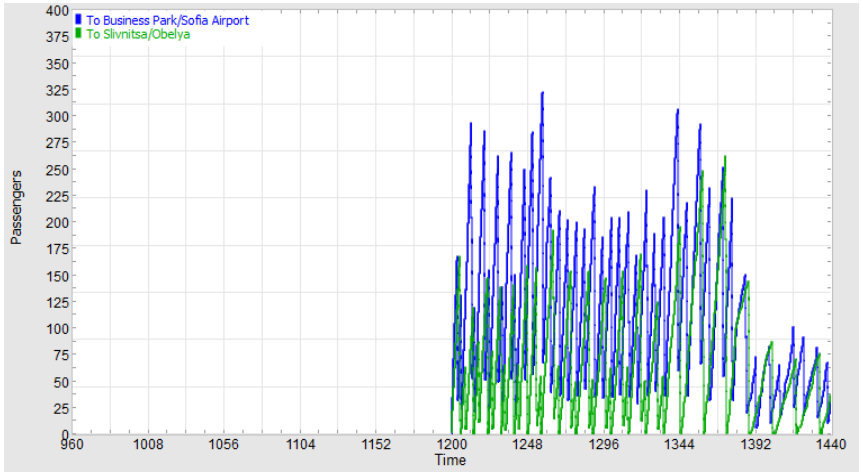


Figure 11: Queue lengths (3C: Serdica, MetroS, both directions, 16:00-24:00)

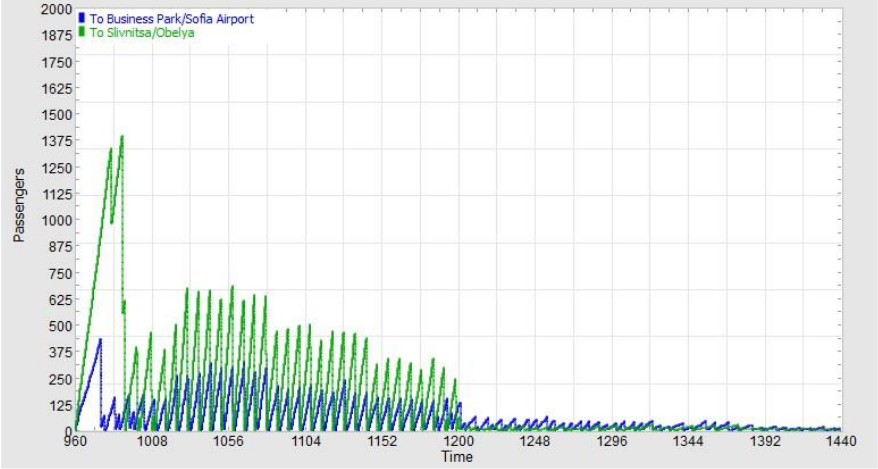


Figure 12: Queue lengths (3C: Opalchenska, MetroS, both directions, 16:00-24:00)

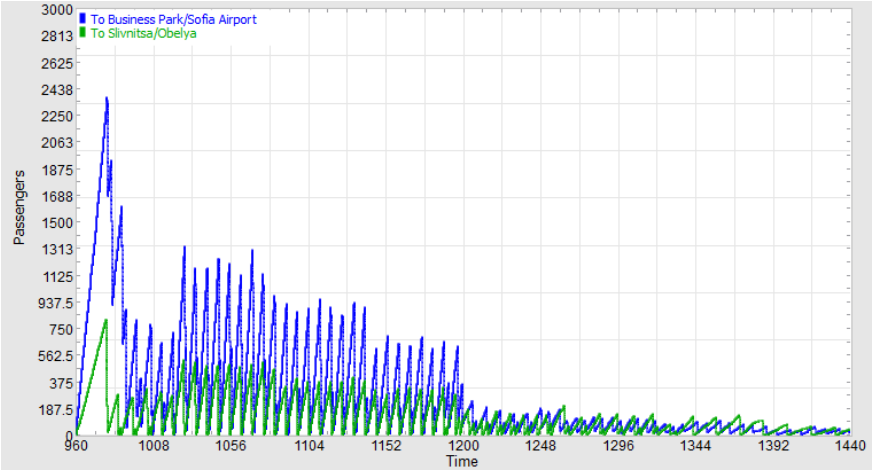


Figure 13: Queue lengths (3C: SU Sv. Kliment Ohridski, MetroS, both directions, 16:00-24:00)

According to **Figure 11** to **Figure 13**, the OD passenger demand at Serdica station is redistributed to the nearby stations (i.e., Opalchenska and SU Sv. Kliment Ohridski) during the disrupted period 16:00-20:00. Similarly, as soon as the Serdica station closes at 16:00, the peaks of the passenger queues are observed at Opalchenska station (**Figure 12**) and SU Sv. Kliment Ohridski station (**Figure 13**). Furthermore, the passenger queue peaks at these nearby stations are constantly higher from 16:00 till 20:00 which gets normalised after 20:00. These peaks are observed while considering the operations of both metro Lines 1 and 4 in both directions of travel. This is due to the restoration of normal operations at Serdica after 20:00 as reflected in **Figure 11** where the queue length of passengers starts to rise after 20:00 till the end of daily operations.

2.3.2.2 NETWORK LEVEL PERFORMANCE OF METRO LINES

This section includes the evaluation of the key KPIs of avg. occupancy, max. resource utilised, and av. operations at network levels of entire metro network. These KPIs are measured based on the following conceptualization.

- ***Avg. occupancy***: it is expressed in percentage in terms of the capacity utilisation of trains once they leave a station after completing boarding/alighting operations and are travelling on a link between two stations. Thus, it is calculated as an avg. value for all operations of a specific metro line, based on the occupancy value at all the links during the analysis time period and from different simulation runs.
- ***Max. resource utilised***: it is expressed in numbers in terms of the number of train resources required to complete the journey from an origin to a destination and then back to the origin. Thus, it calculated as a max. value based on all operations of trains during the analysis time period and from different simulation runs.
- ***Avg. Operations***: it is expressed in numbers in terms of the total number of train services being operated. Thus, it calculated as an avg. value of train services from different simulation runs during the analysis time period.

Table 46: Network level KPI results for AMT (All Scenarios)

KPIS	1A	1C	2A	2C	2E	3A	3B	4A	4B	5A	5B	5D	5E
Brignole To Brin Direction													
Avg. Occupancy	22.48	22.89	4.65	21.44	7.79	21.78	20.38	8.79	66.50	5.27	6.11	3.94	4.55
Max. Resources Utilised	3	4	3	3	5	5	3	11	15	11	12	15	15
Avg. Operations	27	37	9	54	21	57	36	121	160	240	121	320	161
Brin to Brignole Direction													
Avg. Occupancy	37.44	23.96	10.37	20.38	6.18	19.66	29.69	7.14	5.36	4.49	9.1	3.38	6.79
Max. Resources Utilised	4	3	3	4	2	3	4	10	13	10	10	13	13
Avg. Operations	27	36	10	54	18	54	36	121	160	240	120	320	160

Table 47: Network level KPI results for MetroS Line 1 (All Scenarios)

KPIS	1B	1D	2B	2D	2F	3C	4C	4D	5C	5F
Slivnitsa to Sofia Business Park Direction										
Avg. Occupancy	21.45	18.81	11.03	16.14	9.15	20.01	15.01	8.74	15.04	8.81
Max. Resources Utilised	6	5	6	5	5	5	10	13	9	12
Avg. Operations	28	32	11	41	20	32	60	80	60	80
Sofia Business Park to Slivnitsa Direction										
Avg. Occupancy	27.20	27.83	8.03	11.41	8.29	28.51	15.79	15.13	15.33	15,86
Max. Resources Utilised	7	6	3	6	4	6	10	13	9	12
Avg. Operations	27	33	10	41	19	33	60	80	60	80

The key findings are reported under **Table 46** to **Table 50** reflect the metro line level impacts for the single metro line of AMT and all four lines of MetroS. These statistics have been reported while considering the metro operations in both directions of travel. The KPIs as reported under **Table 46** and **Table 50** not only provide the performance indicators for evaluating the individual metro line but also highlights the important informatics for the operators and stakeholder to support the strategic decisions.

Table 48: Network level KPI results for MetroS Line 2 (All Scenarios)

KPIS	1B	1D	2B	2D	2F	3C	4C	4D	5C	5F
Obelya to Vitosha Direction										
Avg. Occupancy	27.40	17.92	11.31	16.89	4.31	18.13	5.09	3.83	5.06	3.84
Max. Resources Utilised	6	4	7	5	4	4	14	20	14	19
Avg. Operations	30	34	14	41	23	34	120	160	120	160
Vitosha to Obelya Direction										
Avg. Occupancy	17.61	25.47	4.69	16.59	7.88	24.97	7.12	4.99	6.63	5.01
Max. Resources Utilised	4	5	3	5	4	5	15	19	15	19
Avg. Operations	29	34	12	41	23	34	121	161	121	161

Table 49: Network level KPI results for MetroS Line 3 (All Scenarios)

KPIS	1B	1D	2B	2D	2F	3C	4C	4D	5C	5F
Hadzhi Dimitar to Vitosha Direction										
Avg. Occupancy	26.53	18.82	8.37	15.52	4.49	18.52	7.4	5.53	7.27	5.59
Max. Resources Utilised	6	7	5	5	7	7	15	19	15	19
Avg. Operations	37	47	15	60	32	47	120	160	120	160
Vitosha to Hadzhi Dimitar Direction										
Avg. Occupancy	18.92	26.99	5.47	15.94	8.68	26.92	10.6	7.38	9.82	7.45
Max. Resources Utilised	7	5	5	6	5	5	14	19	14	19
Avg. Operations	37	47	15	61	32	47	120	160	120	160

Table 50: Network level KPI results for MetroS Line 4 (All Scenarios)

KPIS	1B	1D	2B	2D	2F	3C	4C	4D	5C	5F
Obelya to Sofia Airport Direction										
Avg. Occupancy	29.10	20.26	7.03	20.48	8.44	35.52	13.18	11.33	14.1	13.01
Max. Resources Utilised	9	7	8	7	6	7	13	16	12	16
Avg. Operations	28	33	12	41	20	33	60	81	60	81
Sofia Airport to Obelya Direction										
Avg. Occupancy	17.19	24.99	6.54	20.89	7.48	25.87	13.76	6.93	13.69	6.69
Max. Resources Utilised	9	7	2	7	6	7	12	16	12	16
Avg. Operations	28	33	9	41	20	33	60	81	60	81

Firstly, average occupancy values are determined based on metro lines capacity utilisation status while considering all operations during the time period being analysed. A deeper analysis of **Table 46** to **Table 50** reveals that the average occupancy values are highly dependent on the frequency of trains and slightly on the passenger demands during the specific time period of the day. Similarly, average occupancy values can either decrease or increase while considering different off-peak or peak hour periods since the passenger demand patterns as well as the service frequencies vary during different time of the day. This trend can be observed while comparing the Scenarios 1A, 1C, 2A, 2C, and 2E of AMT (**Table 46**) as well as Scenarios 1B, 1D, 2B, 2D, and 2F for all lines of MetroS (**Table 47** to **Table 50**). However, the average occupancy values increase during the disruption periods (i.e., Scenarios 3A and 3B of AMT (**Table 46**) and Scenario 3C for all lines of MetroS (**Table 47** to **Table 50**). This is due to the high passenger demand and redistribution of this demand to nearby stations, during the disruption scenarios of AMT and MetroS, respectively.

However, once the existing operations (i.e., Scenarios 1, 2, and 3) of both metro networks are switched to GoA4 operations, the average occupancy values would be further reduced under both the normal GoA4 operations (i.e., scenario 4) as well as GoA4 operations during disruptions (i.e., Scenario 5). This is mainly due to the high service frequency where the trains operate at reduced headway meaning that the train resources are not fully utilised. Such an impact can be visualised by analysing the average occupancy values under Scenarios 4A and 4B along with comparing the Scenarios 3A vs 5A and 5B;

as well as 3B vs 5D and 5E, of AMT (**Table 46**). Similar trends can also be observed by visualising the average occupancy values under Scenarios 4C and 4D along with comparing the Scenarios 3C vs 5C and 5F, for all lines of MetroS (**Table 47** to **Table 50**). The average utilisation levels could drop further when headways are reduced from 120 to 90 seconds under the normal GoA4 operations (i.e., Scenario 4) and GoA4 operations during disruptions (i.e., Scenario 5). This can be observed by comparing Scenarios 4A vs 4B, 5A vs 5B, and 5D vs 5E of AMT (**Table 46**) as well as Scenarios 4C vs 4D and 5C vs 5F for all lines of MetroS (**Table 47** to **Table 50**).

To summarise, the average occupancy values shown in **Table 46** to **Table 50** for both metro networks, highlight some important evidence supporting stakeholders, policy makers, and the operators' strategic decision in determining the optimal headways for the concerned metro systems while switching from the existing operations to GoA4 operations. In doing so, a suitable trade-off between either focusing on reducing the passenger waiting times or improving the capacity utilisation of metro services (towards efficient resource utilisation) can be selected. Finally, **Table 46** to **Table 50** also highlight the key information related to the number of train resources required across both the scenarios of current operations (i.e., Scenarios 1, 2 and 3) and GoA4 operations (i.e., Scenarios 4 and 5) to efficiently operate metro services depending on the type of operations during different time of the day. As an additional performance indicator, the last rows of **Table 46** to **Table 50** for both metro networks provide the number of operations which are also dependent upon the considered time periods of the day and the service frequency of metro system during different scenarios. The average number of operation values is not significantly affected by disruption scenarios for both the AMT and MetroS networks. However, their operations under GoA4 operational scenario substantially increase their average number of train services. This is due to the improved frequency of metro service when switching from current operations (i.e., Scenarios 1, 2 and 3) to GoA4 operations (i.e., Scenarios 4 and 5).

2.3.3 INDIVIDUAL STATION CROWD MANAGEMENT LEVEL EVALUATIONS

This section will adopt a crowd management lens to reveal impacts at individual station level passenger movements due to various scenarios, using results from metro station simulation models developed for the two selected stations at De Ferrari for AMT and Serdica for MetroS. The evaluation will mainly focus on Scenarios 1, 2 and 4. Since the impacts to be evaluated from Scenarios 3 and 5 (representing large scale disruptions to the metro networks) are expected to be mostly at network or metro line levels, these scenarios will be excluded from impact evaluations in this section. As the nature of the two selected stations (De Ferrari and Serdica) and their corresponding metro network complexity are quite different from each other, the aspects and model output metrics to be used for crowd management level evaluations will be slightly different as well.

2.3.3.1 DE FERRARI STATION EVALUATIONS

The De Ferrari station serves the one-and-only-one metro line of the AMT metro network, where the passenger volume and characteristics are much simpler. **Table 51** shows the platform level model output statistics, including the average number of passengers, passenger in-station time, average queue length, and average queue time.

Table 51: De Ferrari Station KPI Results Summary

	1A	1C	2A	2C	2E	4A – 120s	4C – 90s
Direction: To Brin							
Ave No. of Pax / hour	127.7	**319.3**	1.1	263.2	17.7	**349.5**	**351.8**
Time in Station (min)	7.6	**14.2**	5.1	12.1	7.1	**7.4**	**7.1**
Ave Queue Time (min)	4.5	**8.9**	2.1	8.1	4.0	**3.1**	**2.9**
Ave Queue Length (# pax.)	4.0	**40.0**	0.4	24.4	7.5	**16.5**	**15.2**
Direction: To Brignole							
Ave No. of Pax / hour	33.8	**137.8**	10.0	79.8	5.7	**107.5**	**105.3**
Time in Station (min)	5.2	**8.9**	5.0	6.2	9.7	**5.1**	**5.1**
Ave Queue Time (min)	2.1	**4.2**	2.0	2.6	6.5	**0.8**	**0.8**
Ave Queue Length (# pax.)	0.5	**9.7**	0.1	2.4	1.7	**1.4**	**1.4**

Existing Performance (Scenarios 1 and 2)

Examining the performance of the current metro system at the De Ferrari station (Scenarios 1A, 1C, 2A, 2C, and 2E), the following key observations could be revealed:

- The “to Brin” direction is clearly much busier than the “to Brignole” direction with a significantly higher passenger volume as reflected in the number of passengers using the platform per hour.
- The afternoon peak (i.e. Scenario 1C: 16:00-20:00) is the most critical period across the whole range of metrics summarised.
- There are on average more than 350 passengers using the “To Brin” platform, which is 2.5 times of the morning peak. This represents an over-crowded condition comparing the available platform capacity of 200 passengers (i.e. a load factor of about +160%)
- The afternoon peak scenario revealed a significantly longer queue and queuing time than any other scenarios from the station model. It indicates that the afternoon peak period is the most congested period across the day where the longest queue (40.0) and queuing time (up to 8.9 minutes) are observed.
- The next busiest period is the inter-peak period (Scenario 2C: 10:00 to 16:00) where the average queue length was 24.4, and the queuing time was 8.1 minutes which was similar to that for the afternoon peak.
- On average, passengers will be able to get served on the platform (i.e. get on board of the train) within 15 minutes during the peak period, and within 5 to 10 minutes most of the time outside the peak.

The above revealed that the current system at De Ferrari station has performed satisfactorily in serving passengers during most of the time on typical day. The most critical period has been the afternoon peak period (16:00-20:00) where the station might be operating at an over-crowded condition. Passengers would be able to get served within 15 minutes during the peak period, and within 5 to 10 minutes during most of the time outside the peak.

Impacts of Fully Automated Operations GoA4 (Scenario 4)

To reveal the station level impact of fully automated operations, model evaluation results simulating the impact of shortening peak period train headways to 120 and 90 seconds (Scenario 4A and 4C respectively) were examined. Comparing the results of Scenario 1C (current operation during 16:00 to 19:00) with Scenarios 4A (GoA4 at 120-second headway) and 4C (GoA4 at 90-second headway),

Figure 14: Simulated hourly average queue time, queue length and in-station time (2C, 4A, 4C)

it is evident that:

- Average queue lengths and queuing time have been significantly reduced by more than 60% at “To Brin” direction and over 80% at “To Brignole” direction.
- Whilst both GoA4 scenarios showed significant improvements over the existing operations, the effect of further shortening train headways from 120 to 90 seconds was not significant.

Figure 14 also shows the same results where the afternoon peak queue length, queue time and in-station time have been significantly reduced in the two GoA4 scenarios.

2.3.3.2 SERDICA STATION EVALUATIONS

The Serdica station serves multiple metro lines of the MetroS network. It is located at the heart of the Sofia, which is a major passenger interchange station between Lines 1 and 4, and Line 2, with significantly large volume of passenger flows. **Table 52** shows the platform level model output statistics, including the average number of passengers, passenger in-station time, average queue length, and average queue time.

Existing Performance (Scenarios 1 and 2)

Examining the performance of the current metro system at the De Ferrari station (Scenarios 1B, 1D, 2B, 2D, and 2F), the following key observations could be revealed:

- The afternoon peak (i.e. Scenario 1D: 16:00-20:00) is the most critical period across all three metro lines with respect to the whole range of metrics summarised.
- There are on average more than 3000 passengers using the platform for Line 2 during the afternoon peak, which is slightly busier than the morning peak. This represents an over-crowded condition comparing the available platform capacity of 3520 (i.e. 1760x2) passengers (i.e. a load factor of around +174%)
- As for Lines 1 and 4, there are on average more than 6200 passengers using the shared platform during the afternoon peak. This represents a close to capacity congested condition (i.e. around 86% of the available platform capacity of 7200).
- The afternoon peak scenario revealed a significantly longer queue and queuing time than any other scenarios. It indicates that the afternoon peak period is the most congested period across the day where the longest queue (166 for Line 2, around 80 for Line 1 and 4).
- The next busiest period is the morning peak (Scenario 1B: 07:00-10:00) where the average queue length was around 80 for Line 2, and around 25 to 40 for Lines 1 and 4.
- As for the average queue time, non-peak periods generally have longer queue time, probably due to the longer train headways outside the peak periods.
- On average, passengers will be able to get served on the platform (i.e. get on board of the train) in no more than 12 minutes.
- The above revealed that the current system at the Serdica station has performed satisfactorily in serving passengers during most of the time for a typical day. The most critical period has been the afternoon peak period (16:00-20:00) where the station might be operating at an over-loaded condition. On average, passengers would be able to get served in no more than 12 minutes.

Table 52: Serdica Station KPI Results Summary

	1B	1D	2B	2D	2F	4B – 120s	4D – 90s
Direction: Line 2 (to Vitosha)							
Ave No. of Pax / hour	2794	**2949**	593	1923	740	**2857**	**2898**
Time in Station (min)	6.8	**7.4**	7.7	7.4	7.9	**5.3**	**5.0**
Ave Queue Time (min)	2.7	**3.3**	3.5	3.5	4.0	**1.3**	**1.0**
Ave Queue Length	64.8	**137.1**	13.2	90.1	56.6	**44.2**	**35.8**
Direction: Line 2 (to Obelya)							
Ave No. of Pax / hour	2625	**3175**	513	1841	756	**3205**	**3262**

Time in Station (min)	7.4	**7.6**	11.6	7.5	8.0	**4.9**	**4.8**
Ave Queue Time (min)	3.3	**3.5**	7.5	3.6	4.1	**0.9**	**0.8**
Ave Queue Length	80.3	**165.9**	24.0	93.9	67.3	**47.6**	**41.3**
Direction: Line 4 (to Obelya)							
Ave No. of Pax / hour	968	**1409**	215	799	395	**1430**	**1399**
Time in Station (min)	7.5	**7.2**	11.3	7.4	8.6	**5.4**	**5.0**
Ave Queue Time (min)	2.7	**3.0**	6.8	3.2	4.9	**1.8**	**1.4**
Ave Queue Length	26.0	**62.0**	8.7	36.7	35.4	**46.6**	**31.9**
Direction: Line 4 (to Sofia Airport)							
Ave No. of Pax / hour	1351	**1718**	291	1015	489	**1726**	**1691**
Time in Station (min)	7.0	**8.2**	7.9	9.0	9.5	**5.8**	**5.3**
Ave Queue Time (min)	1.9	**3.5**	3.3	4.5	5.5	**1.8**	**1.4**
Ave Queue Length	24.0	**84.7**	5.9	58.7	46.8	**58.6**	**38.8**
Direction: Line 1 (to Silv)							
Ave No. of Pax / hour	990	**1421**	198	811	381	**1364**	**1366**
Time in Station (min)	8.3	**8.1**	10.7	8.4	7.2	**5.2**	**4.8**
Ave Queue Time (min)	3.6	**3.8**	6.8	4.3	3.5	**1.6**	**1.2**
Ave Queue Length	33.2	**78.5**	8.9	47.9	27.9	**43.9**	**22.2**
Direction: Line 1 (to Business Park)							
Ave No. of Pax / hour	1349	**1659**	261	1018	461	**1685**	**1684**
Time in Station (min)	8.7	**7.6**	11.1	8.5	9.0	**5.9**	**5.5**
Ave Queue Time (min)	3.6	**2.9**	6.5	4.0	5.0	**1.9**	**1.5**
Ave Queue Length	44.0	**73.0**	12.1	56.3	45.4	**59.2**	**35.1**

Impact of Fully Automated Operations GoA4 (Scenario 4)

To reveal the station level impact of fully automated operations, model evaluation results simulating the impact of shortening peak period train headways to 120 and 90 seconds (Scenario 4B and 4D respectively) were examined. Comparing the results of Scenario 1D (current operation during 16:00-20:00) versus Scenarios 4B (GoA4 at 120-second headway) and 4D (GoA4 at 90-second headway),

Table 53: Serdica Station GoA4 Impact KPI Results Summary

	4B – 120s	4D – 90s	4B vs 4D	4B – 120s	4D – 90s	4B vs 4D
	Direction: Line 2 (to Vitosha)			**Direction: Line 2 (to Obelya)**		
Ave No. of Pax / hour	-3.1%	-1.7%	-1.4%	0.9%	2.8%	-1.8%
Time in Station (min)	-28.2%	-32.0%	5.6%	-35.5%	-36.8%	2.1%
Ave Queue Time (min)	-61.9%	-71.2%	32.5%	-74.3%	-77.1%	12.5%
Ave Queue Length	-67.8%	-73.9%	23.5%	-71.3%	-75.1%	15.3%
	Direction: Line 4 (to Obelya)			**Direction: Line 4 (to Sofia Airport)**		
Ave No. of Pax / hour	1.5%	-0.7%	2.2%	0.5%	-1.6%	2.1%
Time in Station (min)	-25.0%	-30.6%	8.0%	-29.3%	-35.4%	9.4%
Ave Queue Time (min)	-40.0%	-53.3%	28.6%	-48.6%	-60.0%	28.6%
Ave Queue Length	-24.8%	-48.5%	46.1%	-30.8%	-54.2%	51.0%
	Direction: Line 1 (to Silv)			**Direction: Line 1 (to Business Park)**		
Ave No. of Pax / hour	-4.0%	-3.9%	-0.1%	1.6%	1.6%	0.0%

Time in Station (min)	-35.8%	-40.7%	8.3%	-22.4%	-27.6%	7.3%
Ave Queue Time (min)	-57.9%	-68.4%	33.3%	-34.5%	-48.3%	26.7%
Ave Queue Length	-44.1%	-71.7%	97.7%	-18.9%	-51.9%	68.7%

It is evident from **Table 53** that:

- Average queue lengths and queue time have been significantly reduced by more than 60% to 70% for Line 2 and roughly around 20% to 50% for Lines 1 and 4.
- Whilst both GoA4 scenarios have demonstrated significant improvements over the existing operations, shortening train headways from 120 to 90 seconds could achieve further improvements in terms of reducing queue length and queue time for another 12% to 20% (for Line 2) and 25% to 70% (for Lines 1 and 4).

2.4 SUMMARY OF OUTCOMES

Chapter 2 documents the **validation and exploitation of a suite of simulation modelling tool for future metro systems improvement**, with a focus on improving crowd management in metro stations and passenger experience in a network and is positioned within **Task 7.1**. Within this task, the metro network and station models developed in WP4 were taken forward for a robust validation and cross-validation exercise aiming to demonstrate the ability of the suite of metro simulation models in replicating real-world metro system operations. The objective is to exploit the validated metro simulation tools to undertake **a multi-perspective, multi-purpose, and multi-level evaluation of the current (as-is) metro systems in Genoa and Sofia for their strengths and weaknesses, as well as exploring the potential impacts of future (to-be) fully automated (GoA4) operations**.

The suite of **multi-perspective metro system simulation models**, including metro network models (aiming at analysing higher level metro network and line level impacts) and metro station models (for more detailed crowd management level evaluations). These models were calibrated individually against a set of baseline passenger demand and train operational conditions for model accuracy based on commonly used quantitative model output metrics such as passenger flows and train operational parameters. Results demonstrated that the whole range of simulation models performed satisfactorily and broadly consistent across different modelling platforms. The modelled outputs of individual model matched well with their observed counterparts. Cross-comparison of outputs across models developed using different modelling platforms also demonstrated model consistency and thus justified model reliability.

As for model exploitation, a total of **23 scenarios** (organised into 5 themes) were developed to evaluate the operations of the two use case metro systems, AMT and MetroS. These scenarios were developed in close coordination with the two corresponding metro operators, reflecting their real-world practical operational concerns as well as ensuring the relevance of the evaluations and findings to real-life metro operation decision-making. Each scenario represents a dedicated purpose of evaluation including a normal day operation, under disruptions as well as GoA4 operations in the form of train headway reduction to 120 and 90 seconds (i.e. increases in service frequency) during the busiest peak periods of a day.

A **multi-level evaluation approach** was employed during the model exploitation stage, with clear focuses aiming to reveal impacts at (i) macroscopic level – providing an overall network-wide assessment of the two metro systems using the metro network simulation models; (ii) mesoscopic level – focusing on individual metro line and/or station level impacts, considering the operation of the whole metro network, using the metro network simulation models; and (iii) microscopic level – taking a crowd management perspective to reveal impacts on station level passenger movements, using the metro station simulation models.

At macroscopic level assessment (i.e. the overall network operational performance):

- The morning peak (7:00–10:00) was found to be the most stressed band for AMT, with saturation level reaching an average of 32.7%, and occupancy approaching 95 passengers per train on average, representing that train loading approaches comfort-critical levels.
- Waiting time is remarkably stable across all daytime bands (2.72 to 2.96 min), which reflects a consistent and reliable schedule.
- In dealing with disruptions, AMT was found able to maintain zero overcrowding events and zero passengers left behind, demonstrating solid structural resilience of the AMT network even under significant demand spikes.
- Impacts of GoA4 are most visible under disruption, with significant reduction of passenger waiting time for up to 50%, as well as saturation drops. It also dramatically reduces the peak-hour concentration of demand by serving passengers faster across more frequent services.
- However, the findings highlight the interaction between network topology and automation benefits: in a multi-line hub network, **GoA4 is most effective as a disruption-mitigation tool when the disruption affects a single node rather than the entire line**, because the surrounding network can absorb redistributed demand if served at sufficient frequency.

At mesoscopic level assessment (i.e. metro lines and metro station level performance, considering network operations):

- Considering switching from the existing operation to GoA4 operations, a suitable **trade-off between reducing passenger waiting times and improving train capacity utilisation should be made** when determining the optimal train headways.
- The average number of operation values is not significantly affected by disruption scenarios for both the AMT and MetroS networks. However, their operations under GoA4 operational scenario substantially increase their average number of train services

At microscopic level assessment, (i.e. crowd management level evaluation):

- The two metro systems have been performing satisfactorily during most of the time of a typical day;
- The afternoon peak period (16:00-20:00) has been the most critical and congested period across the day for both networks.
- As for De Ferrari station, on average, passengers would get served within 15 minutes during the peak, and within 5 to 10 minutes most of the time outside the peak.
- As for Serdica station, on average passengers would be able to get served in no more than 12 minutes most of the time during a typical day.

- As reflected in the metro station models, **GoA4 operations could effectively reduce passenger queue time and queue lengths**. However, the actual benefits revealed vary from 20% to over 80% reduction in queue lengths and queue time, on a case-by-case basis.

3 VALIDATION AND EXPLOITATION OF AI APPLICATIONS

3.1 OUTLINE OF THE VALIDATION AND EXPLOITATION EXERCISE

Chapter 3 documents the continuation of the work carried out during the first year of the NEXUS project and builds both logically and conceptually on the previously published and publicly available deliverables D6.1 and D6.2. In particular, Chapter 3 presents the further development of the activities initiated in WP5 and WP6 and conducted in WP7.

This section describes the validation and exploitation exercise as required within Task 7.2, which covers:

- the interactions of the three workstreams with focus on energy and CO_2 optimisation, in conjunction with future TCS and the TCO tools
- the validation of three selected use cases on (i) passenger demand forecasting; (ii) timetable creation using GTFS feeds; and (iii) vehicle uncleanliness detection.
- development of AI-supported strategies and measures for improving cybersecurity

The objective of the validation was not limited to verifying the technical functionality of the models, but rather to assess their transferability, structural robustness, operational relevance, and readiness for deployment in real metro systems.

In the first year, a fourth use case was also investigated, namely the “prediction of crowds based on exogenous data sources.” The prediction models for this demonstrator were developed by UNIGE using a comprehensive and heterogeneous dataset covering the entire year 2024. The data was collected to create a holistic picture of urban mobility patterns in Genoa, Italy, with the primary goal of predicting passenger volumes at key metro stations. This use case generates passenger demand information for the simulations in Task 7.1 and is documented in Chapter 2.

3.2 VALIDATION AND EXPLOITATION OF SELECTED AI MODELS AND APPLICATIONS

3.2.1 USE CASE 1 – VALIDATION OF PASSENGER DEMAND FORECAST MODEL

3.2.1.1 GENERAL INFORMATION

3.2.1.1.1 DEMONSTRATOR / USE CASE: SHORT DESCRIPTION

This demonstrator applies and validates a log linear econometric metro demand forecasting model using real operational data from Metro Sofia. The model was originally developed for West Midlands Metro and is tested in Sofia to assess structural robustness, predictive performance, and practical applicability in a different metropolitan context.

The use case evaluates the model's suitability for long term strategic planning, particularly in relation to network expansion and demand response to infrastructure investment. The validation supports subsequent exploitation of the model for scenario-based forecasting of future metro extensions.

3.2.1.1.2 OPERATIONAL CONTEXT / PILOT CITY / METRO NETWORK

The pilot city is Sofia, Bulgaria. Metro Sofia provides an appropriate validation environment due to the availability of consistent monthly ridership data and documented network expansion phases over the period 2011 to 2024.

(Source: metro Sofia)

Figure 15: MetroS network map showing Lines 1, 2, and 3 and staged extensions

The network experienced major structural changes during the validation period, including:

- Completion of Line 2 and extension of Line 1 in 2012, adding approximately 13 km and 13 stations
- Airport and Business Park extensions in 2015
- Introduction of Line 3 in 2020
- Significant ridership disruption during the COVID period followed by recovery

These events create distinct structural regimes that allow testing of model behaviour under infrastructure expansion and external shocks. The spatial configuration of the Metro Sofia network, including the phased extensions of Lines 1, 2, and 3, is illustrated in **Figure 15**.

MetroS has been developed with substantial support from European Union funding instruments, including the European Regional Development Fund and the Cohesion Fund. Planned extensions to Line 3 between 2022 and 2026 further reinforce the operational relevance of a validated forecasting framework.

The network therefore provides an appropriate setting for assessing how a structurally interpretable demand model performs under long term expansion dynamics in a European metropolitan rail system.

3.2.1.1.3 MODEL SPECIFICATION

The passenger demand model is specified as a log linear regression of the following form:

$$ln(D_t) = \beta_0 + \beta_1 ln(GDP_t) + \beta_2 ln(POP_t) + \beta_3 ln(NETKM_t) + \beta_4 ln(FARE_t) + \gamma t + \delta_m + \varepsilon_t$$

Variables Definitions:

D_t = Passenger demand in month t

t = trend term

δ_m = seasonal effect for month, m

γt = deterministic time trend into capture long-term growth

ε_t = random error term at time t

GDP_t = Employment growth rate in month t

POP_t = Population in month t

$NETKM_t$ = size of metro network in kilometres in month t

$FARE_t$ = Average fare in month t

The model follows a multiplicative demand structure expressed in logarithmic form. The coefficients β_1 *to* β_2 are interpreted as elasticities:

- β_1: elasticity of demand with respect to economic activity
- β_2: elasticity with respect to population
- β_3: infrastructure elasticity with respect to network length
- β_4: fare elasticity

The deterministic time trend allows the model to capture underlying growth not explained by observed variables, while seasonal dummy variables account for recurring intra-year demand patterns.

3.2.1.1.4 Model Variants Used in Validation

All model variants share the same log linear functional specification. Differences arise solely from the treatment of parameter estimation.

Transfer Model
All parameters $(\beta, \gamma, \delta)$are fixed using estimates obtained from the West Midlands Metro model and applied directly to the Sofia dataset without re-estimation. This tests strict parameter transferability.

Partial Adaptation Model

Elasticities (β)are fixed to the West Midlands values, while the intercept, time trend, and seasonal components are re-estimated using Sofia data. This allows adjustment to local temporal patterns while maintaining transferred behavioural relationships.

Sofia Refit Model
All parameters, including elasticities, intercept, trend, and seasonal effects, are estimated using the Sofia dataset. This tests whether the functional form remains valid under full local calibration.

3.2.1.2 VALIDATION OBJECTIVE

3.2.1.2.1 VALIDATION QUESTION

This validation assesses whether the log linear metro demand forecasting model developed (in WP6) for the West Midlands Metro in the UK can be applied to MetroS with acceptable predictive performance and whether the key demand elasticities remain structurally consistent when transferred to a different metropolitan and institutional context.

Specifically, the validation will test:

- predictive accuracy relative to statistical benchmarks;
- ability to reproduce demand changes following infrastructure expansion; and
- stability and transferability of behavioural elasticities.

3.2.1.2.2 KEY PERFORMANCE INDICATORS (KPIS): DEFINITION / BASELINE / TARGET / RESULT

The following KPIs are used to evaluate model performance as illustrated in **Table 54**.

Table 54: KPIs for the passenger demand forecast use case validation

KPIs	Definition	Baseline	Target
KP1: Forecast accuracy	Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and symmetric Mean Absolute Percentage Error (sMAPE) computed on monthly ridership.	Seasonal naive and trend plus seasonal benchmark models.	The structural model should perform comparably to or better than benchmark models when locally calibrated.
KPI 2: Expansion event performance	Forecast accuracy measured within defined post opening evaluation windows following major network extensions.	Seasonal naive in same windows	The model should reproduce the direction and approximate magnitude of demand uplift following infrastructure expansion.
KPI 3: Transferability of elasticities	Comparison of sign and magnitude of estimated elasticities between development and validation contexts.	Locally estimated Sofia coefficients.	Infrastructure elasticity retains directional consistency. Macroeconomic and pricing elasticities remain within plausible behavioural ranges.

3.2.1.2.3 EVALUATION LOGIC

Validation distinguishes between:

- Structural validity of the functional form
- Transferability of parameter value

A model may demonstrate structural robustness even if elasticity magnitudes require local calibration. Conversely, strong predictive performance alone does not imply behavioural portability. Detailed empirical results corresponding to these KPIs are presented in **Section 3.2.1.4**.

3.2.1.3 DATA BASIS

3.2.1.3.1 DATA SOURCES

The datasets for MetroS were obtained from official statistical and transport sources. Each variable used is documented below:

- Passenger demand – Total monthly passenger demand for MetroS was sourced from the operator covering the period from February 2011 to December 2025. January 2011 is excluded because no official passenger demand value is available. No imputation was applied since the missing value occur at the start of the series and does not affect continuity.
- Population – Resident population of Sofia Capital administrative district as of 31 December each year was sourced from National Statistical Institute of Bulgaria, annual district level population estimates for Sofia. Annual values were collected, converted to monthly inputs by assigning each annual value to all twelve months of the corresponding year, consistent with West Midlands model development approach. The data covers 2011 to 2024, as population figures are published only through 2024 at the time of analysis. Therefore, the validation period is limited to February 2011 to December 2024 to maintain consistency across regressors
- Economic Activity – Employment rate for persons aged 15 to 64 years in Sofia Capital as collected from National Statistical Institute, Labour Force Survey covering 2011 to 2024. In the West Midlands development model, employment rate was used as a proxy for economic activity. The same definition is used for Sofia to ensure consistency of the GDP proxy variable.
- Network Length – Total operational metro route length in kilometres was retrieved from Metro Sofia's official expansion documentation and operational reports. Network length is encoded as a stepwise series that increases at the exact month when new sections became operational. Major expansions milestones included: September 2012, April 2015, May 2015
- Fare – Representative single ride fare used as proxy for average fare levels and modelled as a step function documenting tariff change. The standard single ride ticket was used as the most frequent fare category, consistent with the development model approach for West Midlands. Detailed ticket mix data are not publicly available. Therefore, the most common fare category is used as a representative proxy, consistent with the methodology applied in the development model.

3.2.1.3.2 DATA PERIOD

The Sofia validation dataset covers monthly observations from February 2011 to December 2024. As mentioned above, January 2011 is excluded because no official passenger demand value is available in the source series. The omission is limited to this single boundary month and no imputation was applied, as the missing value occurs at the start of the sample and does not affect continuity of the time series.

Although passenger demand observations are available through December 2025, population data for Sofia Capital are officially published only up to 31 December 2024 at the time of analysis. To ensure

consistency across regressors, the validation period is therefore restricted to February 2011 through December 2024.

3.2.1.3.3 DATA VOLUME

Number of monthly observations: Approximately 167 months

Number of explanatory variables: 4 logged regressors, trend variable, 11 monthly seasonal indicators

Total model parameters include intercept, elasticities, trend coefficient and seasonal effects.

3.2.1.3.4 GROUND TRUTH DEFINITION

Ground truth demand is defined as the observed monthly passenger demand for MetroS over the period February 2011 to December 2024. The demand variable reflects system wide monthly ridership as reported by the official source. No smoothing or model-based adjustment is applied to the observed demand values used for validation

Data quality checks

Completeness – Only one missing observation, January 2011, located at the start of the series. No internal gaps are present from February 2011 onward.

Consistency – Network length step changes verified against official opening dates. Fare step changes verified against documented tariff decisions. Population and employment rate series verified for monotonic plausibility and absence of structural inconsistencies.

Transform readiness – All variables confirmed strictly positive prior to log transformation.

Outliers and shocks – COVID period retained as observed without smoothing to test robustness under structural demand disruption.

3.2.1.3.5 DATA LIMITATIONS

Population timing – Sofia population is reported as end year values, whereas West Midlands used midyear estimates. Annual values are treated as representative for each year to maintain methodological consistency.

Fare proxy – The model uses a representative single ticket fare due to lack of detailed ticket mix data. This may not capture weighted average fare effects.

Economic proxy – Employment rate is used as a proxy for economic activity. This reflects labour market conditions but does not capture full income or productivity dynamics.

Structural shocks – The COVID period introduces non-economic behavioural shifts that may affect elasticity stability.

3.2.1.4 RESULTS OF VALIDATION

3.2.1.4.1 BASELINE

To evaluate whether the proposed log linear demand forecasting model adds predictive value, its performance was compared against statistical benchmark models applied to the Sofia demand series.

Table 55: Baseline Model Performance

Model	RMSE (Passengers)	MAE (Passengers)	sMAPE (%)
Seasonal naive	1,898,293	1,200,665	20.5
Historical mean	1,807,290	1,355,690	21.4
Trend plus season	1,588,117	1,179,758	18.9

The “trend plus season” model provides the strongest statistical baseline and is used as primary comparison benchmark as shown in **Table 55**.

3.2.1.4.2 EVALUATION METHODOLOGY

Validation was conducted using time aware evaluation as opposed to random splitting to preserve temporal structure. Because metro ridership evolves sequentially and responds to infrastructure, economic and policy changes over time, random cross validation would violate the underlying data generating process and inflate predictive performance. Two validation strategies were therefore applied.

Full period performance comparison against baseline models

In this strategy, overall predictive performance was evaluated across the full validation sample from February 2011 to December 2024. Model fitted values were compared against observed ridership and benchmarked against three statistics:

- Seasonal naïve model
- Historical mean model
- Linear trend with seasonal dummies

Performance was assessed using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and symmetric Mean Absolute Percentage Error (sMAPE). These metrics provide complementary information: RMSE penalises large deviations, MAE captures average magnitude of errors, and sMAPE normalises errors relative to observed levels (**Table 54**).

This full sample evaluation tests whether the structural log linear specification provides predictive value beyond simple time series extrapolation.

Event-based reverse forecasting around major network expansion

A targeted reverse forecasting approach was used to evaluate model behaviour during discrete infrastructure shocks. The major MetroS extensions completed in September 2013 and April-May 2015 provides a quasi-natural experiment in which network length increased abruptly.

For each expansion event:

- A training sample was defined using only data available prior to opening date.
- The model was estimated using pre-expansion data.
- Ridership was forecast for the post expansion period
- Forecast were compared to observed demand

The design tests whether the model can reproduce the magnitude and timing of demand uplift associated with network growth, rather than merely fitting long run trends.

Because 2012 training window exhibited limited variation in network length and fare, elasticities were fixed to West Midlands model's estimates and only intercepts and trend component were locally estimated. For the 2015 window, sufficient variation existed in network length to estimate the infrastructure elasticity directly, while variables without variation were excluded to avoid rank deficiency. Specifically, the following modelling efforts were conducted:

- *Structural model variants* – Three structural model variants of the log linear specification were tested to evaluate transferability and stability:
- *Transfer model* – West midlands elasticities were applied directly to Sofia regressors without re-estimation. This tests strict parameter portability
- *Partial adaptation model* – Elasticities were fixed to West Midlands values, while intercept and seasonal components were re-estimated on Sofia data. These tests whether behavioural parameters are transferable once local temporal structure accounted for.
- *Sofia refit model* – All parameters were estimated using Sofia data. This tests whether the multiplicative functional form remains valid when locally calibrated.

3.2.1.4.3 OVERALL PERFORMANCE VS. BASELINE

The fitted models and the associated KPI evaluation statistics are summarised in **Table 56**.

Table 56: Structural Model Performance

Model	RMSE (Passengers)	MAE (Passengers)	sMAPE (%)
Transfer	3,849,536,877	2,492,181,594	188.0
Partial adaptation	9,940,196	6,314,432	57.2
Sofia refit	1,646,390	1,213,520	19.2

Direct transfer of West Midlands elasticities produces extremely large forecast errors, with sMAPE exceeding 180 percent as shown in **Table 56**. This indicates that behavioural parameter values are not directly portable across the two metro systems.

Allowing partial adaptation through local estimation of intercept and seasonal component would substantially reduce error, but performance is relatively lower than that of the statistical baselines. This suggests that correcting only temporal structure is insufficient when core elasticities differ.

Full re-estimation of log linear specifications on metro Sofia data yields an RMSE of 1.65 million and sMAPE of 19.2 percent, which is comparable to the strongest statistical benchmark. This indicates that while parameter values are context dependent, the multiplicative demand structure remains empirically valid when calibrated locally.

Figure 16 support this interpretation. The Sofia refit model tracks long term ridership growth and seasonal patterns closely, though deviations are visible during structural breaks such as 2015 extension and the COVID period. The visual fit confirms that most residual error is concentrated around periods of abrupt demand disruption rather than during stable growth phases.

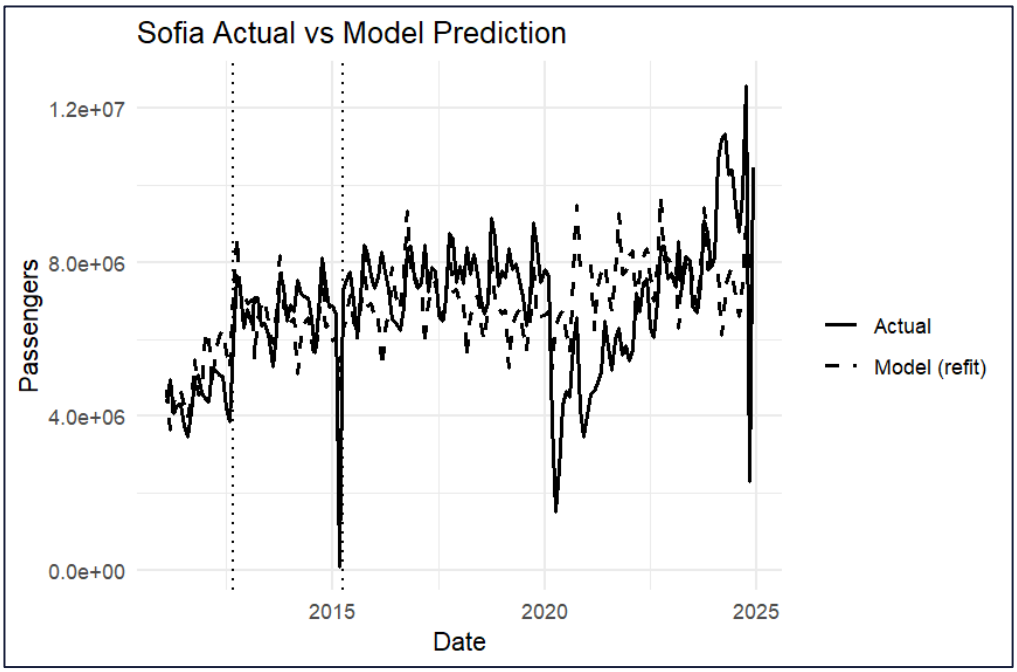


Figure 16: Actual vs Predicted Monthly Ridership for MetroS

3.2.1.4.4 PERFORMANCE IN EDGE CASES: STABILITY AND ROBUSTNESS

Expansion Event 1: September 2012 Extension

Training period: February 2011 to August 2012

Testing period: September 2012 to December 2013

Table 57: Event Window: 2012 Performance

Model	RMSE (Passengers)	MAE (Passengers)	sMAPE (%)
Fixed elasticity event model	1,037,512	909,522	14.8

The 2012 expansion window demonstrates stable and accurate post-opening prediction. Despite limited pre-expansion variation in network length, the fixed elasticity specification reproduces the magnitude and timing of demand uplift with relatively low percentage error. This suggest that the infrastructure response mechanism embedded in the log linear formulation is structurally plausible across systems, at least for centrally located extensions with immediate ridership integration.

Most importantly, error is not concentrated at the opening date itself, indicating that model captures both step increase and subsequent stabilisation phase well.

Expansion Event 2: April-May 2015 Extension

Training period: February 2011 to March 2015

Testing period: April 2015 to December 2016

Table 58: Event Window: 2015 Performance

Model	RMSE (Passengers)	MAE (Passengers)	sMAPE (%)
Reduced event model	4,166,590	4,073,691	76.7

In contrast, predictive accuracy deteriorates substantially for the 2015 expansion. The elevated percentage error indicates that the post opening demand trajectory diverges from the structural response implied by the model. This may reflect heterogenous demand composition, gradual airport patronage builds up, spatial redistribution effects or nonlinear dynamics not captured by constant elasticity specification.

The asymmetry between 2012 and 2015 performance highlights that infrastructure elasticity may vary depending on corridor characteristics, catchment maturity and modal competition, rather than operating as a uniform system wide multiplier.

3.2.1.4.5 ELASTICITY STABILITY AND TRANSFERABILITY

As indicated in **Table 59**, only network length elasticity retains directional consistency across systems. While its magnitude declines substantially in the Sofia refit; its positive sign supports the core hypothesis that network expansion contributes to demand growth.

On the other hand, macroeconomic and pricing elasticities exhibit sign reversals. This indicates that behavioural responses to income conditions and fare variation are highly context specific and likely influenced by local institutional arrangements, fare integration structure and baseline model shares.

These results suggest that structural form may generalise more reliably than parameter values. Transferability therefore appears conditional rather than universal.

Table 59: Elasticity Comparison

Variable	West Midlands	Sofia Refit	Sign Consistent
log_gdp	4.15	-1.12	No
log_pop	1.23	-2.32	No
log_netkm	9.26	0.73	Yes
log_fare	-1.30	0.32	No

3.2.1.4.6 ROBUSTNESS OBSERVATIONS

The validation retains the COVID period without smoothing to assess structural resilience under exogenous disruption. While this increases overall error magnitude, it provides more stringent test of model stability.

Population timing conventions do not materially alter coefficient direction, suggesting limited sensitivity to annual reference alignment.

Fare elasticity instability is consistent with the limited temporal variation in tariff levels during the sample and may reflect identification constraints rather than true positive fare responsiveness.

Residual diagnostics from the Sofia refit indicate no strong serial correlation patterns, supporting the adequacy of the deterministic trend and seasonal structure.

3.2.1.5 OPERATIONAL IMPACT & EXPLOITATION

3.2.1.5.1 EXPECTED BENEFITS

The validated log linear framework provides a transparent and interpretable approach to long term metro demand forecasting. Unlike purely statistical time series models, the specification links ridership explicitly to infrastructure provision, demographic conditions, economic activity, and pricing structure.

The principal operational benefits include:

- Quantification of infrastructure elasticity, enabling planners to estimate demand uplift from network expansion
- Scenario testing capability for population growth, economic change, and fare adjustments
- Transparent parameter interpretation suitable for public sector audit and funding justification
- Reproducible methodology applicable across metro systems

The model therefore supports strategic decision making rather than short term operational forecasting alone.

3.2.1.5.2 ESTIMATED IMPACT

Although direct parameter transfer between West Midlands and Sofia was not supported, the structural specification remains valid when locally calibrated. This indicates that:

- The multiplicative demand framework is robust across metropolitan contexts.
- Infrastructural elasticity remains directionally consistent
- Model for generalises more reliably than parameter values

For infrastructure investment appraisal, this implies that demand response to network expansion can be structurally modelled using a consistent framework, provided local calibration is performed.

In practical terms, the approach can reduce on purely extrapolative models and improve the transparency of demand forecasts used in cost benefit analysis and funding applications.

3.2.1.5.3 TRANSFERABILITY

In terms of transferability, the validation demonstrates conditional transferability. In terms of structural transferability, the log linear form generalises across metro systems. As for parameter transferability, the elasticities are context dependent and require local estimation.

In particular, infrastructure elasticity appears more stable than macroeconomic and fare elasticities, suggesting that physical network effects may generalise more reliably than behavioural pricing responses.

This supports a framework where (i) Model structure shared; (ii) parameters are locally estimated; and (iii) results are benchmarked across cities

3.2.1.5.4 RISK / LIMITATION

Several limitations are acknowledged in this model:

- Limited variations in fare within certain periods constrained identification of price elasticity
- Structural breaks such as COVID introduce non-economic behavioural effects
- Airport related extensions may exhibit gradual adoption rather than immediate step change demand
- Administrative population boundaries may not perfectly align with metro catchment areas.

These factors affect parameter stability but does not invalidate the structural specification in the model.

3.2.1.5.5 RECOMMENDATIONS

To strengthen operational deployment, the following actions are recommended:

- Extend validation to additional European metro systems to assess multicity parameter stability.
- Incorporate lagged infrastructure response terms to capture gradual demand ramp up
- Explore interaction effects between network length and population density
- Integrate real fare deflation and broader economic indicators for improved elasticity identification
- Develop pooled panels with city interaction terms for cross system benchmarking.

Such extension would enhance the model's suitability for infrastructure appraisal and strategic network extension planning.

3.2.2 USE CASE 2 – VALIDATION OF TIMETABLE CREATION USING GTFS FEEDS

3.2.2.1 BACKGROUND

Data-driven strategies are becoming increasingly fundamental in transport operations and planning. General Transit Feed Specification (GTFS) presents a variety of benefits and versatility for operators, agencies and users if produced to a high standard, validated and verified. **Figure 17** demonstrates how GTFS operates, providing a link between public transport timetables and passengers, planners, governance organisations, and third-party applications (GTFS, 2026). Task 7.2 involves the validation and exploitation of the applications and strategies developed in WP6. As for timetable creation using GTFS feeds, it requires multiple time-dependent OD matrices and archives of real-time GTFS updates as inputs for determining periods of homogenous demand (Van Der Knaap et al., 2024). However, only a fixed daily OD demand matrix and pre-determined fixed train service scheduled were available from the two metro operators. Therefore, validation of the complete cycle, as outlined in D6.1, – data readiness, pattern discovery and prediction, optimisation and integration, as well as deployment and real-time adjustment – was not possible. Therefore, this use case of Timetable Creation using GTFS Feeds was presented as a feasibility study and that a physical timetable was not created as part of this work. Consequently, this section will explore the feasibility and validity of the methodology for creating

timetables using GTFS feeds as well as other algorithmic methods used for timetable creation. This paper aims to continue and progress the work completed in Work Package 6. Including both a literature review and use-case scenario, these prior studies illustrated the benefits and opportunities presented by GTFS data for metro operators to enhance existing timetables and for optimising service operations. Specifically, this paper will highlight various case studies focusing firstly on advanced algorithmic timetable creation and latterly, illustrating the uses of leveraging GTFS data.

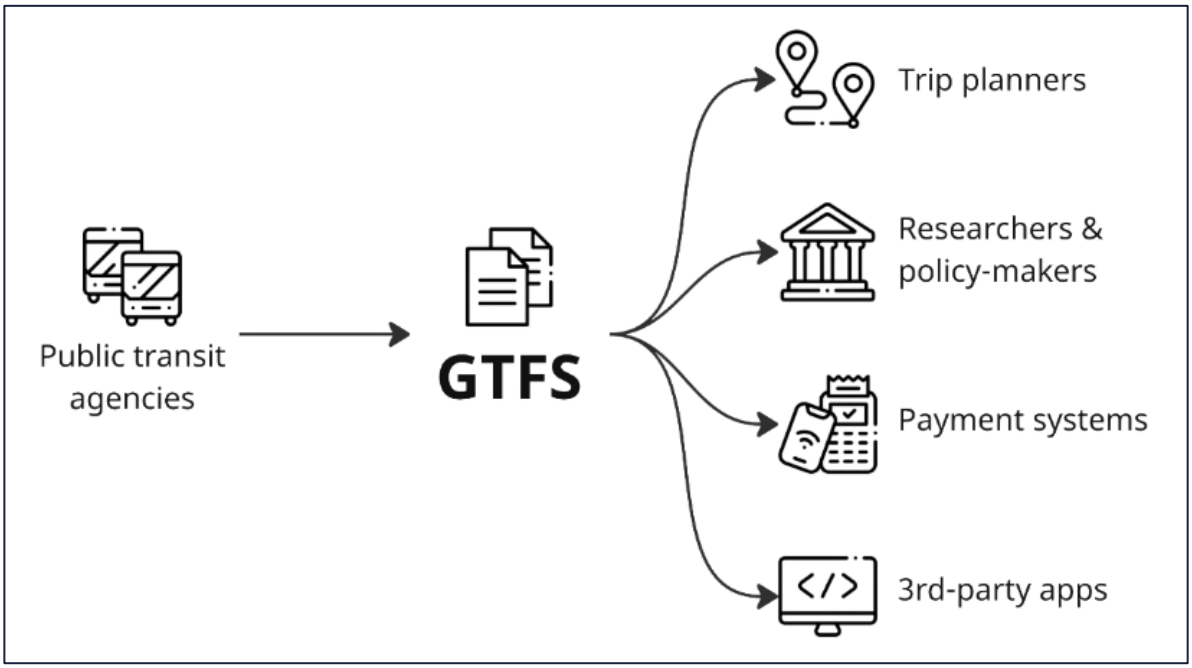


Figure 17: An overview of GTFS - (GTFS, 2026)

3.2.2.2 FEASIBILITY & VALIDATION CASE STUDIES

The following case-studies demonstrate the feasibility and validity of timetable creation using GTFS feeds. Despite this, it is important to recognise that there is no official case study or announcement from metro operators on the adoption of advanced scheduling algorithms in creating timetables – however this work aims to illustrate the benefits and processes behind this. The following case studies focus on a variety of operational factors, including disruption recovery, modal integration, energy consumption, passenger waiting times, crowding and capacity, fleet utilisation and efficiency.

3.2.2.2.1 REAL-TIME RE-SCHEDULING

Across all metro networks globally, an inevitable aspect of operations is disruptions and disturbances along the route. This case study employs a unique two-step optimisation framework to address real-time train rescheduling. The first step involves managing traffic flow by optimising dispatching techniques, such as retiming, cancellation, short-turning, and backup rolling stock utilisation, in order to realign train operations with the original planned timetable. The second stage introduces stop-skipping and further adjusts the train schedule to ensure the best possible service quality throughout the delay recovery period. Mixed-integer nonlinear programming models are used to formulate both stages, with

the second-step model being heuristically decomposed. This framework was demonstrated using data from the Beijing Yizhuang Metro Line and a small-scale scenario. The suggested two-step optimisation framework can meet the performance requirements set by the metro operator, as well as, exceeding the performance of the existing methodology used to re-plan metro services in the event of disruption. It is further highlighted that the applicability of the framework is widespread, with the potential to address various disruption scenarios at various locations. (Su, et al., 2025)

3.2.2.2.2 TIMETABLING RAIL-METRO INTERCHANGES

Typically, in large cities with metro networks, mainline railway stations are usually served by the metro, facilitating connectivity between the rail network and the wider city region. The case describes how large volumes of passengers transfer to metro services when intercity rail services arrive at an interchange station. Consequently, the injected passenger demand cannot be adequately handled by the metro timetables with regular headway currently used in real-world operations. Conversely, during gaps between intercity services, passenger demand is much reduced, therefore, to enhance the service quality, timetable optimisation is suggested. The case devises a model to describe the time-varying demand of passengers arriving at an interchange station. This case develops a mathematical model based on the arrival times of intercity trains and the interchange process for passengers transferring from intercity services to metro services. Following this, a model timetable is proposed to optimise train departure time of a metro line with a mainline station interchange, given the time-varying passenger demand and capacity of metro trains. The case aims to reduce passenger waiting times at the interchange station and an adaptive large neighbourhood search technique is used to solve the suggested timetable model. Consequently, the case used the Beijing Metro Line 9 at Beijing West Station to test the optimised timetable and the findings of the research show that the proposed model can reduce the average passenger waiting time at interchange stations by 28.47%. (Guo, et al., 2021)

3.2.2.2.3 WAITING TIME OPTIMISATION

Across many metro networks, peak-hour capacity is often a significant constraint and a source of frustration for many passengers, with some passengers left behind due to overcrowding. Consequently, this case proposes to match passengers with different trains and aims to reduce passenger waiting times by optimising metro timetables. The objective of the first aspect of the research is to match passengers with various services using both timetable and fare data. For the second aspect of the research, the objective is to reduce passenger waiting times, this uses frequency and dwell times as decision variables. The model was tested on Beijing Metro Line 1 with the outcomes revealing that the newly developed timetables resulted in various improvements for passengers, with waiting times dropping by 15.5%, while the number of unique matched passengers rose by 37.7%. (Huang et al., 2021)

3.2.2.2.4 SHORT TURNING

A common practice seen across metro networks is short turning, this involves the early termination of a train at a preceding station, and the train is subsequently reversed in order to recover the original timetable. This case study focuses on this technique, however, looks to implement short turning as a regular service within the timetable. The study utilises a mixed integer nonlinear programming (MINLP) model for timetabling, in which the predetermined headway is used based on passenger demand to optimise both full-length and short-turning train services. Based on data from Beijing Subway Line 4,

two case studies with varying scales are created to evaluate the effectiveness of train schedules using the short turning technique. According to the modelling data, there has been a 20.69% decrease in fleet utilisation, consequently, freeing up more vehicles for increased capacity or maintenance. (Zhang et. al., 2018)

3.2.2.2.5 CROWDING MITIGATION

Peak time congestion is a recurring constraint for metro operators globally, with passengers left behind on platforms and poses significant operational challenges. Consequently, to combat these operational challenges, this case study focuses on using service timetabling and passenger flow control coordination problem with changeable dwell time on congested metro lines. This research develops a mixed-integer nonlinear programming model. The study aims to reduce platform congestion yet, maintaining the capacity of the metro line. Creating a timetable that is service oriented and using an efficient passenger flow control method. Latterly, numerical tests demonstrate a 40.8% reduction in passenger waiting times overall, which can successfully lessen train traffic congestion and confirm the model's efficacy. (Wu & Yang, 2023)

3.2.2.2.6 ENERGY OPTIMAL TIMETABLES

A significant, yet often overlooked, aspect of metro operations is energy consumption. This first case study focuses on energy-optimised timetables for metro operations, the approach exploits the use of regenerative braking, whilst simultaneously minimising the overall energy consumption of the train. Consequently, the model developed facilitates GoA2+ trains under Communications-Based Train Control supervision to maximise energy savings on the train in real-time, producing energy-optimised timetables, with metro network constraints considered. The impact and potential of this approach was demonstrated on Shanghai Metro Line 8, the results revealed a 20.93% and 28.68% energy savings as compared to current real-world timetables. Furthermore, this method is easily transferrable to other metro networks and can be quickly implemented to various lines. (Gupta et al., 2023)

3.2.2.2.7 ENERGY CONSUMPTION

From both a business and an environmental perspective, many metro operators are exploring strategies to reduce energy consumption. This case study focuses on an optimised model of train control, with an optimised timetable and headway management to minimise energy consumption. Vehicle features such as regenerative braking and driving techniques such as coasting are also incorporated into the timetable, with many modern metro systems operating vehicles that have these capabilities, optimising the use of regenerative braking and coasting can have a significant impact on energy efficiency and reduction. Furthermore, varying train masses are also incorporated into the model, with regular passenger movements on and off the trains. When compared to the other train control systems in timetable optimisation, testing the proposed timetable on Beijing Metro Line 5 showed that the proposed enhanced train control strategy can reduce traction energy consumption by 20% in certain areas of the network with steep downhill slopes. Additionally, the overall net energy consumption is lowered by 4.97% because of the integrated model's ability to greatly extend the overlapping duration between driving, coasting and braking trains. (Zhou et. al., 2018)

3.2.2.2.8 PEAK DEMAND OPTIMISATION

During peak hours, significant tidal passenger flow characteristics, namely the imbalance of passenger numbers in both directions, have been brought about by the recent rapid rise of cities and their inhabitants. This disparity frequently results in an inefficient use of capacity in one direction and a shortage of train capacity in the other. To reduce operating costs and improve the quality of passenger service, this case suggests a schedule optimisation strategy that integrates several strategies to meet the tidal passenger flow need of urban rail transportation. The case proposes multi-strategy optimisation technique, consisting of the unpaired operation strategy and the express/local train operation strategy. Both methods are flexible enough to adjust to changing passenger demand over time. A mixed integer linear programming model is developed based on the decision variables of headway, running time between stations, and dwell time. To demonstrate the efficacy and accuracy of the suggested multi-strategy method and model, simulations under various passenger needs are carried out using the Shanghai Suburban Railway airport link line as an example. The findings show that the multi-strategy optimisation approach successfully reduces train congestion and reduces overall expenses for both the operator and the passengers by 38.59%. (Jin et. al., 2024)

3.2.2.2.9 ENERGY CONSUMPTION & PASSENGER WAITING TIMES

Optimising both passenger waiting times and energy consumption simultaneously is a complex challenge for metro operators. This case study focuses on achieving reductions in these two areas simultaneously, recognising the complexity of this challenge because of a variety of factors that reflect the urban railway's operating state, including the number of passengers boarding and alighting, the number of passengers on board each train, and the overall train operation status along the route. However, the most influential factor is the fluctuating and constantly changing nature of passenger demand, consequently, these factors combined, present a challenge to optimise both passenger waiting times and energy consumption. Consequently, to solve this challenge, this case study, using Seoul Metro Line U as a use case, creates a real-time metro operation optimisation model based on recurrent neural networks that is developed using data that represents the train departure times from the stations. To find the best daily metro schedule, data is taken from and recreated from a variety of simulated operation plans. As a result, the suggested model determines the best operating strategies in real time for trains leaving the next station in an average of 0.18 seconds. When compared to the existing train operating methods, the results of metro operation simulations employing the suggested optimal operation techniques show an efficiency gain of 7–14%. (Oh et al., 2024)

3.2.2.2.10 INTEGRATION & OPTIMISATION

This next case study takes a tri-lateral approach, considering the plan of the line, timetable and rolling stock schedule, using Chongqing Metro Line 3 as an example to validate the proposed timetable and algorithm. This case study examines the challenges of integration in urban rail transit timetabling and operations, using passenger flow demand, line fixed facilities, and equipment conditions as inputs. To reduce corporate operating costs and passenger wait times, an integer nonlinear programming model was developed. To achieve a collaborative optimisation of line planning, timetabling, and rolling stock scheduling, the model was broken down, and a deterministic method was created. When compared to a train operation plan created by stages using manual experience, computational analysis demonstrates that this integrated optimisation model can more effectively lower the operational costs

of operators. It is possible to cut the total cost by 13%, the variable cost by 23%, and the fixed cost by 24%. The integrated optimisation model can provide a more reasonable train capacity allocation based on the distribution characteristics of the line passenger flow, even though passenger waiting times increase. The train operation plan's passenger service level was improved as the minimum section load factor rose from 9.27% to 14.83% and the maximum section load factor dropped from 109.44% to 94.31%. Simultaneously, the model considered the rolling stock schedule, timetable, and line planning. Consequently, the model can precisely determine the size of the fleet, which can simplify the implementation for operators. The resulting optimised operational timetable is periodic to ensure a successful headway between trains within the timetable. Alongside this, the model facilitates a reduction in the fleet size and the running distance of trains. In addition to being consistent with the real demand, this periodicity helps to rationalise the optimised train operation plan. (Li et. al., 2023)

3.2.2.2.11 IMPLEMENTATION OF ALGORITHMIC TIMETABLING SOLUTIONS

Despite having no official examples of advanced algorithms being employed to create metro timetables, an example of interest is seen in Norway. Unlike other case-studies, this use case goes further and implements an algorithm for dispatching and timetabling on an Iron Ore Railway Line between Narvik, Norway and Luleå, Sweden. The "Iron Ore Line," a crucial section of the railway network that connects Sweden to the Norwegian coast, is the subject of the implementation that is addressed in this case. The line's capacity has been put to the test in recent years due to an increase in iron ore trains, some passenger services and other goods trains. To make sure that the operational capacity is utilised to its maximum potential, dispatching services needs to be optimised. Bridging the gap from theoretical study to implementation presents several significant challenges and many oversimplifying assumptions need to be abandoned. Consequently, this case outlines how to overcome these challenges to effectively dispatch trains on a railway in northern Sweden and Norway. Interacting with real-time systems to obtain accurate data in real-time is another significant challenge that must be mitigated. Collaborative working with dispatchers is crucial because they have the final say over whether the algorithm's recommendations are approved. It is important to gain an understanding as to why dispatchers the choices that they do. In a real-time environment, the algorithm's input data is refreshed every ten seconds before it is performed once more. The method described in this case has been evaluated on real-time data over a long period of time and has been able to deliver solutions in 2 seconds, as desired. Dispatchers are given solutions covering a 2-hour period and the solutions are constantly updated, depending on the current status of the trains, consequently, it is important to ensure that solutions are delivered rapidly as existing solutions may be already outdated against real-time data. The Norwegian National Research Council provided funding for the system's implementation, and it must be recognised that, it was only functional on the Norwegian part of the route, assisting the Norwegian signallers stationed at Narvik. However, it must be noted that, there is a second control centre in Boden for the Swedish portion of the line. It is important to note that the poor synchronisation between the line's two control centres causes a number of issues. Only until the trains are getting close to the border do the dispatchers at Narvik learn of the scheduling decisions made at Boden, and they are powerless to make alterations to trains currently in service. To conclude, the case recognises that greater benefits would be achieved if a similar system was implemented at Swedish signalling centres. (Bach et al., 2019)

3.2.2.2.12 SUMMARY

Consequently, the case studies highlighted in this section illustrate the benefits and feasibility of creating timetables using advanced algorithms, such as using GTFS data. As illustrated in the various case studies, tangible results were discovered, using data from real metro lines, with reductions in energy consumption and passenger waiting times, enhanced passenger capacity, and integration with other services. Finally, disruption recovery and fleet utilisation and peak hour utilisation can also be enhanced by using advanced algorithmic timetables. The Norwegian example shows how implementation can be achieved, although on a small scale, the case study recognises the gap between theory and practical implementation and the challenges that are required to be met to bridge this gap. It is important to note however, that most of the case-studies highlighted are still at a research level and there has been no unambiguous evidence of any official adoption of advanced optimisation algorithms to enhance metro operations from any metro operator. Illustrated below in a summary table (**Table 60**), are some of the factors (**Figure 18**) that can be influenced and enhanced by advanced timetable creation alongside some key findings and the metro lines the case-studies were tested on:

Table 60: A summary table highlighting timetable creation case-studies

Operational Aspect	Metro Case Study	Findings	Reference
Real-Time Rescheduling	Beijing Yizhuang Line	Meets operational requirements of Yizhuang Line.	Su et al., 2025
Interchange Integration	Beijing West Station	28.47% reduction in passenger waiting times	Guo et al., 2021
Waiting Times	Beijing Metro Line 1	Waiting times dropping by 15.5%	Huang et al., 2021
Crowding	Unspecified	40.8% reduction in passenger waiting times	Wu and Yang, 2023
Short Turning	Beijing Metro Line 4	20.69% decrease in fleet utilisation	Zhang et. al., 2018
Energy Optimisation	Shanghai Metro Line 8	Between 20.93%-28.68% energy savings	Gupta et al., 2023
Energy Consumption	Beijing Metro Line 5	20% Reduction in Energy Consumption	Zhou et. al., 2018
Peak Hour Demand	Shanghai Suburban Airport Link	38.59% cost reduction	Jin et. al., 2024
Energy Consumption & Waiting Time Optimisation	Seoul Metro Line U	7-14% Efficiency Increase	Oh et al., 2024
Integration & Optimisation	Chongqing Metro Line 3	13% Total Cost Reduction, Optimised train loadings.	Li et. al., 2023
Implementation	Narvik, Iron Ore Line	2-second solutions for train timetabling	Bach et al., 2019

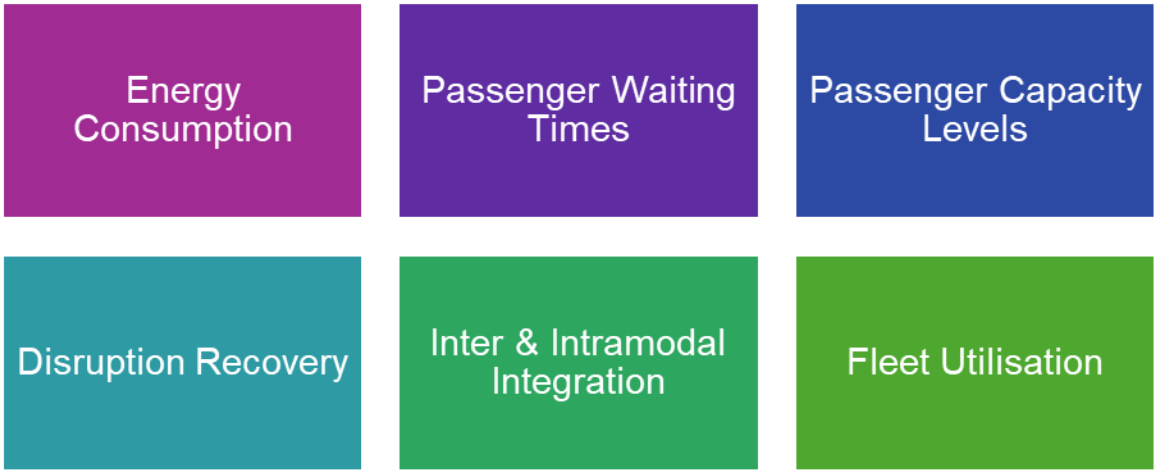


Figure 18: Operational Factors that advanced timetabling may enhance.

3.2.2.3 LEVERAGING GTFS DATA

GTFS data has a wide scope for using and can be leveraged in many ways, this section highlights some of these areas of operation that GTFS can be leveraged. Firstly, focusing on Network Analysis, followed by Performance Analysis and Planning and Monitoring, as well as a practical tool for timetable validation. These case studies illustrate the potential opportunities presented by leveraging GTFS data for metro operators and the benefits that may be exploited from this data as a result.

3.2.2.3.1 NETWORK ANALYSIS

The first area of focus is network analysis; this is a key factor for ensuring efficient and reliable transport operations. The first study presents GTFS2STN, a tool that transforms static GTFS data into spatiotemporal networks. This change makes it possible to analyse transit system accessibility in detail. Moreover, the case further presents a web-based application version of the tool that enables users to create spatiotemporal networks online and carry out simple analyses, such as calculating travel time variability between origin-destination pairs over time and creating isochrone maps from a given origin. Tested using data from the Nashville transit network, comparative investigation reveals that the presented tool demonstrates similar findings to existing open-source tools for producing isochrone maps from GTFS inputs. A significant advantage of the tool is the ability to enable users to upload and analyse historical or suggested GTFS feeds from any transit agency, it provides researchers and planners with more flexibility to assess transit plans than alternative tools. Consequently, GTFS2STN is a useful tool for transit system planning and research because of this capability, which makes it easier to evaluate accessibility and travel time variability in transit networks over long periods of time. (Liu et. al., 2024)

3.2.2.3.2 PERFORMANCE ANALYSIS

Service reliability and performance are significant aspects of operating an efficiency and dependable transport system. This second case proposes a method for extracting transit performance indicators from GTFS Real-Time (GTFS-RT) grouping them to roadway segments. This data is then analysed

using a framework in terms of random variation on a segment-by-segment basis (stochastic delays) and consistent, predictable delays (systematic delays). Every technique and dataset utilised can be applied to transit systems that report vehicle positions using GTFS-RT characteristics, consequently providing a network-wide screening tool that may be used to find areas where potential solutions (e.g., schedule padding) or proactive infrastructural upgrades (e.g., bus-only lanes, transit signal priority) may be beneficial at enhancing efficiency and service reliability, on any transit network. A case study involving a one-year set of GTFS-RT data obtained from the King County Metro bus network in Seattle is conducted to illustrate this concept. By calculating and allocating systematic and stochastic delays to network segments, regional trends in efficiency and dependability were shown. Findings imply that high-pace segments present a potential for stochastic speedups, while the network may contain excessive schedule padding. To demonstrate continuing performance metrics in the case study area, an online interactive visualisation tool was created in addition to the static analysis covered in the study. To promote more generalisable development on the GTFS-RT standard, all code is available as open-source. (Aemmer, Ranjbari and MacKenzie, 2022)

3.2.2.3.3 PLANNING AND MONITORING

Furthermore, network planning and monitoring is another crucial aspect of transport operations, ensuring recovery from delays and reducing potential bottlenecks is a significant challenge for transport operators. To help with the collection, elaboration, storage, and visualisation of GTFS-RT data, this third study proposes an open-source framework with key logic unit design. By giving academics and planners the tools to comprehend and start extracting GTFS data, it closes the gap for those who desire real-time access to GTFS. The framework is used on actual data from Translink in Brisbane, Australia, to demonstrate its applicability. The extracted data is then used for dashboard visualisation and performance metrics. The codes can be found online and are open-sourced. (Lim, et. al., 2019)

3.2.2.3.4 PRACTICAL VALIDATION

GTFS provides a wide variety of benefits for operators, passengers and wider stakeholders. Once a timetable is prepared and completed, the official GTFS website provides a tool for validating and cross-checking completed data, prior to or following publication. This tool highlights errors within the data, providing insights into possible solutions. Consequently, facilitating a simple, yet effective tool for practical validation of GTFS data evaluating the provided data against GTFS Reference standard as well as, best practices. (Mobility Data, 2026)

3.2.2.3.5 SUMMARY

As illustrated in this **Section 3.2.2.3**, the versatility and benefits of leveraging GTFS data can provide significant operational enhancements for metro operators, including network and planning analysis, as well as, planning and monitoring. Illustrated in the Table 61 is a summary of this section:

Table 61: GTFS Summary

Leverage	Purpose	Reference
Network Analysis	A tool that transforms static GTFS data into spatiotemporal networks	Liu et. al., 2024
Performance Analysis	Extracting transit performance indicators from GTFS-RT	Aemmer et al., 2022
Planning & Monitoring	Open-source framework to extract GTFS data	Lim, et. al., 2019
Validator	Allows the user to validate a timetable prior to the publication of a timetable using GTFS	Mobility Data, 2026

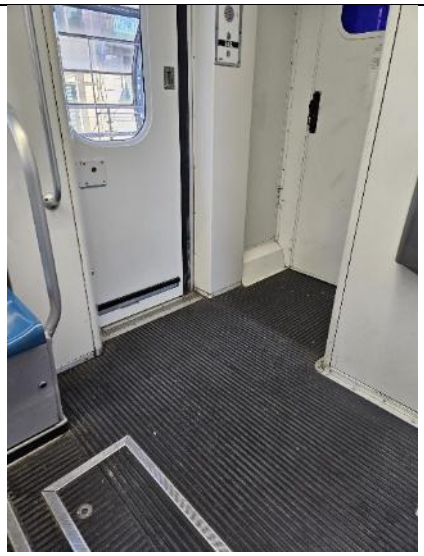
Despite there being no official utilisation of advanced algorithmic methods to develop timetables, this use case illustrates the potential feasibility of using advanced algorithms to create timetables and validates the methodology utilised by various studies. As highlighted, there are numerous aspects of metro operations that can be enhanced by advanced timetabling. The strong feasibility and clear benefits of generating timetables using advanced algorithms, is presented across the various case studies, and some of which tested based on real-life metro systems operational data around the world. With the Norwegian case-study taking this further and implementing an algorithmic timetable on a railway line. Across energy savings, improved passenger experience, enhanced capacity, disruption resilience, and more efficient fleet usage, the evidence demonstrates tangible operational improvements. Therefore, algorithmic timetable creation, alongside validated GTFS feeds, is both practical and advantageous for modern metro systems. Furthermore, leveraging GTFS data can provide further valuable insights for metro operators, with network analysis, performance analysis as well as, planning and monitoring all aspects of operations that GTFS data can support with enhancing metro operations.

While a validation of the complete cycle for Timetable Creation using GTFS feeds was not possible in the use case developed in this section, train dwell time and service frequency optimisation (essentially equivalent to optimising train timetables) will be undertaken as part of the comprehensive metro service operation improvement campaign in Task 7.4, and will be documented in D7.2.

3.2.3 USE CASE 3 – VALIDATION OF UNCLEANLINESS DETECTION

3.2.3.1 GENERAL INFORMATION

To support the validation of our proof of concept for uncleanliness detection, we received a set of eight images from the Metro Genova network. These samples were utilised to conduct a preliminary assessment of both our classification and object detection models in a real-world transit environment. This initial data exchange has been instrumental in testing the system's ability to identify various states of cleanliness and locate specific areas of concern within the station infrastructure.

Clean	Unclean
	
3	1
	
4	2

5

6

8

7

3.2.3.2 VALIDATION OBJECTIVE

The primary objective of this validation phase is to conduct a qualitative assessment of the model's performance within the specific environment of the Metro Genova network. With a focused set of eight images, the current evaluation serves as an important initial proof of concept rather than a basis for broader statistical metrics like precision or recall, which typically require a larger sample size to be representative. Instead, this stage allows us to confirm the model's functional alignment with real-world conditions and provides a valuable baseline for its predictive logic. These early observations are instrumental in guiding a future project's trajectory and defining the parameters for more extensive data collection and analysis in the future.

3.2.3.3 DATA BASIS

For the validation phase, we processed a total of eight images provided for testing. We applied our unmodified training process to these samples to generate initial predictions and assess performance in a real-world context. While the preliminary outputs offer interesting insights, the current dataset size is not yet sufficient to draw statistically significant or conclusive results. We look forward to expanding this evaluation as more data becomes available, which will allow for a more robust validation of the model's accuracy and reliability. Since one of our models is a classifier, we labelled them as clean and unclean, as seen below. For the object detection results, we provide only the images with detections.

3.2.3.4 RESULTS OF VALIDATION

3.2.3.4.1 CLASSIFICATION RESULTS

In our initial evaluation, the model correctly classified seven out of the eight provided images (**Table 56**). While this represents a high success rate for this specific batch, we recognise that the limited sample size serves primarily as a preliminary proof of concept rather than a comprehensive performance metric. These encouraging results provide a positive starting point, suggesting that the model's current architecture is well-positioned for more extensive validation as larger datasets become available in the future. Such expanded testing will be essential for establishing a statistically robust measure of the system's long-term accuracy and reliability.

Table 62: Vehicle interior uncleanliness detection model classification results summary

Image ID	Prediction	Confidence	Correctness
1	Unclean	0.790750384330749	✅
2	Unclean	0.751765847206115	✅
3	Unclean	0.737206339836120	❌
4	Clean	0.635131001472473	✅
5	Clean	0.799001514911651	✅
6	Unclean	0.835015892982482	✅
7	Unclean	0.717746078968048	✅
8	Clean	0.739433348178863	✅

3.2.3.4.2 OBJECT DETECTION

In addition to classification, we performed initial object detection on the eight provided images, where the model correctly identified the target components in seven instances. The single discrepancy observed was a false positive located at a ground-level anchoring point (or fastening assembly), which the system incorrectly flagged as a target object. This happened in multiple images, as can be seen below, pointing to a specific case that could be improved with more data. While these preliminary results

are encouraging, the occurrence of this false-positive highlights the need for a more extensive dataset to further refine the model's spatial precision and minimise detection errors.

Table 63: Vehicle interior object detection model results summary

Image ID	Prediction	Correctness
1	Unclean	✅
2	Unclean	✅
3	Clean	✅
4	Clean	✅
5	Clean	✅
6	Unclean	✅
7	Unclean	✅
8	Unclean	❌

See below the images where our object detection model produced bounding boxes around items of uncleanliness. Please note that we deem image 8 a misclassification.

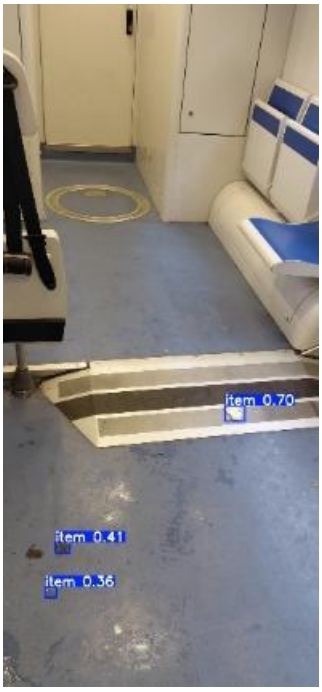
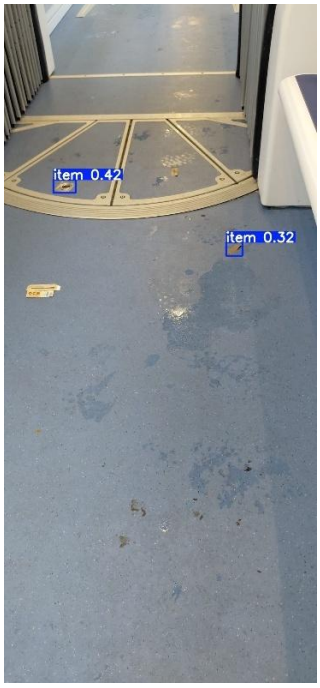
Image ID	IMAGE
1	

Image ID	IMAGE
2	

6

7

8

3.2.3.5 OPERATIONAL IMPACT & EXPLOITATION

The validation of our proof of concept demonstrates that the underlying methodology is ready for deployment in a semi-automated setting. We believe that the model's current performance provides a solid foundation for a human-in-the-loop learning framework, where human feedback can be used to iteratively refine and improve the system's accuracy over time. This approach, which we will detail in the following sections, will allow the model to learn from its mistakes and adapt to the specific nuances of the Metro Genova environment, ultimately leading to a more robust and reliable uncleanliness detection system.

3.2.3.6 VALIDATION – HUMAN IN THE LOOP

Validation is defined as the process of evaluating a machine learning model (Bishop and Nasrabadi, 2006). For this purpose, a dataset that has not previously been used in training is used to determine the performance of that model. This process is of paramount importance as it demonstrates whether the model has acquired the capacity for generalisation in relation to the problem it is attempting to solve.

There are a variety of techniques for the application of validation. However, it is most commonly divided into the following categories: K-fold cross-validation, Holdout method, and Nested cross-validation (Bhagat and Bakariya, 2025; Wu et al., 2022). In circumstances where the available dataset is limited, K-fold cross-validation is a frequently employed approach. In this regard, the dataset is segmented into K sub-sets. The model is trained K times using K-1 data sets, while validation is performed on K single sets. In this sense, each subset will be evaluated separately. Another technique that can be employed is the Holdout method, which is in fact the most common choice. The dataset is divided into training, validation and testing subsets in a given ratio. The final technique is Nested cross-validation. This configuration indicates the presence of two loops. The initial loop pertains to evaluation, wherein the trained model is evaluated on an unused dataset. The subsequent loop is employed for training, that is to say, the selection of parameters. This technique is more commonly used in hyperparameter tuning methods.

3.2.3.6.1 HUMAN IN THE LOOP

Human-in-the-loop (HITL) is the process in which a human is actively involved in the machine learning process to integrate knowledge (Mosqueira-Rey et al., 2023). The fundamental function of the model is to regulate the learning process in a manner that rectifies erroneous predictions, thereby enhancing the model's accuracy and refining its performance (Jakubik et al., 2023). In this regard, the responsibility for identifying anomalies in the prediction typically falls upon an expert in the relevant domain. This role assumes particular importance in the context of sensitive tasks, such as medical imaging (Holzinger, 2016). As the machine model remains a "black box", erroneous predictions are frequently uninterpretable. In this sense, human expertise can facilitate the evaluation of model predictions, thereby ensuring their accuracy. Furthermore, human input can provide supplementary explanatory and contextual information, thus promoting the enhancement of model development and facilitating a more profound comprehension of the data (Kang et al., 2021).

Nevertheless, it must be noted that this approach also has its drawbacks. The primary concern pertains to the significant temporal expense associated with incorporating a human element into the process. Data sets are often very large, especially when it comes to images. In this sense, it is incumbent upon the individual to undertake a manual review of all errors and to effect the necessary corrections, a process which has the effect of delaying the retraining time. It is evident that sensitive data, such as X-ray images, necessitates the involvement of experts from the respective domain, which in turn, results in a significant delay in the overall process. Moreover, the possibility of human error is heightened in circumstances where the data and the context of problem resolution are of a particularly complex nature (Pandey et al., 2022). In this regard, there may be a certain degree of bias towards human predictions that may ultimately prove to be erroneous.

HITL often implies three types of processes, namely: reinforcement learning, supervised learning, and active learning. In all three approaches, data annotation correction is also done. However, active learning requires at least human involvement in the process. It is a type of learning in which the model flow selectively selects an unlabelled dataset that is moved to a given pool, and annotation is requested. In this example, the uncertainty is defined by some percentage, indicating that a label needs to be added so that the model can make an accurate prediction.

3.2.3.6.2 IMPROVING MACHINE LEARNING OPERATIONS WITH HITL

MLOps brings together machine learning, DevOps, and data engineering to make it easier to build, deploy, track, and maintain ML models (Kreuzberger et al., 2023). Its main goals are to improve efficiency, ensure reliability and reproducibility, automate repetitive tasks, and manage both models and data effectively throughout their lifecycle. The involvement of human operators in these processes has been demonstrated to yield substantial benefits. A prevalent issue in practical settings pertains to the paucity of data that is conducive to the training of models. It is within this process that human involvement through online annotation can be facilitated, thereby enhancing the efficiency of the model.

The initial step in this process is to generate a data set for the purposes of testing, validation, and training. The fundamental supposition is that the data has been annotated in a manner that incorporates class and bounding boxes as components of object classification and detection. This initial training and validation are regarded as the primary golden standard, to which all subsequent versions and validations will be compared. Subsequent to the initial training and validation, the preliminary evaluation is conducted. In the classification process, the primary focus is on the F1-score metric, with particular emphasis on precision and recall. The confusion matrix must be incorporated into the internal evaluation process.

Following the completion of the training and testing of the model, and the storage of all the necessary evaluation metrics on the system, the model can be integrated into the CCTV system. Since predictions are made in real-time, to reduce system load, it is sufficient to make a prediction at intervals of between 30 and 60 minutes by selecting one frame from the video streams. For the specified frame, the model generates the following predictions: binary class, building box, and confidence score.

The confidence score is the main metric we use during online testing (Geifman and EI-Yaniv, 2017). We need to pick a value that tells us when a human needs to step in or not. For example, the model gives a confidence score of 0.3 for a frame. That is just 30% sure about its prediction, whether it is a 0 or a 1. When that happens, this frame has to be extracted and saved, along with its class, bounding box, and confidence score, and saved in a separate database. But to make human annotation worth it, it should not do this for just one frame. There needs to be at least N frames as a batch before involving someone to check or label them.

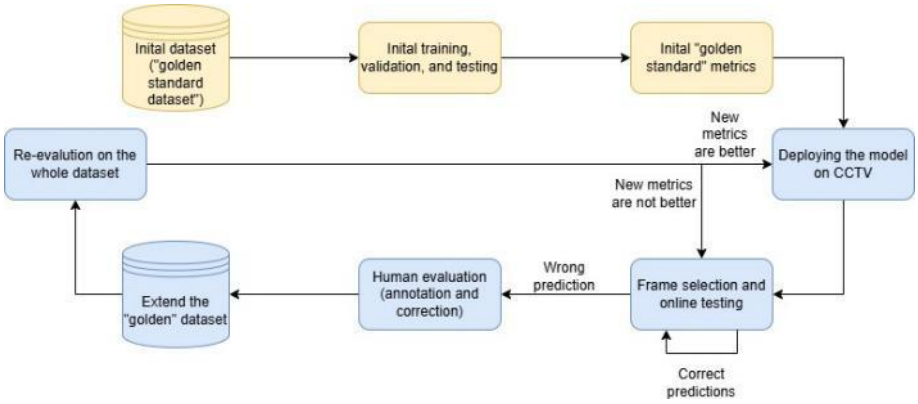
In the subsequent phase, human intervention is requisite. The human reviews all the data in the system and confirms or corrects it as required. It is possible to make corrections to the classification or correction of the building box. Furthermore, in instances where the building box is not found, the addition of a label is necessary. Subsequent to the implementation of the correction, all annotations are migrated to a predefined file, which is frequently a CSV file. Concurrently, a new training and validation with a more substantial data set is initiated automatically. Should the newly derived validation values prove

satisfactory following initial validation, the updated dataset will be versioned into the system. Following this, the new version of the model will be re-integrated into the CCTV system.

This process is repeated periodically, which means that in each iteration the same steps are performed: annotation and correction of new data, versioning of the new data set, evaluation of the model, and integration of the new model into the system. The frequency with which this process is repeated depends on the system's configuration and the internal decisions made. The whole described pipeline can be seen on **Figure 19**.

In order to initiate the process of optimisation and empirical validation, it is recommended that the proposed system be deployed at a single metro station as part of a controlled pilot implementation. Restricting the deployment to one localized environment enables systematic monitoring, minimises operational risk, and allows for structured testing of the technical infrastructure under real-world conditions. Consequently, all online testing procedures, in addition to the integration and iterative refinement of newly developed models, should be conducted within this operational setting.

The pilot phase should extend over a clearly defined evaluation period, for ex-ample one month, to ensure that the system is exposed to varying operational conditions and environmental influences. This duration is deemed sufficient to collect both quantitative and qualitative data, which can then be used to conduct a meaningful performance assessment.



Note: The yellow colour represents the steps in the first iteration, while the blue colour represents other iterations.

Figure 19: The human-in-the-loop validation pipeline.

Following the pilot testing phase, a comprehensive evaluation should be con-ducted. The evaluation must include all previously defined metrics, such as the model's accuracy, precision, recall, and false alarm rates. Additionally, feedback from operational staff should be obtained. Employees should be invited to evaluate the accuracy, actionability, and benefit of the cleaning alerts in improving workflow efficiency. If the evaluation showed good performance, then the system can subsequently be scaled and implemented across all metro stations and associated cameras.

3.3 EXAMPLARY RESULTS ON ENERGY AND CO_2 OPTIMISATIONS

3.3.1 INTRODUCTION

The transition to Grade of Automation 4 (GoA4) metro systems—where trains operate without onboard staff — presents a fundamental strategic question for transit authorities: should automation savings be captured as cost and emissions reductions, or reinvested into enhanced service? This section demonstrates that demand-responsive scheduling with short depot breaks, enabled by full automation opens up optimisation space to pursue either objective.

The goal of this work is not to present exact simulation values, but rather to:

- **Outline a future use case** of intelligent metro systems enabled by automation
- **Demonstrate how automation unlocks** a previously uneconomic optimisation area: short depot stops during demand lulls
- **Promote the development of an online tool** where metro operators and researchers can configure line parameters and optimisation targets to generate their own context-specific results

The analysis centres on energy consumption and (the lack of) personnel costs—two factors fundamentally influenced by automation through optimised train control algorithms, intelligent power peak management, and coordinated regenerative braking (Wang, et al., 2016).

3.3.2 METHODOLOGY

3.3.2.1 TOTAL COST OF OWNERSHIP (TCO) TOOL

To enable cost-based optimisation, an internal Total Cost of Ownership (TCO) tool was integrated into the simulation framework for this particular task. This tool captures granular cost components as optimisation targets (**Table 64**):

Table 64: Summary of cost components in the Total Cost of Ownership (TCO) tool

Cost Component	Description
Power draw by state	kW consumed during accelerating, decelerating, cruising, dwelling, and storage states
State-specific costs	Cost incurred from power consumption in each operational state
Power peak effects	Estimated values for caused power peaks (penalties) and prevented peaks (savings)
CO_2 emissions	Carbon footprint based on energy consumption and regional grid intensity
Maintenance costs	Track wear, wheel degradation, and component lifecycle costs

This TCO framework allows optimisation algorithms to target specific cost functions rather than simplified proxies, enabling more realistic economic analysis.

3.3.2.2 SIMULATION APPROACH: GENOVA METRO CASE STUDY

In our single-line Genova metro simulation, implemented as a discrete-event model in SimPy. SimPy is a powerful, process-based framework for discrete-event simulations built on Python. System evolution is governed by the environment clock while trains and passengers are encoded as interacting processes. Each train process advances through successive blocks of motion and station operations by yielding timeouts for running times between stations and dwell events for boarding and alighting, while enforcing line constraints such as minimum headways and platform occupancy. Passenger demand is generated as station-specific arrival processes and realized as origin–destination trips, with passengers joining platform queues, boarding subject to service availability and capacity, and completing trips upon alighting at their destinations. Throughout the 24-hour horizon, the model records event timestamps and state transitions, enabling the computation of spacetime trajectories, passenger completion outcomes, travel-time distributions, and operational/energy-relevant indicators such as concurrent acceleration episodes and per-train state-time allocations.

The simulation is based on passenger flow analysis from a previous deliverable, using the Genova metro line as the test case:

- Passenger Generation: 45,000 synthetic passengers generated with realistic arrival time-of-day distributions, origins, and destinations. Identical for all simulations.
- Train Spawning: Trains spawned randomly according to configurable parameters.
- Simulation Duration: Full 24-hour cycle with continuous data recording.
- Dwell Behaviour: Trains dwell at stations for passenger boarding and alighting.
- Conflict Resolution: If a train enters the final 10% of an incoming segment, any dwelling train in the station ahead is evicted; otherwise, trains dwell until target dwell time is reached.

For the demand distribution we assumed the same distribution for each station and approximated the distribution that was found in Deliverable 6.1 and presented in the chapter 2.1.1 in that deliverable.

In the simulation a tick is defined to last one minute, however arbitrary sizes of subticks and therefore any subminute or even subsecond events are happening and are recorded accurately.

3.3.2.2.1 PASSENGER DISTRIBUTION

The origin-destination heatmap in **Figure 20** highlights the spatial distribution of travel demand, revealing the most heavily utilised station pairs. By identifying these high-traffic corridors, the simulation provides critical insights into where platform congestion is most likely to occur and where service capacity is most vital. Note that stations with and without “backwards” in their names simply encode the direction of the platform but are the same destination. 0-entries simply encode inefficient trips that require a detour to complete, which are not part of our simulation.

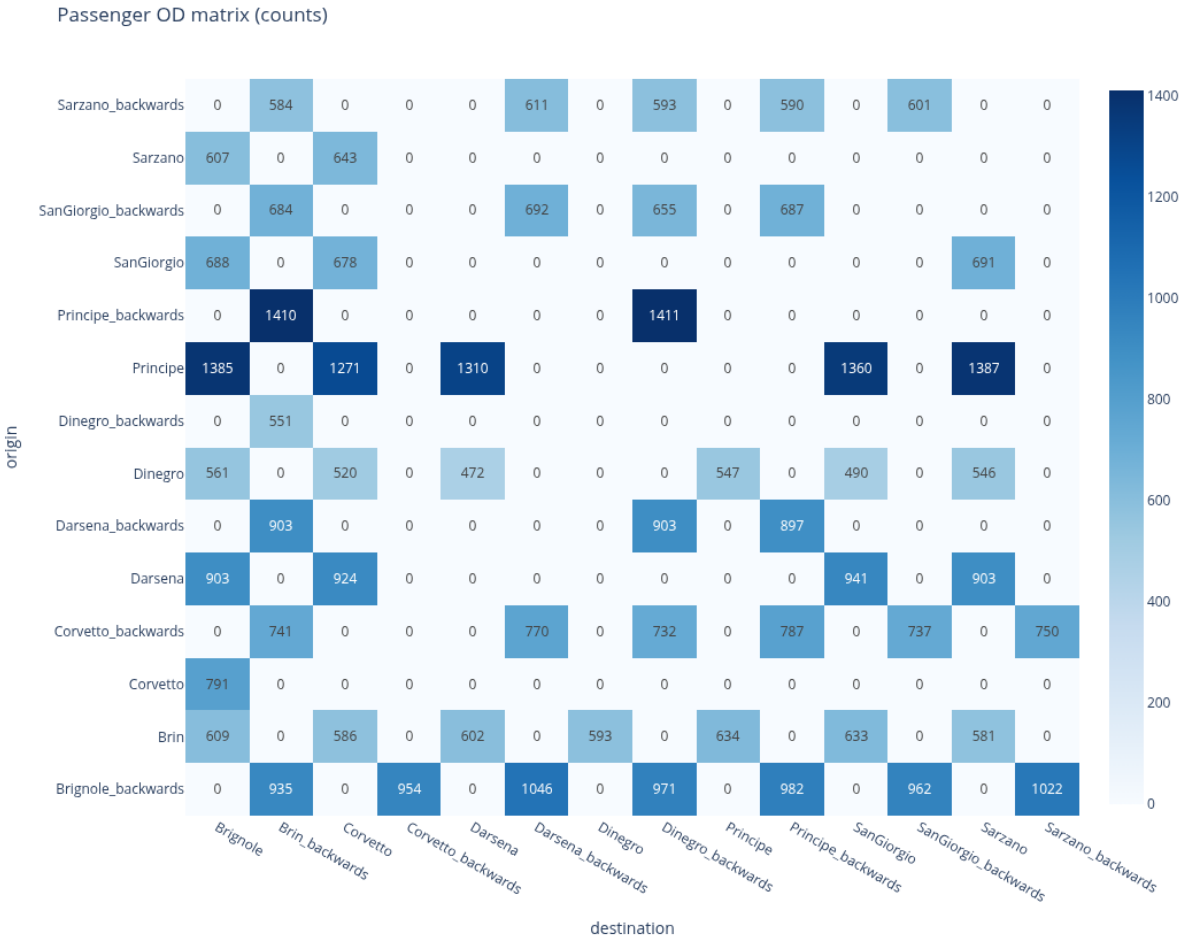


Figure 20: OD Heatmap for cost optimisation purposes

3.3.2.3 OPTIMISATION PARAMETERS

The following parameters were subject to optimisation (**Table 65**):

Table 65: TOC optimisation parameters

Parameter	Description
Target dwell time	Base dwelling duration at station
Passenger threshold for depot entry	Trains enter depot when passengers waiting ahead falls below this threshold percentage of capacity
Depot check frequency	How often trains in depot evaluate the threshold above to re-enter service
Number of trains	Fleet size deployed in simulation

For our demonstration we created a sample optimisation target. However, when we publish the simulation, we intend to make it as easy as possible to calculate custom optimisation targets from the metrics recorded by trains and passengers. We designed our optimisation target using the following components:

Table 66: Summary of Optimisation Target Components

Metric	Description
S: Total Passenger Travel time in seconds	Cumulative travel time of all passengers
C: Total Cost of Ownership	Cumulative costs calculated by our TCO tool
U: Uncompleted trips	Uncompleted trips by simulation termination
T: Number of trains	Count of trains deployed in the simulation

We employ the Bayesian Optimisation algorithm (Kim and Choi, 2023; Gardner, 2014; Garrido-Merchán and Hernández- Lobato, 2020) to minimize our target loss function. The cost function that was then minimized is the following:

$$\mathcal{L} = S \cdot (\alpha) + C \cdot (1 - \alpha) + U + T \cdot 450$$

Notably, this cost function uses an alpha parameter to balance between passenger travel time and cost. The value we used was experimentally chosen for stability. Choosing a good trade-off between cost-savings and passenger travel times, as a proxy for service quality, is central to get realistic and desirable results. Defining the optimisation objective is a governance decision rather than just a technical result, as it requires a strategic and intentional policy choice between maximizing service quality and minimizing total costs (Li et al., 2023).

We ran the optimiser for 100 iterations, with each iteration simulating a full day under a new parameter configuration. A total of 100 iterations was selected to provide a sufficiently broad sample of the parameter space while remaining computationally feasible, allowing us to evaluate a diverse range of configurations and identify those that produced the strongest performance.

3.3.3 RESULTS

After running our simulations and optimising at each iteration we found the following best-scoring solution. Some of the discussed qualities of the solution relate directly to the metro system and are thus tied to metro operations, while the passenger-related metrics serve to gauge the quality of service and to assess whether our simulation approximates reality.

3.3.3.1 SPACETIME DIAGRAMS

The spacetime diagrams illustrate the continuous movement of trains across the network, where the slope of each trajectory represents the operational speed between stations. This visualization confirms the consistency of headways and identifies the rhythmic nature of the 24-hour service cycle, including peak and off-peak frequency adjustments.

In our simulation of the Genova line, depots are located in the stations Brin_backwards and Brignole. No trips cross through either of these stations, so trains are always empty when they enter the depot. As can be seen from the spacetime diagram of this single train, it enters the depot at Brin_backwards for almost 200 minutes during the morning hours, as demand is low.

In **Figure 22**, we see another train, which happened to enter the depot for multiple short stays, between 5 and about 15 minutes long. **Figure 23** is the full spacetime diagram for completeness, however as a static plot it is less readable.

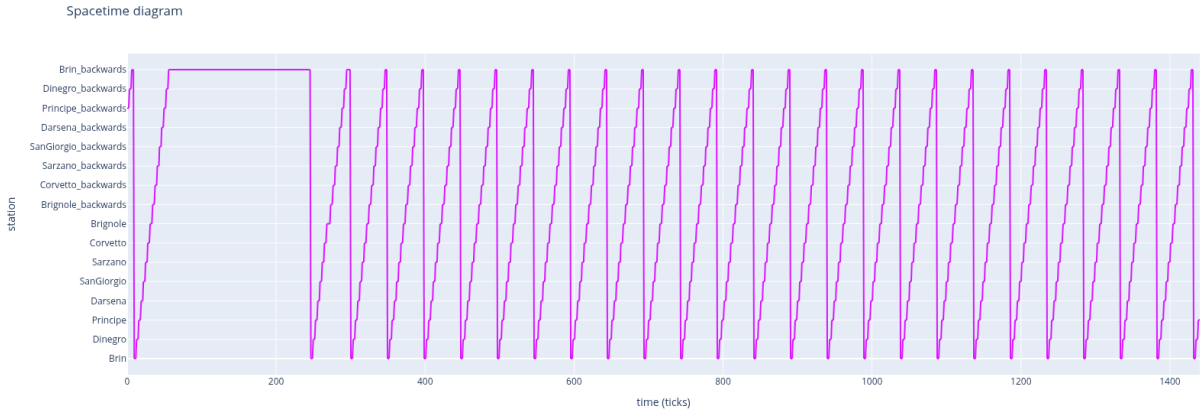


Figure 21: Single Train Movement Spacetime Diagram (Example 1)

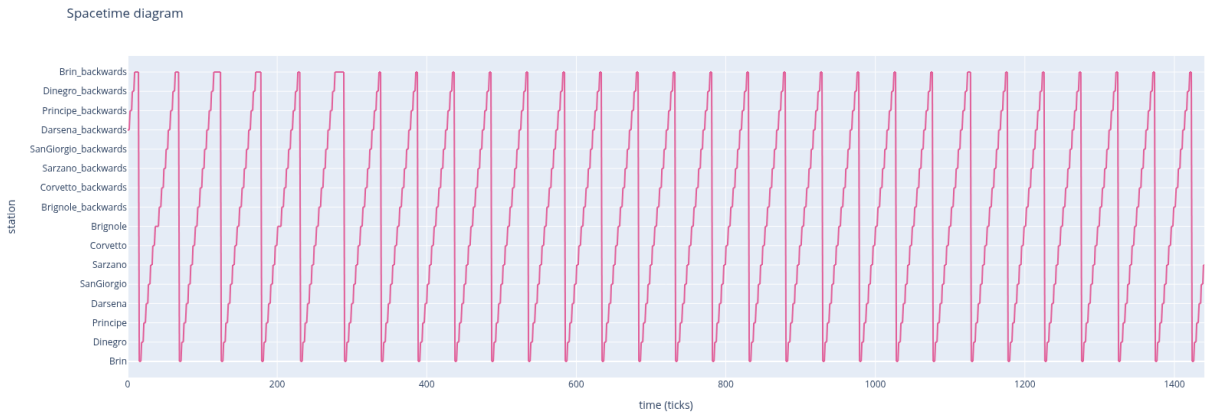


Figure 22: Single Train Movement Spacetime Diagram (Example 2)

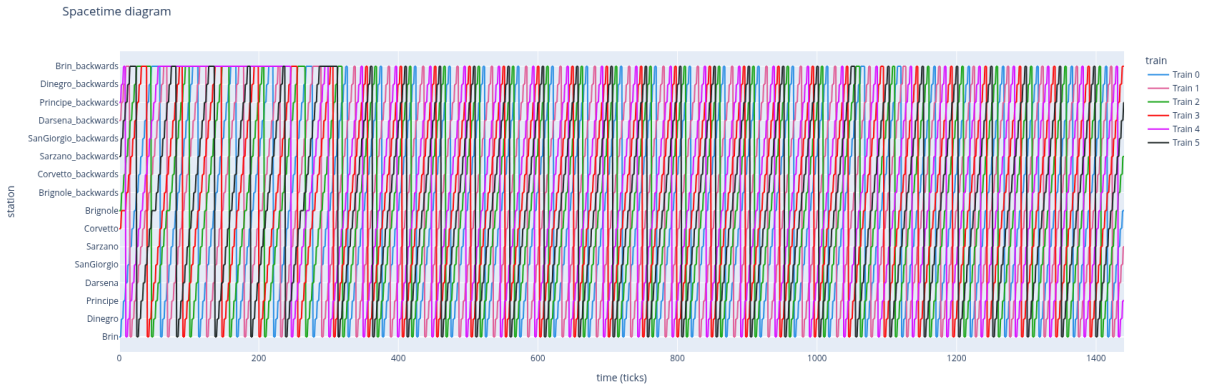


Figure 23: Multiple Trains Movement Spacetime Diagram

3.3.3.2 TRAIN STATE TIME DISTRIBUTION

Figure 24 categorises the total simulation time for each train into distinct operational states, such as cruising, dwelling, and accelerating. This breakdown informs TCO calculations by highlighting the proportion of time spent in energy-intensive states versus passive storage or dwelling.

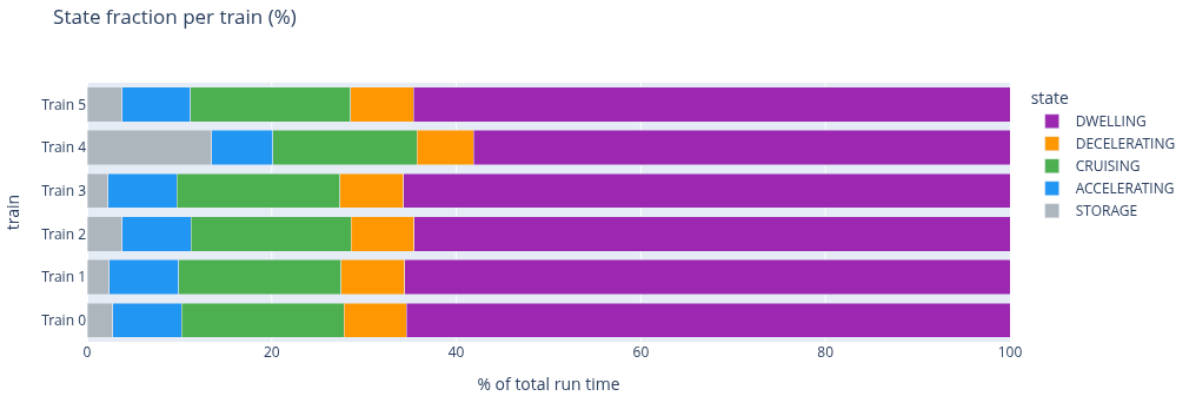


Figure 24: Train operational state time distribution diagram

As we can see, trains efficiently travel through the segments, minimizing the risk of an incident happening within a tunnel segment. Most time is spent dwelling for each train, however all trains manage to enter storage during the day, for 2.2% up to 13.5% of a 24-hour day. Achieving longer dwell times also helps reduce cost and allows passengers more time to alight and board.

3.3.3.3 PASSENGER METRICS

The average total time of passengers entering their station of origin and leaving at their destination was 689.5 seconds or slightly under 12 minutes. This can of course be tuned in any direction by adjusting the optimisation targets parameters.

Figure 25 displays the statistical distribution of waiting times, in-vehicle journey times, and total trip durations across the simulated population. These histograms quantify the quality of service, showing the variance in passenger experience and the efficiency of the schedule in minimizing idle time at platforms.

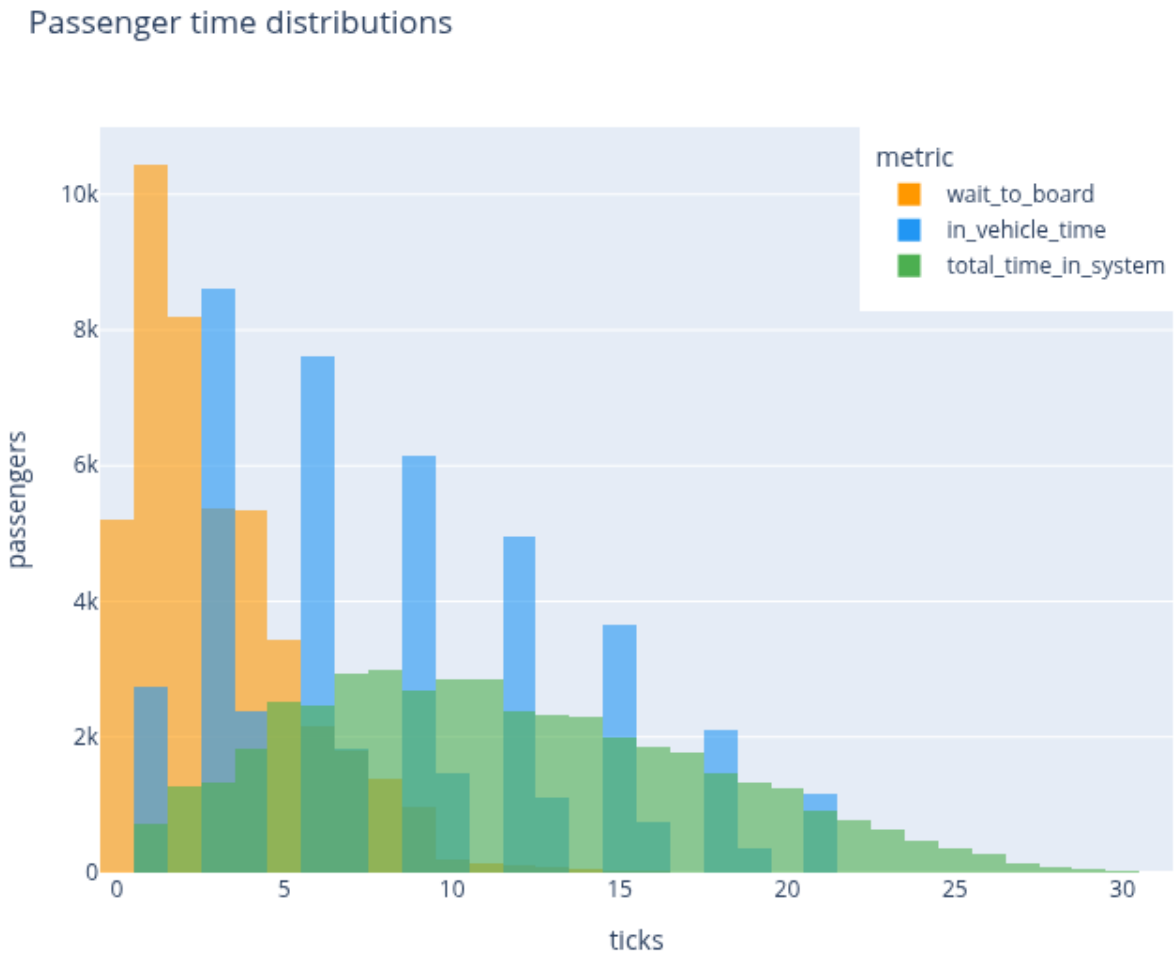


Figure 25: Time distributions of passenger activities (waiting times, in-vehicle journey times and trip durations)

3.3.3.4 CONCURRENT ACCELERATION ANALYSIS

The analysis of simultaneous train accelerations in **Figure 26** identifies critical periods of high instantaneous power demand. Because concurrent starts drive significant peaks in electricity consumption and infrastructure stress, this plot is essential for evaluating the economic and technical feasibility of the power supply system. We can see that despite operating 5 trains at the same time, besides rare instances, at most 2 trains accelerate at the same time. Concurrent decelerations – where regenerative breaking could be used – are not visualized in this plot.

To summarise, this section presented a comprehensive case for automated metro operations by evaluating the intersection of TCO and demand-responsive scheduling. The simulation results demonstrate that when optimising for a balanced objective of operational expenditure and service quality, unconventional scheduling patterns—such as frequent, short-duration depot withdrawals—emerge as an optimal strategy. While such manoeuvres were historically precluded by the high marginal costs of human labour, they arise naturally in an automated context and are preserved by the

optimisation algorithm as a means of precisely aligning supply with fluctuating demand. Ultimately, the released simulation framework will hopefully provide useful as a robust analytical tool for transit authorities; by incorporating site-specific cost parameters, empirical passenger flows, and precise infrastructure layouts, stakeholders can derive high-fidelity insights to inform the transition toward next-generation, autonomous urban mobility systems.

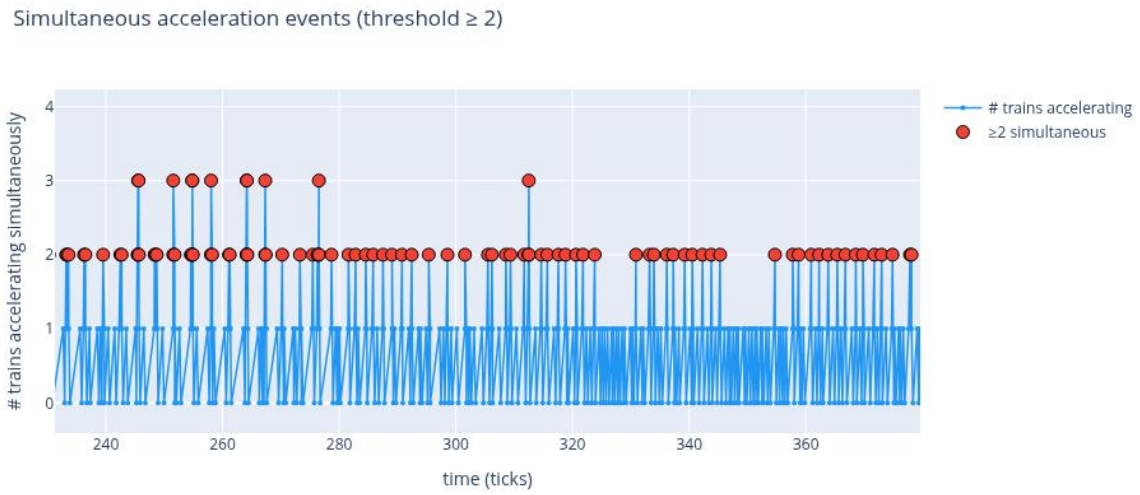


Figure 26: Trains simultaneous acceleration events analysis diagram

3.3.4 SUMMARY

We set out to determine whether brief intermittent depot stops, an optimization avenue newly unlocked by GoA4 operation, constitute a genuinely viable cost-optimization strategy. By deploying an unbiased optimization algorithm that was given no a priori preference toward this behaviour, we have shown that the strategy is not merely plausible in theory but is independently discovered and actively employed when the solution space is searched for high-quality solutions. The fact that an impartial procedure converges on this strategy provides strong evidence that it confers a real advantage rather than reflecting any assumption built into our analysis. We therefore conclude that intermittent depot stops warrant serious consideration as a practical cost-optimization measure in real-world GoA4 operations, and that the broader methodology demonstrated here offers a principled means of surfacing and validating such previously inaccessible strategies.

3.4 CYBERSECURITY

This section, in alignment with the validation and exploitation activities described, builds on the completed cybersecurity baseline assessment and the defined priority improvement areas, and provides a forward-looking security validation layer for the AI applications and simulation-driven operational concepts addressed

In practical terms, the security work products defined in 3.4 are intended to be applied consistently across each AI/Data Science demonstrator and use case described in the D7.1 structure (e.g., General Information, Validation Objective, Data Basis, Results of Validation, and Operational Impact & Exploitation). For each use case the section contributes to: (a) an AI lifecycle and trust-boundary view (data ingestion, training, deployment, inference) aligned with the demonstrator's data basis and interfaces; (b) a set of MITRE-informed adversarial scenarios (MITRE ATLAS for AI systems, and MITRE ATT&CK/ICS where IT/OT interfaces are relevant) to be treated as "must-test" items within validation; and (c) concrete safeguards and gating criteria that determine when a use case can move from experimentation to pilot operation and, subsequently, to production-like exploitation.

The section also leverages the wider D7.1 context on simulation models and future operational characteristics (including disrupted conditions, increased automation, and stronger reliance on telemetry) as realistic stress conditions for security validation. This ensures that cybersecurity recommendations are operationally meaningful for a large metro operator: they focus on maintaining integrity, availability, auditability and safe rollback under both normal and degraded operational scenarios, while enabling evidence-based adoption of AI-enabled capabilities.

3.4.1 OBJECTIVES AND SCOPE

As mentioned, the analysis builds upon the completed cybersecurity assessment and gap analysis, which established the current security baseline and identified priority improvement areas. The present work is forward-looking and focuses on the cybersecurity implications of adopting AI models and data-driven applications for future metro operations management and planning.

The analysis adopts a threat-driven perspective, examining how AI-enabled capabilities may alter the attack surface, trust boundaries, and risk profile across metro IT and operational environments. AI use is treated in a model-agnostic and vendor-agnostic manner, considering generic AI/ML lifecycles (data ingestion, training, deployment, inference) and their interfaces with operational systems and future train control characteristics.

3.4.2 CYBERSECURITY EXPLORATION AND VALIDATION OF AI-ENABLED SYSTEMS

The objectives of this activity are to:

- Identify and describe plausible AI-enabled use cases relevant to metro operations management and planning (including exploration activities related to future train control characteristics).
- Analyse adversarial behaviours and threat patterns that are specific to AI systems and their data pipelines and compare the current environment with a future AI-enabled scenario.
- Derive strategies and measures for secure data handling, reliable AI operation, and cybersecurity safeguards that ensure safe validation, deployment and exploitation of AI applications.
- Provide actionable recommendations that can be integrated into future architecture, procurement, and operational processes.

The analysis leverages on the results of the full baseline control assessment already delivered and extends them with AI-specific threat modelling and mitigation guidance.

3.4.3 APPROACH AND METHODOLOGY

The methodology is structured in four steps, aligned with a security audit / risk assessment workflow:

1) Outline the AI-enabled target scenario(s): define representative AI use cases, data sources, integration points, and operational dependencies.
2) Identify trust boundaries and attack surfaces introduced or expanded by AI: data flows, model lifecycle assets, interfaces to OT/ICS and future train control components.
3) Threat-driven analysis using MITRE-based AI threat patterns: map adversarial behaviours across the AI lifecycle and the metro environment.
4) Derive strategies and measures: translate threat patterns into layered safeguards (governance, process, technical controls, monitoring, and assurance testing).

For threat patterning, the analysis uses:

- MITRE ATT&CK® (including the ICS Matrix, where OT/ICS interfaces are relevant) to capture adversarial behaviour in traditional IT/OT environments.
- MITRE ATLAS™ (Adversarial Threat Landscape for AI Systems) to capture AI-specific tactics and techniques (e.g., data poisoning, model theft/exfiltration, adversarial inputs, prompt injection for LLM-based assistants).

The output is a structured comparison between (a) the current baseline environment and (b) a future scenario with increased AI adoption, highlighting deltas in exposure and the compensating controls required.

3.4.4 AI-ENABLED METRO OPERATIONAL SCENARIOS

To remain model-agnostic, AI adoption is described using a small number of representative scenario archetypes. These archetypes can be mapped to specific demonstrators and future deployments as details mature.

Cross-reference to D7.1 use-case template: the scenario archetypes below can be instantiated per demonstrator under the “Data Basis” and “Results of Validation” fields, ensuring consistent security validation inputs and evidence across all use cases.

Table 67: AI enabled scenarios

SCENARIO ARCHETYPE	TYPICAL PURPOSE	PRIMARY DATA INPUTS	OPERATIONAL TOUCHPOINTS / DEPENDENCIES	SECURITY RELEVANCE
A1 Demand & service planning analytics	Ridership forecasting, timetable optimisation, scenario planning	Historical ridership, events, weather, service logs, station counts	Planning tools, data lake/warehouse, reporting platforms	Data integrity, data governance, access control, auditability
A2 Asset & maintenance optimisation	Predictive maintenance, failure prediction, condition-based interventions	Sensor/telemetry streams, maintenance history, work orders	CMMS/EAM, OT data brokers, maintenance operations	Telemetry integrity, model drift, safety impact, supply chain security

SCENARIO ARCHETYPE	TYPICAL PURPOSE	PRIMARY DATA INPUTS	OPERATIONAL TOUCHPOINTS / DEPENDENCIES	SECURITY RELEVANCE
A3 Operational decision support (near-real time)	Decision support for dispatch, incident handling, crowding control	Real-time train positions, headways, station occupancy, alarms	Operations control centre (OCC), SCADA/ICS interfaces, comms systems	Trust boundaries with OT/ICS; resilience and monitoring; tamper detection
A4 AI assistant for operations knowledge	Search/assist with procedures, incident playbooks, engineering knowledge	Documents, runbooks, tickets, knowledge base, logs (sanitised)	SOC/OCC workflows, service desk, engineering teams	Prompt injection/data leakage; access control; logging and oversight

Key architectural building blocks typically required across the above archetypes include: data ingestion pipelines, storage layers (data lake/warehouse), feature engineering, model training/validation environments, model registry/versioning, deployment/runtime services (APIs), and monitoring/telemetry (for both security and model performance).

3.4.4.1 FUTURE METRO CONTROL CHARACTERISTICS: ASSUMPTIONS AND SECURITY IMPLICATIONS

As the project includes exploration activities on the characteristics of future metro train control systems, the cybersecurity analysis adopts a small set of conservative, technology-agnostic assumptions to avoid speculative design. These assumptions are used solely to stress-test the threat model and to identify where the baseline improvement areas must be reinforced to safely accommodate increased data exchange, automation and third-party integration. Where D7.1 evaluates performance under disrupted conditions and higher automation levels, the same stress conditions should be used to validate the security assumptions, monitoring coverage and rollback procedures described

- Increased connectivity and telemetry: Future train control and operations platforms are expected to exchange more telemetry and diagnostic data (including near-real-time feeds) to support optimisation, condition monitoring and improved decision support. As a security implication the OT/IT boundary becomes more data-intensive and therefore requires brokered access, integrity protection and monitoring for anomalous transfers.
- More distributed / edge components: To reduce latency and improve resilience, parts of data processing or inference may occur closer to the edge (stations, depots, wayside equipment). Security implication: The asset inventory and configuration management must extend to edge nodes; therefore, the secure remote administration and HW/SW elements patching becomes even more critical.
- Tighter coupling between operational decision support and control-adjacent functions: Even when AI is ‘advisory’, its outputs may influence dispatching, incident handling and maintenance decisions. Therefore integrity and auditability of AI outputs become safety-relevant and fail-safe behaviours and rollback procedures must be specifically adapted.
- Expanded supplier ecosystem and remote support: Modern train control ecosystems often involve multiple vendors, integrators and remote support channels. Security implication:

third-party access paths, software supply chain, and contractual evidence requirements become more important for AI components, datasets and platform services.

- Stronger governance expectations (safety, compliance, assurance): As systems become more connected and data-driven, governance will require demonstrable evidence of security controls and validation outcomes. From the security perspective documentation, monitoring evidence, and acceptance criteria (gates) must be treated accordingly and proactively

A critical requirement is the integration of Cybersecurity with Operational Safety. Any output generated by AI models—even when serving in an advisory capacity—must not have the authority to override the Hard Logic or Fail-safe fallback mechanisms of the train control systems. In the event of detected cyber-anomalies or 'AI hallucinations,' the system must automatically revert to a predefined Safe State, ensuring that metro operations remain strictly within the network's safety parameters.

3.4.5 APPLICATION OF GATING CRITERIA AND MITRE THREAT TAXONOMY TO VALIDATED AI USE CASES

To ensure consistency between the AI validation activities presented in Section 3.2 and the trustworthiness assessment framework described in this section, the defined gating criteria and MITRE ATLAS-based threat taxonomy were applied to the three AI use cases validated within NEXUS: (i) passenger demand forecasting, (ii) timetable creation and optimisation, and (iii) vehicle uncleanliness detection.

The passenger demand forecasting use case supports operational planning through predictive analytics based on historical and operational transport data. The use case satisfies the defined gating criteria, as it does not involve automated decision-making affecting individuals, relies on structured operational data, and allows for human oversight of all resulting recommendations. From a threat perspective, the most relevant MITRE ATLAS threats relate to training data poisoning, data manipulation and model evasion attacks, which could affect forecast accuracy and subsequently influence operational planning decisions. Appropriate data quality controls, model monitoring and periodic retraining procedures are therefore recommended.

The timetable creation and optimisation use case provides decision-support capabilities for metro service planning. Although optimisation outputs are intended to assist planners rather than automatically control operational systems, the use case has a more direct influence on service provision and resource allocation. Consequently, stronger human oversight and validation procedures are required before implementation of any optimisation results. Relevant MITRE ATLAS threats include data poisoning, adversarial manipulation of optimisation inputs and model integrity attacks that could influence generated schedules. Human review of optimisation outputs and operational validation prior to deployment constitute key mitigation measures.

The vehicle uncleanliness detection use case applies computer vision techniques to identify cleanliness-related issues in rolling stock. The use case operates as a decision-support and inspection-assistance tool and therefore presents limited operational or safety impact. Human operators remain responsible for verifying detected issues and initiating maintenance actions. The most relevant AI-specific threats include data poisoning during model training and evasion attacks through manipulated

image inputs. The human-in-the-loop validation process, together with continuous collection of new image data, contributes to maintaining model robustness and improving performance over time.

Overall, the application of the gating criteria and MITRE ATLAS threat assessment demonstrates that all three validated AI use cases can be considered suitable for continued development and validation within the NEXUS framework. While the use cases differ in their operational impact and risk profile, all benefit from the presence of human oversight mechanisms and from the adoption of cybersecurity measures aimed at mitigating AI-specific threats. Future validation activities under WP7 will continue to assess both technical performance and trustworthiness aspects as the solutions mature towards operational deployment.

3.4.6 ALIGNMENT WITH THE EUROPEAN AI ACT

The AI trustworthiness assessment framework presented in this deliverable was primarily developed based on cybersecurity, privacy and operational risk considerations, including compliance with applicable data protection requirements such as the General Data Protection Regulation (GDPR). In addition, the principles established by the European Union Artificial Intelligence Act (AI Act) were considered to assess the potential regulatory implications of the validated AI use cases.

Based on the current scope and maturity of the solutions evaluated within NEXUS, none of the three AI use cases (passenger demand forecasting, timetable creation and optimisation, and vehicle uncleanliness detection) are expected to fall under the prohibited AI practices defined by the AI Act. Furthermore, all use cases operate as decision-support tools and maintain meaningful human oversight throughout the decision-making process. Nevertheless, future operational deployments may require a more detailed assessment of the applicable AI Act obligations, particularly in cases where AI systems contribute to operational planning, resource allocation or critical transport service management.

The AI Act introduces several principles that complement the existing gating criteria defined in this deliverable. These include requirements related to human oversight, transparency, technical robustness, cybersecurity, data governance, accuracy monitoring, logging and traceability. Consequently, future versions of the NEXUS trustworthiness assessment framework may incorporate additional gating criteria derived from the AI Act to ensure alignment with evolving European regulatory requirements.

In particular, future assessments may evaluate whether AI systems:

- maintain appropriate human oversight and intervention capabilities;
- provide sufficient transparency regarding system outputs and limitations;
- implement adequate data governance and data quality controls;
- demonstrate technical robustness and resilience against cyber and AI-specific threats;
- maintain logging and traceability mechanisms supporting accountability and auditing; and
- support continuous monitoring of performance, accuracy and reliability throughout their operational lifecycle.

The inclusion of these considerations will further strengthen the responsible and trustworthy deployment of AI technologies within automated metro transport environments and contribute to alignment with emerging European regulatory expectations.

3.4.7 THREAT LANDSCAPE: HOW AI CHANGES THE ATTACK SURFACE

Compared to traditional IT applications, AI-enabled systems introduce additional high-value assets and novel failure modes. The attack surface expands across (i) data, (ii) models, (iii) pipelines and orchestration, and (iv) human interaction points:

Table 68: Threat landscape changes

AI LIFECYCLE ELEMENT	NEW / AMPLIFIED ATTACK SURFACE	TYPICAL ATTACKER OBJECTIVES	REPRESENTATIVE THREAT EXAMPLES (NON-EXHAUSTIVE)
Data ingestion & labeling	External data sources, ETL jobs, labeling tools, ground truth pipelines	Corrupt outputs, degrade decision quality, create unsafe recommendations	Data poisoning; label flipping; compromise of upstream data feeds; tampering with ground truth
Training & experimentation	Training environments, notebooks, GPU compute, secrets in pipelines	Steal IP, implant backdoors, access sensitive datasets	Model backdooring; training-time compromise; credential theft from notebooks; dependency attacks
Model registry & CI/CD (MLOps)	Model artefact stores, versioning, promotion gates, signing	Deploy malicious model, downgrade controls, persist access	Model swapping; registry compromise; bypass of approval gates; artefact tampering
Inference / runtime services	Public/private APIs, edge deployments, integration with operational tools	Manipulate outputs, cause service disruption, exfiltrate data/model	Adversarial inputs; API abuse; model extraction; denial-of-service on inference endpoints
Human-facing AI (assistants)	Prompts, conversation memory, plugins/tools, knowledge bases	Data leakage, unsafe actions, privilege escalation via tool use	Prompt injection; indirect prompt injection via documents; retrieval data exfiltration; tool hijacking

3.4.8 MITRE-BASED THREAT MAPPING FOR AI-ENABLED METRO ENVIRONMENTS

This subsection addresses the threat-driven view by mapping plausible adversarial behaviours to MITRE patterns. The purpose is not to predict a single threat actor, but to ensure systematic coverage of 'how' AI systems can be compromised or abused, and what that means for metro operations.

To ensure this work is grounded in an established threat knowledge base for AI systems, the analysis explicitly leverages MITRE ATLAS™ (Adversarial Threat Landscape for AI Systems) as the reference catalogue of AI/ML adversarial tactics and techniques. ATLAS is used as the primary source for AI-specific threat patterns (e.g., training-data manipulation, model artefact tampering, model extraction, adversarial inputs, and misuse of human-facing AI interfaces). Where AI systems interface with

operational technology or safety-relevant functions, MITRE ATT&CK® (including the ICS Matrix) is used in parallel to capture the broader IT/OT adversary behaviours that enable or amplify those AI-specific attacks (initial access, lateral movement, persistence, and impact)

3.4.8.1 MITRE ATLAS OVERVIEW AND HOW IT IS USED

MITRE ATLAS™ (Adversarial Threat Landscape for AI Systems) is a curated knowledge base that documents adversarial behaviours targeting AI/ML-enabled systems. It is conceptually aligned with the MITRE ATT&CK® approach: adversary actions are described as repeatable tactics and techniques, enabling systematic coverage of 'how' attacks occur, independent of a specific vendor or model type. ATLAS is particularly useful for this activity because it addresses attack classes that are not well captured by traditional IT/OT threat models, such as training-data manipulation, model backdooring, model extraction/theft, adversarial inputs at inference time, and attacks that exploit human-facing AI interfaces (e.g., prompt injection for LLM-like assistants).

Within this section, ATLAS is used in three practical ways: (a) as a checklist to ensure each AI lifecycle stage (data → training → deployment → inference) is covered by relevant threat patterns; (b) as a mapping layer to link AI-specific threats to the baseline improvement areas already identified (logging/monitoring, IAM, secure configuration, vulnerability management, incident response); and (c) as an input to validation planning, defining 'must-test' adversarial scenarios for pilots (e.g., tampering with datasets, replacing model artefacts, abusing inference APIs, and manipulating assistant behaviour through crafted inputs).

A practical deliverable is a threat mapping matrix that aligns AI pipeline stages to likely tactics/techniques, and links them to defensive measures and validation activities (including red teaming where appropriate).

Table 69: Threat mapping matrix

AI PIPELINE STAGE	MITRE ATLAS / AI THREAT PATTERN FOCUS	METRO-SPECIFIC IMPACT CONSIDERATIONS	SUGGESTED VALIDATION / ASSURANCE ACTIVITY
Data collection & ingestion	Poisoning, data manipulation, supply chain of data sources	Incorrect planning outputs; unsafe operational recommendations; integrity loss in telemetry-derived analytics	Data integrity controls, provenance checks, anomaly detection on input distributions; source authentication
Training / model development	Training-time compromise, backdoors, insecure dependencies	Hidden failure modes; safety-related errors; long-lived compromise via model artefacts	Secure build pipeline, isolated training envs, dependency scanning, model evaluation for backdoors where feasible
Model packaging & deployment	Artefact tampering, registry compromise, insecure promotion	Malicious model in production; inability to roll back safely; integrity loss of decision support	Model signing, approval gates, SBOM for ML artefacts, reproducible builds; staged rollout with rollback
Inference / operations	Adversarial examples, model extraction, API abuse	Service disruption, degraded decisions, leakage of sensitive operational insights Inference Denial-of-Service (DoS): This involves the malicious overloading of AI inference APIs with massive request volumes to induce latency or service	Rate limiting, authN/authZ, monitoring for abuse, output validation/guardrails; adversarial robustness testing

AI PIPELINE STAGE	MITRE ATLAS / AI THREAT PATTERN FOCUS	METRO-SPECIFIC IMPACT CONSIDERATIONS	SUGGESTED VALIDATION / ASSURANCE ACTIVITY
		interruption. In a metro context, targeting systems like crowding management could lead to operational paralysis and significant passenger safety risks due to the loss of real-time guidance and situational awareness	
Human-in-the-loop (AI assistants)	Prompt injection, data leakage via retrieval, tool misuse	Exposure of sensitive procedures/security information; unsafe actions if assistants can trigger workflows	Prompt injection testing, least-privilege tool access, content filtering, logging and review, safe-by-design prompts

3.4.9 SAFEGUARDS AND MEASURES

The measures below are expressed as layered safeguards that can be adopted incrementally. They are designed to integrate with the improvement areas already identified in the baseline assessment (e.g., logging/monitoring, vulnerability management, incident response), while extending them to AI-specific assets and workflows.

Table 70: Safeguards and measures

DATA GOVERNANCE & HANDLING	DATA CLASSIFICATION FOR AI DATASETS; PROVENANCE AND INTEGRITY CONTROLS; MINIMISATION; RETENTION; SECURE DISPOSAL; CONTROLLED SHARING WITH PARTNERS	DEFINE DATASET OWNERS AND APPROVAL WORKFLOWS; MAINTAIN DATASET INVENTORY AND LINEAGE; ENFORCE ENCRYPTION AND ACCESS LOGGING
MLOps security (build & deployment)	Isolated training environments; secrets management; model registry with signing/versioning; promotion gates; SBOM for code and ML artefacts	Treat models as production software artefacts; require peer review and change control; maintain rollback and reproducibility evidence
Runtime protection & monitoring	Strong API security; rate limiting; anomaly detection; monitoring for drift and abuse; security logging for inference calls	Integrate with SOC monitoring; define alert thresholds; ensure logs include model version, input metadata, caller identity
Human-in-the-loop safeguards	Least-privilege access for assistants; retrieval scope control; prompt injection resilience; redaction of sensitive data; user training	Document acceptable use; require confirmations for risky actions; maintain conversation audit trails where allowed
Testing & assurance	Adversarial testing where feasible; tabletop exercises including AI failure modes; supplier security requirements for AI components	Include AI scenarios in penetration tests and incident response drills; require evidence from vendors and integrators
Resilience & safety alignment	Fail-safe modes; output sanity checks; separation between advisory vs control actions; rigorous change management for OT/ICS interfaces	Clearly define whether AI is advisory or can influence control; enforce segregation of duties and safety constraints

3.4.10 REINFORCEMENT OF BASELINE RECOMMENDATIONS FOR AN AI-ENABLED FUTURE

This activity explicitly builds on the baseline improvement areas and proposed actions identified in the completed cybersecurity assessment. The intent is not to re-baseline the organisation, but to clarify which of the already proposed measures become gating prerequisites once AI systems and data

pipelines are introduced, and how those measures must be extended to cover AI-specific assets (datasets, feature stores, model artefacts, training environments, orchestration components and inference endpoints).

From an auditor perspective, AI adoption should be treated as an expansion of the organisation's crown jewels and the operational decision chain. Even for advisory AI, compromised integrity can lead to systematically wrong planning outputs, degraded service, or unsafe operational recommendations. Therefore, the baseline improvements in monitoring, audit logging, identity/access management, secure configuration and vulnerability management become more time-critical and require tighter evidence and assurance. The subsections below translate baseline recommendations into concrete reinforcement actions to support safe validation and progressive exploitation of AI-enabled applications.

3.4.10.1 NETWORK MONITORING, DEFENCE AND VISIBILITY UPLIFT (BASELINE PRIORITY)

The baseline assessment prioritised Network Monitoring and Defence and Audit Log Management as top exposure areas. In an AI-enabled scenario, network visibility must extend to the full AI pipeline, including: data ingestion paths, training environment connectivity, model registry access, CI/CD interactions, and inference API traffic. This requires defining explicit network zones for AI components (training, staging, production, OT-facing integrations), restricting east-west traffic by default, and ensuring that monitoring covers both north-south and east-west flows. Where AI systems consume OT-derived telemetry, the OT/IT boundary becomes a high-risk trust boundary and should rely on brokered data access.

Concrete reinforcement actions include establishing NDR coverage at key aggregation points, enabling netflow/traffic telemetry for AI zones, adding detections for data exfiltration attempts from training environments, and monitoring for abuse patterns against inference APIs (credential stuffing, high-rate queries, anomalous geographies). Evidence expectations include up-to-date network diagrams with zones and allowed flows, documented firewall rulesets, and SOC use-cases tied to AI lifecycle risks.

3.4.10.2 AUDIT LOGGING AND SIEM USE-CASES EXTENDED TO AI TELEMETRY

The baseline recommendation to centralise and actively monitor audit logs becomes a prerequisite for AI validation. AI systems introduce new decision and integrity events that must be logged and correlated: dataset version changes, model training runs, model promotion approvals, model version deployed to production, inference requests (caller, model version, input metadata where permitted), and administrative actions in orchestration platforms.

Reinforcement actions include defining a minimum set of AI log sources (data pipeline/orchestrator, model registry, runtime gateway, IAM/PAM, container platform/cloud audit logs), ensuring time synchronisation and retention consistent with incident response needs, and engineering SOC detections for high-risk events such as unsigned model deployments, sudden model version changes, repeated inference errors, anomalous access to training datasets, or bulk export of model artefacts. Auditor evidence should include a log source inventory mapped to AI components, retention settings, and examples of correlation rules / alerts tested during pilots.

3.4.10.3 IDENTITY, ACCESS CONTROL AND PRIVILEGED ACCESS MANAGEMENT FOR AI PIPELINES

Baseline IAM improvements (IDM, MFA for privileged and remote access, PAM, periodic access reviews) must be extended to cover AI roles and service identities. AI pipelines create additional privileged personas (data engineers, data scientists, MLOps engineers) and machine identities (ETL jobs, training runners, deployment automation, inference services). Without strong identity governance, these identities become durable attack paths.

Reinforcement actions include enforcing MFA for all privileged interactive access to AI tooling; onboarding AI roles into a formal RBAC model; applying least privilege to datasets, feature stores, registries and orchestration; adopting just-in-time privileged elevation for sensitive operations (model promotion, secrets access, production deployment); and rotating credentials/keys for service identities. Auditor evidence includes RBAC matrices, privileged role inventories, access review records, and PAM/MFA coverage metrics for AI environments.

3.4.10.4 SECURE CONFIGURATION, VULNERABILITY AND SUPPLY-CHAIN MANAGEMENT FOR AI STACKS

The baseline prioritisation of secure configuration and continuous vulnerability management becomes more complex with AI. AI stacks often involve container images, notebooks, orchestration platforms, third-party libraries, GPU drivers, and managed cloud services. This expands supply-chain exposure and the likelihood of misconfiguration.

Reinforcement actions include establishing hardened baselines for AI hosts and container platforms; restricting notebook execution environments; scanning images and dependencies (including ML libraries) with an SBOM approach; integrating vulnerability findings with ticketing and ownership; and defining patch windows for both IT and AI platform components. For third-party models or datasets, supplier controls should demand provenance, integrity guarantees, and update transparency. Evidence includes baseline build standards, scanning reports, remediation SLAs, and SBOM artefacts for pilots.

In addition to the standard Software Bill of Materials (SBOM), the organization should mandate an AI Bill of Materials (AIBOM) from all vendors. The AIBOM must provide comprehensive documentation regarding data provenance, model versions, and dependencies on open-source libraries. This level of transparency is essential for continuous monitoring and the timely identification of vulnerabilities across the AI supply chain

3.4.10.5 INCIDENT RESPONSE UPLIFT TO INCLUDE AI-SPECIFIC INCIDENT CLASSES

Incident response management is foundational and becomes more critical with AI adoption. AI introduces incident classes requiring specific triage and containment steps: suspected data poisoning, model artefact tampering, model extraction attempts, abuse of inference endpoints, and manipulation of human-facing assistants. Without playbooks, pilots may overreact (unnecessary rollback) or underreact (continued use of compromised outputs).

Reinforcement actions include extending IR playbooks to cover AI artefacts (preservation of datasets/models/logs), defining rollback criteria (switch to last known good model, revert to manual decision processes), ensuring SOC scoping through AI telemetry in the SIEM, and running tabletop exercises featuring AI scenarios. Evidence includes updated playbooks, exercise reports, and defined RACI for AI incident handling.

3.4.10.6 SECURITY AWARENESS AND SPECIALISED TRAINING FOR AI-RELATED RISKS

Where the baseline highlighted security awareness as an improvement priority, AI adoption raises the bar. Staff interacting with AI decision support, dashboards or assistants must understand how outputs can be manipulated and what constitutes suspicious behaviour (sudden changes in recommendations, inconsistent outputs, unexpected assistant instructions, anomalous data inputs).

Reinforcement actions include role-based training for operational users and supervisors (trustworthy use of outputs and escalation triggers), engineering teams (secure MLOps and data handling), and leadership (risk ownership and go/no-go criteria). Evidence should include training materials, attendance records, and periodic awareness exercises relevant to AI use cases.

3.4.10.7 CLASSIFICATION OF AI USE: ADVISORY / ACTION-INFLUENCING

For governance and risk management, AI-enabled capabilities should be classified upfront according to how outputs are used. This classification determines the minimum assurance depth, the strictness of gating criteria, and the operational controls required. The intent is to avoid treating all AI use cases as equal: a planning dashboard has different risk dynamics than an AI component that can influence near-real-time operational actions.

Table 71: Classification of AI use (Advisory / Action influencing)

CLASSIFICATION	TYPICAL EXAMPLES IN METRO CONTEXT	PRIMARY RISK	MINIMUM SAFEGUARDS (SUMMARY)
Advisory / decision-support (human-in-the-loop)	Planning optimisation, demand forecasting, maintenance recommendations, operator dashboards	Integrity and drift leading to systematically wrong decisions; data leakage	Strong data governance; auditability of inputs/outputs; monitoring for drift/anomalies; clear user guidance and escalation triggers; rollback to manual process
Action-influencing / automation-adjacent	Near-real-time dispatch support with automated suggestions, automated routing decisions, closed-loop optimisation that changes parameters	Operational disruption or unsafe outcomes if outputs are manipulated; amplified impact of errors	All advisory safeguards plus stricter gating, fail-safe design, separation of duties, stronger monitoring/alerting, staged rollout and rigorous change control

In practice, advisory use cases can progress with proportionate controls and strong auditability. Any action-influencing use case should be treated with higher criticality and should not proceed without explicit sign-off by operations/safety stakeholders ensuring the evidence and validation of fail-safe and rollback mechanisms/

Also critical is considered the Human-in-the-Loop (HITL) and Accountability area: To mitigate the risks of automated decision-making in safety-critical metro environments, a Human-in-the-Loop (HITL)

approach is adopted as a core safeguard. AI-driven recommendations must be presented to operators in an interpretable format (Explainable AI), ensuring they can validate the output against real-world conditions. This human oversight serves as a final security layer, allowing operators to override AI suggestions if a cyber-threat, model bias, or operational anomaly is suspected, thereby maintaining clear lines of accountability.

3.4.10.8 GATING CRITERIA

To avoid 'pilot-to-production' security gaps, the programme should define explicit gating criteria that must be met before (a) initiating an AI pilot with operational data and (b) using AI outputs in production-like operational decision processes. The criteria below are intentionally practical and evidence-driven; they translate the baseline recommendations and the AI-specific measures into concrete go/no-go checks that can be verified by engineering leads, security, and (where relevant) safety/operations stakeholders.

- AI asset inventory established: Datasets, feature stores, model artefacts, registries, training environments and inference endpoints are recorded with owners, sensitivity classification and retention rules; lineage is documented for datasets used in pilots.
- Identity and privilege governance in place: MFA enforced for privileged access; RBAC defined for AI roles; service identities are scoped and rotated; PAM/JIT elevation is used for sensitive actions (model promotion, secrets access, production deployment).
- Secure MLOps pipeline with approval gates: Model registry enforces versioning and signing; promotion to pilot/production requires approvals; artefact integrity is verifiable; reproducible build evidence exists; rollback to a known-good model is documented and tested.
- Baseline security monitoring extended to AI telemetry: Central logging includes AI pipeline and runtime events (dataset changes, model deployment, inference access); key detections are implemented (unsigned model deploy, anomalous dataset access, inference API abuse) and tested end-to-end with the SOC.
- Network zoning and data-path controls validated: AI components are placed in defined network zones; OT/IT boundaries for telemetry are brokered and monitored; firewall rules are reviewed; egress controls from training environments are enforced to reduce exfiltration risk.
- Vulnerability and supply-chain controls operational: Container images and dependencies are scanned; SBOM artefacts are available for pilot components; remediation SLAs are defined; notebook and orchestration platforms are hardened; third-party models/datasets have provenance and integrity assurances.
- Data governance and privacy constraints satisfied: Training and operational datasets meet minimisation requirements; sensitive fields are protected or redacted; sharing with vendors/research partners is controlled and logged; retention and disposal processes are defined for pilot artefacts.

Table 72: Gating criteria

GATE	OWNER (EXAMPLE)	EVIDENCE EXAMPLES	STATUS / SIGN-OFF
AI asset inventory & lineage	Data Governance / Platform	Inventory export; dataset lineage; classification	
IAM/PAM coverage for AI	IAM / Security	RBAC matrix; MFA/PAM reports; access reviews	
Signed models & promotion gates	MLOps / Platform	Registry logs; approvals; signing proof; rollback test	
SOC monitoring for AI telemetry	SOC / Security	SIEM use-cases; alert tests; log source list	
Network zoning & OT/IT boundary controls	Network / OT Security	Diagrams; firewall rules; NDR coverage evidence	
Supply chain & vuln management	SecOps / Platform	Scan reports; SBOM; remediation tickets; AIBOM;	
AI IR readiness	Security / Operations	Playbooks; tabletop report; RACI	
Pilot validation tests	Project / Security	Test plan; results; remediation evidence	
Operational safe-use constraints	OCC / Safety / Ops	Procedures; user guidance; sign-off	

To ensure full compliance with GDPR and protect passenger privacy, the adoption of Privacy-Enhancing Technologies (PETs) is mandatory. Techniques such as Differential Privacy or on-the-fly anonymization must be applied to telemetry data and CCTV feeds before they are ingested into AI training or inference pipelines. This ensures that the utility of the data for AI optimisation is balanced with the stringent protection of personal identifiable information (PII)

- AI incident response readiness confirmed: AI-inclusive IR playbooks exist (poisoning, model tampering, extraction, prompt injection/assistant abuse); evidence collection steps are defined; rollback criteria are agreed with operations; a tabletop exercise has been performed.
- Validation and assurance tests executed: A defined test plan is run prior to pilot go-live, including abuse testing of inference APIs, attempts to tamper with model artefacts, data integrity checks, and (where feasible) adversarial input testing; results and remediation actions are documented.
- Operational safe-use constraints agreed (human-in-the-loop): For pilots, AI is clearly classified as advisory vs action-influencing; manual override and fall-back procedures exist; users receive guidance on interpreting outputs and escalating anomalies; acceptance criteria are signed off by relevant stakeholders.

For governance and audit tracking, the above criteria can be maintained as a simple checklist with assigned owners, evidence references and sign-off dates.

In relation to the priority areas derived from the assessment, the following should be addressed:

Table 73: Focus areas

BASELINE PRIORITY AREA (FROM ASSESSMENT)	AI-ENABLED REINFORCEMENT FOCUS	MINIMUM EVIDENCE / ACCEPTANCE CRITERIA (EXAMPLES)
Network Monitoring & Defence; Audit Log Management	Extend visibility to AI pipelines, registries and inference APIs; monitor OT/IT boundary where AI consumes telemetry	Updated zoning diagram; NDR/SIEM coverage for AI zones; tested alerts for API abuse and data exfiltration
Access Control / Account Management	RBAC for AI roles and service identities; MFA/PAM coverage; JIT elevation for sensitive actions	RBAC matrix; access review records; MFA/PAM metrics; key/secret rotation evidence
Secure Configuration & Vulnerability Management	Hardening baselines; container/dependency scanning; SBOM for ML artefacts; remediation SLAs	Baseline standards; scan reports; remediation tickets; SBOM artefacts for pilots; change control records
Incident Response Management	AI-specific playbooks (poisoning, model tamper, extraction, assistant abuse); rollback criteria	Playbooks; tabletop exercise results; evidence retention plan for AI artefacts/logs
Security Awareness & Skills Training	Role-based training for AI users, engineers and leadership; escalation triggers	Training plan; attendance logs; periodic exercises; documented escalation guidance

3.4.11 AI AWARE COMPENSATING MEASURES

The table below provides a concise ‘delta’ view: what changes when AI is introduced and what must be strengthened to maintain or improve overall risk posture. It is intended to guide planning, procurement and architecture decisions.

Interpretation guidance (auditor view). The ‘delta’ table should be read as a prioritisation aid: it identifies which baseline improvement areas become gating prerequisites for safe AI validation and which minimum safeguards must be added specifically for AI assets. In practice, AI adoption increases dependency on data integrity, identity controls, and monitoring. Even where baseline actions are already planned (e.g., central logging, stronger access control, vulnerability management), AI introduces additional assets that must be placed under the same governance and monitoring regime: datasets, feature stores, model artefacts, training environments, orchestration tools and inference endpoints.

A key distinction is between advisory AI (decision support) and AI that can influence operational actions. For advisory use cases, integrity and auditability of data/model outputs are paramount. For scenarios that could affect operational decisions in near-real time, resilience and safe failure behaviour become additional requirements: the system must remain secure and predictable under attack or malfunction, and must support rapid rollback to a known-good state.

This comparison does not replace detailed solution design. It defines the minimum compensating measures that should be demanded in architecture reviews and procurement specifications before moving from experimentation to pilot operation.

Table 74: Compensating measures per security area

SECURITY AREA	CURRENT BASELINE (REFERENCE TO COMPLETED ASSESSMENT)	FUTURE AI-ENABLED DELTAS	MINIMUM REQUIRED COMPENSATING MEASURES
Asset inventory & classification	Baseline covers conventional IT/OT assets	New asset class: datasets, models, feature stores, prompts, model endpoints	Extend inventory to AI assets; maintain lineage; classify AI artefacts and data
Logging & monitoring	Improvement areas already identified (e.g., central logging, monitoring)	Need to monitor AI-specific events (model version changes, inference abuse, drift)	Security telemetry for AI pipeline and inference; SOC use-cases mapped to AI threats
Vulnerability management	Traditional OS/app vulnerability scanning priorities already identified	ML dependencies, notebooks, pipelines, container images add new supply chain risk	Dependency/SBOM management for ML stacks; hardening of notebooks and orchestrators
Access control	Conventional IAM and privileged access needs remain	Additional privileged roles (data scientists, MLOps) and service identities	Least privilege for registries, data stores, orchestrators; strong separation of environments
Incident response	Baseline improvement actions include IR capability uplift	New incident classes: model compromise, data poisoning, prompt injection, drift-based failure	AI-inclusive playbooks; evidence collection for ML artefacts; rollback procedures
Third-party / supplier	Existing supplier controls applicable	Increased reliance on AI vendors, datasets, cloud AI services	Explicit AI security requirements (signing, provenance, audit logs, red teaming evidence)

3.4.12 RECOMMENDED WORK PRODUCTS AND DELIVERABLES

The activity will result in a set of concrete outputs that enable validation and exploitation of AI models while managing cybersecurity and data-handling risks:

Cross-reference: the outputs in this subsection are intended to be attached/linked per AI demonstrator to the “Results of Validation” and “Operational Impact & Exploitation” subsections of D7.1, so that cybersecurity readiness becomes part of the overall validation evidence package.

Deliverable elaboration. Each work product is intended to be usable by both engineering teams and governance/audit stakeholders. The emphasis is on evidence-based outputs (diagrams, matrices, and acceptance criteria) rather than generic statements.

- AI-enabled architecture and data-flow views: These diagrams should explicitly mark trust boundaries (IT zone, OT zone, vendor/cloud zone), data sensitivity classes, and control points (authentication, encryption, logging, approval gates). They serve as the reference for threat modelling and future design reviews.
- Threat mapping matrix (AI lifecycle × MITRE patterns): The matrix lists the most relevant ATLAS techniques per lifecycle stage and maps them to likely impacts in metro operations (availability, integrity of operational decisions, confidentiality of sensitive operational data). It should also capture detection opportunities (log sources/telemetry) and indicate which threats are ‘must-test’ items during pilots.

- Catalogue of measures: Distinguish between (a) extensions of baseline improvement actions already identified (e.g., expanding SOC/SIEM use-cases to include AI events) and (b) new AI-specific requirements (e.g., model signing, dataset provenance, model registry approval gates). For each measure, define minimum expectations, ownership, and evidence to be produced.
- Validation and assurance plan: Define acceptance criteria for pilots, including security tests (API abuse, credential misuse, artefact tampering), data integrity checks, and operational safety checks (output sanity checks, rollback procedures). State what evidence is collected (logs, configuration exports, test reports, approvals) and what constitutes pass/fail.
- Roadmap aligned to baseline priorities: Sequence actions so enabling foundations (identity, logging, vulnerability management, incident response readiness) are implemented before higher-risk AI integrations. Highlight 'gating items' required before moving from experimentation to production-like environments (e.g., SOC monitoring coverage for AI, IR playbooks including AI incidents, registry signing and approval workflows).
- AI-enabled architecture and data-flow views (high-level) highlighting trust boundaries and security control points.
- Threat mapping matrix (AI lifecycle × MITRE ATLAS/ATT&CK patterns) with prioritised risk themes.
- A catalogue of security measures (governance/process/technical) tailored to the identified AI scenario archetypes.
- Validation and assurance plan: recommended tests, evidence expectations, and acceptance criteria for pilot deployments.
- A roadmap of short-, medium-, and longer-term actions, aligned with the baseline improvement priorities already agreed.

3.4.13 LIMITATIONS AND ASSUMPTIONS

This section is based on a model-agnostic description of AI/ML pipelines and typical metro operational integrations. As concrete use cases, vendors, and future train control characteristics are specified, the threat mapping and measures should be refined to reflect the actual architecture, data sensitivity, and safety constraints.

The increasing use of AI both in the technologies used but also in the tactics and tools utilised by the attackers is calling for specific strengthening measures in most areas of the defence strategies as explained in this section. The adoption of the measures described should be aligned with the IT landscape characteristics of the Metro operators and should be considered as a high priority project. Furthermore, all IT security tools used should be of the latest generation and continuously updated in order to be aware of the AI threats.

3.5 SUMMARY OF OUTCOMES

Chapter 3 documents the **validation and exploitation of AI applications developed in WP6** and is positioned within **Task 7.2**. It is structured around three core pillars: the validation of selected AI use cases, exemplary energy and CO_2 optimisation analysis, and the development of AI-supported cybersecurity strategies. The chapter highlights the interaction of three workstreams, with a particular focus on **energy and CO_2 optimisation in conjunction with future Train Control Systems (TCS)**

and the Total Cost of Ownership (TCO) tool, the structured validation of AI models, and the integration of dedicated cybersecurity measures. The objective is not only to demonstrate technical feasibility, but to assess **transferability, robustness, operational relevance, and readiness for real-world exploitation** in metro environments.

Three use cases were validated: (1) passenger demand forecasting, (2) timetable creation using GTFS feeds, and (3) vehicle uncleanliness detection. Together, they represent different operational layers of metro systems, ranging from strategic planning to service quality monitoring.

The passenger demand forecasting demonstrator validates a **log-linear econometric metro demand model** originally developed for West Midlands Metro using operational data from **Metro Sofia**. The Sofia network provided a robust validation environment due to significant structural developments, including network extensions, the introduction of a new metro line, airport and business park expansions, COVID-related ridership disruption and recovery, and the availability of longitudinal monthly ridership data from 2011 to 2024. These structural changes created distinct demand regimes, enabling the model to be tested under conditions of infrastructure expansion, external shocks, and changing policy or funding contexts. The key validation question was whether the model could be transferred across metropolitan contexts while maintaining acceptable predictive performance and stable structural demand elasticities. The analysis therefore focused on **transferability and structural robustness rather than mere statistical fit**. The exploitation potential lies in **long-term strategic planning, scenario-based forecasting of network extensions, and infrastructure investment analysis**, supporting evidence-based decision-making in future metro developments. The second use case evaluates the feasibility and methodological validity of **timetable creation using GTFS feeds combined with algorithmic optimisation approaches**. No physical timetable was implemented within WP7; instead, the study assessed the methodological soundness of GTFS-based timetable generation and the applicability of optimisation frameworks for metro operators. The chapter synthesizes literature and case studies demonstrating real-time rescheduling under disruption conditions and timetable optimisation at interchanges with time-varying demand.The findings indicate that advanced optimisation approaches can significantly improve disruption recovery, enhance modal integration, reduce passenger waiting times, and optimise capacity and fleet utilisation. In addition, GTFS validation tools enable structured cross-checking of timetable data prior to publication, strengthening data reliability. Overall, the results confirm that **algorithmic timetable creation supported by validated GTFS feeds is both practical and advantageous for modern metro systems**, with clear pathways toward performance analytics, energy optimisation, and improved operational resilience.

The third use case addresses **AI-based classification and object detection for uncleanliness detection** using real images from Metro Genova. The validation was conducted as a qualitative proof-of-concept using a dataset of eight images processed through the unmodified training pipeline. Although the limited dataset does not allow statistically robust evaluation through conventional metrics such as precision or recall, the model correctly classified seven out of eight images. These results, while not statistically representative, confirm **functional alignment with operational conditions, model architecture viability, and readiness for expanded validation**.

From an exploitation perspective, the system has the potential to support automated cleanliness monitoring, targeted dispatch of maintenance personnel, and service quality assurance. However,

further development requires larger datasets, deeper operational integration, and comprehensive performance evaluation under real operating conditions.

Chapter 3 also presents exemplary results linking AI applications to **energy and CO_2 optimisation**, particularly in combination with the TCO tool and future TCS interactions. This analysis establishes a connection between AI-enabled optimisation and operational cost reduction, energy efficiency improvements, sustainability targets, and future automation scenarios. Although still preliminary, these findings demonstrate a **direct link between AI exploitation and measurable environmental impact**, reinforcing the strategic relevance of AI integration in metro systems.

A substantial part of the chapter is dedicated to **cybersecurity in AI-enabled metro systems**. The cybersecurity analysis follows a threat-driven, model-agnostic approach covering the entire AI lifecycle, from data ingestion and model training to deployment and inference. The methodology is structured into four steps: (1) Definition of AI-enabled target scenarios, (2) identification of trust boundaries and expanded attack surfaces, (3) threat mapping based on MITRE ATT&CK® and MITRE ATLAS™, (4) derivation of layered safeguards and mitigation strategies. By integrating MITRE ATT&CK® for IT/OT environments and MITRE ATLAS™ for AI-specific threats such as data poisoning, model theft, adversarial inputs, and prompt injection, the analysis provides a structured comparison between the current security baseline and a future AI-enabled exposure profile.

To remain vendor-agnostic, AI adoption is structured into defined scenario archetypes covering demand and service planning analytics, asset and maintenance optimisation, operational decision support, and AI assistants for operations knowledge. For each archetype, primary data inputs, operational touchpoints, and security implications — such as data integrity, telemetry protection, and OT/IT boundary security — are systematically assessed. The chapter further defines **concrete gating criteria for safe AI deployment**, including integration of AI telemetry into central logging systems, clear network zoning and OT/IT boundary controls, supply-chain transparency through scanning and SBOM artefacts, GDPR compliance supported by Privacy-Enhancing Technologies such as differential privacy, AI-specific incident response playbooks, and human-in-the-loop operational constraints. These measures ensure that **AI exploitation is embedded within governance, safety, and operational resilience frameworks rather than implemented as an isolated technical layer**. Overall, Chapter 3 demonstrates that **AI applications can deliver tangible strategic and operational value for metro systems**, including improved planning capabilities, advanced timetable optimisation, enhanced service quality, energy and CO_2 reduction, and greater resilience under disruption. At the same time, it clearly shows that successful exploitation requires structured validation, clearly defined acceptance criteria, rigorous cybersecurity hardening, and comprehensive governance and lifecycle controls. The chapter therefore marks the transition from experimental AI development in WP6 toward **operationally validated, secure, and strategically exploitable metro AI solutions**, integrating performance, sustainability, and security within a coherent validation framework.

4 CONCLUSIONS

Deliverable D7.1 achieved a key objective of the NEXUS project, namely to employ scenario-based simulation methods to validate and exploit all the tools, concepts and applications developed in the preceding WPs, and hence attending to requirements of WS1, WS2 and WS3. A total of 23 simulation modelling scenarios (organised into 5 themes) were carried out to ensure that the metro system and station models, concepts and applications are accurate, consistent, complete, and free from: errors, ambiguities, and contradictions. All the metro system models, concepts and applications were properly validated and cross-validated to ensure accurate reflection of the real-world system and its specific requirements and are useful for decision-making purposes without reservation. Specifically, this deliverable (D7.1) documented the outcomes of Task 7.1 and Task 7.2, including the validation and exploitation of metro simulation tools and AI applications, exemplary assessment of metro energy and CO_2 optimisation, as well as AI-support cybersecurity strategies.

Within Task 7.1, a **multi-perspective** (using a suite of validated metro simulation models), **multi-purpose** (according to 23 scenarios), and **multi-level** (looking at macro-, meso-, and microscopic level impacts) evaluation approach was adopted to identify current system's strengths and weakness, as well as the possible impacts of switching to GoA4 operations. Scenarios were developed in close coordination with corresponding metro operators of the two use cases in Genoa and Sofia to ensuring practical metro operational concerns were incorporated. Impacts were revealed at network-wide, metro line, and metro station crowd management levels. It was found that **GoA4 is most effective as a disruption-mitigation tool when the disruption affects a single node rather than the entire line.** When **c**onsidering switching to GoA4 operations, **a suitable trade-off between reducing passenger wait times and improving train capacity utilisation should be made** when determining the optimal train headways.

The core activity of Task 7.2 was three-folded, including use case validations, a preliminary analysis linking AI applications to **energy and CO_2 optimisation**, particularly in combination with the TCO tool and future TCS interactions, and the AI-support cybersecurity strategy assessment. This analysis established a connection between AI-enabled optimisation and operational cost reduction, energy efficiency improvements, sustainability targets, and future automation scenarios. Although still preliminary, these findings demonstrate a **direct link between AI exploitation and measurable environmental impact**, reinforcing the strategic relevance of AI integration in metro systems.

A cybersecurity baseline assessment was completed by focusing on the future-oriented validation and safe exploitation of AI-enabled systems for metro operations and planning. Using a threat-driven, model-agnostic approach, this task examined how AI/ML lifecycles (data ingestion, training, deployment, inference) could expand the attack surface and introduced new trust boundaries across IT and operational environments. MITRE-based AI threat patterns (MITRE ATLAS, complemented by ATT&CK/ICS where relevant) were used to define "must-test" adversarial scenarios and to compare current vs AI-enabled exposure. The outcome was a set of practical strategies and measures for secure data handling, trustworthy AI operation, and cybersecurity safeguards. Clear gating criteria are provided

to support evidence-based progression from experimentation to pilots and production-like exploitation for a large metro operator.

To summarise, this deliverable (D7.1) documented a baseline assessment of the strengths and weaknesses of the current metro system performances as well as the possible impacts of switching to GoA4 operations, AI-supported applications in various aspects of metro operations and their links to energy and CO_2 optimisation, as well as the associated cybersecurity concerns. Deliverable D7.2 will extend this effort further by exploiting the validated models to simulate various proposed future innovative metro improvement measures envisaged in Task 7.3 and Task 7.4 for optimised train vehicle concepts and service optimisation purposes.

5 REFERENCES

Aemmer, Z., Ranjbari, A. and MacKenzie, D., 2022. “Measurement and classification of transit delays using GTFS-RT data.” *Public Transport*, 14(2), 263-285. https://doi.org/10.1007/s12469-022-00291-7.

Bach, L., Mannino, C. and Sartor, G., 2019. “MILP approaches to practical real-time train scheduling: the Iron Ore Line case.” In *International Network Optimisation Conference*, 78-82. DOI:10.5441/002/inoc.2019.15.

Bhagat, M. and Bakariya, B. 2025. “A comprehensive review of cross-validation techniques in machine learning,” *IJSAT-International Journal on Science and Technology* 16(1). https://doi.org/10.71097/IJSAT.v16.i1.1305.

Bishop, C. M. and Nasrabadi, N. M. 2006. “Pattern recognition and machine learning. Vol.4.” Springer. ISBN: 978-1-4939-3843-8.

Gardner, J. R., Kusner, M. J., Xu, Z. E., Weinberger, K. Q., and Cunningham, J. P. 2014. “Bayesian optimisation with inequality constraints.” *ICML'14: Proceedings of the 31st International Conference on International Conference on Machine Learning* 32: 937-945.

Garrido-Merchán, E. C., & Hernández-Lobato, D. 2020. “Dealing with categorical and integer-valued variables in bayesian optimisation with gaussian processes.” *Neurocomputing*, 380(7): 20-35. https://doi.org/10.1016/j.neucom.2019.11.004.

Geifman Y. and El-Yaniv, R. 2017. “Selective classification for deep neural net-works.” *Advances in neural information processing systems*, 30 (NIPS 2017).

General Transit Feed Specification. 2025. “What is GTFS?” Available at: https://gtfs.org/getting-started/what-is-GTFS/ (Accessed: 26 January 2026).

Guo, H., Bai, Y., Hu, Q., Zhuang, H. and Feng, X., 2021. “Optimisation on metro timetable considering train capacity and passenger demand from intercity railways.” *Smart and Resilient Transportation* 3(1), 66-77. https://doi.org/10.1108/SRT-06-2020-0004.

Gupta, S.D., Van Parys, B.P. and Tobin, J.K., 2023. “Energy-optimal Timetable Design for Sustainable Metro Railway Networks.” arXiv preprint arXiv:2309.05489. https://doi.org/10.48550/arXiv.2309.05489

Holzinger, A. 2016. “Interactive machine learning for health informatics: When do we need the human-in-the-loop?” *Brain informatics* 3(2) 119–131. Doi: 10.1007/s40708-016-0042-6.

Huang, J., Zhang, T. and Wei, R., 2021. “Urban railway transit timetable optimisation based on passenger-and-trains matching – A case study of Beijing metro line.” *Promet - Traffic and Transportation* 33(5): 671-687. DOI: 10.7307/ppt.v33i5.3736.

Jakubik, J., Weber, D., Hemmer, P., Vössing, M. and Satzger, G. 2023. “Improving the efficiency of human-in-the-loop systems: Adding artificial to human experts.” in *International Conference on Wirtschaftsinformatik* 131–147, Springer. https://doi.org/10.1007/978-3-031-80122-8_9.

Jin, W., Sun, P., Yao, B. and Ding, R., 2024. “A Train Timetable Optimisation Method Considering Multi-Strategies for the Tidal Passenger Flow Phenomenon.” *Applied Sciences” 14(24): 11963 https://doi.org/10.3390/app142411963.

Kang, Y., Chiu, Y.-W., Lin, M.-Y., Su, F.-Y. and Huang, S.-T. 2021. “Towards model-informed precision dosing with expert-in-the-loop machine learning.” in *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)* 342–347, doi: 10.1109/IRI513352021.00053.

Kim, J., and Choi, S. 2023. “BayesO: A Bayesian optimisation framework in Python.” *Journal of Open Source Software* 8(90), 5320. https://doi.org/10.21105/joss.05320.

Kreuzberger, D., Kühl, N. and Hirschl, S. 2023. “Machine learning operations (mlops): Overview, definition, and architecture.” *IEEE access* 11 31866–31879. 10.1109/ACCESS.2023.3262138

Li, C., Tang, J., Zhang, J., Zhao, Q., Wang, L., & Li, J. 2023. „Integrated optimisation of urban rail transit line planning, timetabling and rolling stock scheduling.” *PLoS ONE* 18(5), e0285932. https://doi.org/10.1371/journal.pone.0285932.

Lim, A., Sharma, S., Bhaskar, A. and Arkatkar, S., 2019. “An open-source framework for GTFS data analytics: Case study using the Brisbane TransLink network” In *Proceedings of the 41st Australasian Transport Research Forum (ATRF)* Australasian Transport Research Forum (ATRF).

Liu, D., Guo, J., Gu, Y., King, M., Han, L.D. and Brakewood, C. 2024. “GTFS2STN: analyzing GTFS transit data by generating spatiotemporal transit network.” *Information* 16(1): 24. https://doi.org/10.3390/info16010024.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J. and Fernandez-Leal, A. 2023. “Human-in-the-loop machine learning: A state of the art.” *Artificial Intelligence Review* 56: 3005–3054. https://doi.org/10.1007/s10462-022-10246-w.

Mobility Data. 2026. “Canonical GTFS Schedule Validator.” Available at: https://gtfs-validator.mobilitydata.org/

Oh, Y., Kwak, H.C. and Kang, S., 2024. “Development of optimal real-time metro operation strategy minimizing total passenger travel time and train energy consumption.” *IET Intelligent Transport Systems* 18(12): 2440-2458. https://doi.org/10.1049/itr2.12582.

Pandey, R., Purohit, H., Castillo, C. and Shalin, V.L. 2022 “Modeling and mitigating human annotation errors to design efficient stream processing systems with human-in-the-loop machine learning.” *International Journal of Human-Computer Studies* 160: 102772. https://doi.org/10.1016/j.ijhcs.2022.102772.

Su, B., Wang, F., Su, S., Wang, Z. and Tang, T., 2025. “A Two-Step Optimisation Framework for Real-Time Train Rescheduling in an Urban Rail Transit Line.” In *IEEE Transactions on Intelligent Transportation Systems.* 26(6): 9065-9079. doi: 10.1109/TITS.2025.3541275

Wang, Y., Zhang, M., Ma, J. and Zhou, X. 2016. “Survey on Driverless Train Operation for Urban Rail Transit Systems.” *Urban Rail Transit 2* 106–113. https://doi.org/10.1007/s40864-016-0047-8.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T. and He, L. 2022. “A survey of human-in-the-loop for machine learning.” *Future Generation Computer Systems* 135: 364–381.

Wu, Z. and Yang, K., 2023. “Collaborative Optimisation of Train Scheduling and Passenger Flow Control with Adjustable Dwell Time on a Congested Metro Line.” In *International Conference on Electrical and Information Technologies for Rail Transportation* Singapore: Springer Nature Singapore: 540-548.Van Der Knaap, R. J. H., De Bruyn, M., Van Oort, N., Huisman, D., & Goverde, R. M. P. (2024). Clustering railway passenger demand patterns from large-scale origin–destination data. Journal of Rail Transport Planning & Management, 31, 100452. https://doi.org/10.1016/j.jrtpm.2024.100452

Zhang, M., Wang, Y., Su, S., Tang, T. and Ning, B., 2018. “A short turning strategy for train scheduling optimisation in an urban rail transit line: the case of Beijing subway line 4.” *Journal of Advanced Transportation* 2018: 5367295. https://doi.org/10.1155/2018/5367295

Zhou, Y., Bai, Y., Li, J., Mao, B. and Li, T., 2018. “Integrated optimisation on train control and timetable to minimise net energy consumption of metro lines.” *Journal of Advanced Transportation* 2018: 7905820. https://doi.org/10.1155/2018/7905820