Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p26079-cd

From Inspection to Action: Probabilistic LLM-Based Planning for Resource-Aware Railway Maintenance

E. Aldao

Aerolab, Instituto de Física, Computación e Ciencias Aeroespaciais, Universidade de Vigo, 32004, Ourense, Galicia, Spain E-mail: enrique.aldao.pensado@uvigo.gal

E. Ríos-Otero

Aerolab, Instituto de Física, Computación e Ciencias Aeroespaciais, Universidade de Vigo, 32004, Ourense, Galicia, Spain E-mail: eduardo.rios@uvigo.gal

F. Veiga-López

Aerolab, Instituto de Física, Computación e Ciencias Aeroespaciais, Universidade de Vigo, 32004, Ourense, Galicia, Spain E-mail: fernando.veiga@uvigo.gal

H. González-Jorge

Aerolab, Instituto de Física, Computación e Ciencias Aeroespaciais, Universidade de Vigo, 32004, Ourense, Galicia, Spain E-mail: higiniog@uvigo.gal

High-speed rail is becoming the main transport option for medium-distance travel, thanks to its efficiency and convenience. However, the increased frequency of operations causes significant infrastructure wear, raising both the requirements and costs of maintenance. For these reasons, the AIRIS (Autonomous Intelligent Railway Inspection System) project, jointly developed by COPASA and the University of Vigo, aims to automate inspections using onboard optical sensors and artificial intelligence tools for the autonomous characterisation of railway infrastructure. The system includes LiDAR sensors as well as global-shutter cameras for the automatic detection of track defects, such as damaged concrete sleepers, deformations in the ballast bed, and issues with fastener elements. All this defect information is integrated into a Large Language Model (LLM) decision-making framework, which, based on pathology analysis data from the sensors, generates maintenance plans that account for factors such as resource availability, damage severity, and uncertainties in defect detection. The output is a user-friendly tool that provides maintenance operators with an inspection plan and recommended maintenance actions. This work presents the main results of the AIRIS project, focusing on the system's outcomes and uncertainties, as well as a quantitative analysis of the proposed decision-making framework, evaluating the reliability of LLM-based solutions for planning and decision-making in railway maintenance.

Keywords: Railway Maintenance, Artificial Intelligence, Operation Planning, Large Language Model (LLM)

1. Introduction

High-speed railway has become one of the preferred options for medium-distance travel, driven by improvements in infrastructure, increasing demand for sustainable mobility, and competitive door-to-door travel times (Ye et al. (2022); Li et al. (2023)). However, the increase in operating frequency, together with the higher loads imposed by heavy and high-speed vehicles, leads to significant infrastructure wear, thereby increasing maintenance and inspection requirements (Zhan et al. (2022); Consilvio et al. (2019)).

Traditionally, railway inspections have been performed manually by human operators, who visually verify on-site the condition of track elements. This process is time-consuming and resource-intensive, often requiring track closures and thereby considerably increasing maintenance costs. With the growing inspection requirements associated with higher traffic volumes and operating frequencies, these costs are expected to rise further, potentially compromising the economic sustainability of the railway sector (Odolinski and Boysen (2019); Binder et al. (2023)).

To mitigate this issue, autonomous inspection systems have been proposed in recent years to automate these tasks using sensors embedded in track vehicles such as trains or inspection draisines. These solutions typically rely on optical sensors, such as cameras or LiDAR, with computer vision and artificial intelligence algorithms for the automatic detection and characterisation of defects. Recent works have successfully proposed systems for the autonomous inspection of track components, including fasteners, catenary systems, ballast morphology, and the condition of rail and fastening elements (Aldao et al. (2023); Arain et al. (2024); Jeong et al. (2025); Olivier et al. (2025)).

Although these models can reliably quantify infrastructure damage, the information they provide must be integrated into decision-making frameworks, and optimisation methods must be established for planning maintenance operations (Sedghi et al. (2021); Buurman et al. (2023)). Traditionally, task scheduling has been addressed using Mixed-Integer Linear Programming (MILP) models, which mathematically define decision variables (such as whether to perform a repair), objective functions, and operational constraints to optimise resource allocation. This approach has been widely applied in previous studies, successfully optimising maintenance scheduling in both construction and railway contexts (de Weert et al. (2024); van Hövell et al. (2022); Costa et al. (2024)).

However, the rigid mathematical formulation of MILP approaches may not fully capture the uncertainty and adaptability inherent in real-world maintenance operations. By defining hard constraints, these models can sometimes produce infeasible solutions, and formulating a single objective function that balances cost, risk, and service continuity is often challenging. Computational complexity also grows rapidly with problem size, limiting scalability for large infrastructures (He et al. (2023); Figueroa et al. (2024)).

In contrast, large language models (LLMs) have recently emerged as an alternative for task scheduling. LLMs are AI models trained on vast amounts of text and structured data to understand, interpret, and generate human-like reasoning. In the context of maintenance planning, they can process inspection results, network information, and resource availability, and use this knowledge to autonomously generate flexible and adaptive maintenance schedules. By reasoning over multiple data sources and constraints, LLMs provide a scalable approach that can prioritise actions, handle uncertainty, and support decision-making beyond the limits of traditional MILP formulations (Gao et al. (2024); Chen et al. (2025); Maggio (2025); Bai et al. (2026)).

This work demonstrates the use of a large language model (LLM) for autonomous railway maintenance planning. The study is set within the AIRIS project (Adaptive and Intelligent Railway Inspection System), a collaboration between the University of Vigo and COPASA to develop optical-based autonomous inspection and maintenance systems. A regional case study in Galicia, Spain integrates infrastructure defect maps annotated by severity, railway network layouts, and operational constraints such as maintenance windows and vehicle availability.

The remainder of the paper is structured as follows: Section 2 introduces the methodology, presenting the training dataset generation and the noise field reconstruction framework. In Section 3, the results are discussed and analysed. Finally, the main conclusions of this work are included in Section 4.

2. Methodology

Figure 1 depicts the workflow diagram of the AIRIS autonomous inspection and maintenance planning system. First (Section 2.1), infrastructure damage is detected and geolocated using an autonomous defect detection system introduced in previous works (Aldao et al. (2025); Aldao et al. (2025)). Next (Section 2.2), the severity of detected defects is assessed, and finally (Section 2.3), maintenance operations are planned using a state-of-the-art LLM. The rest of the sections explain the different steps of this methodology in detail, providing an overview of the AIRIS workflow.

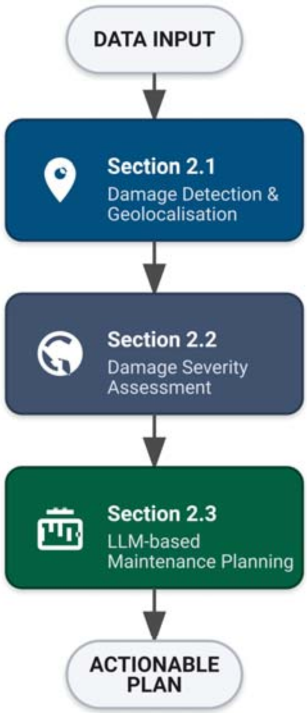

Fig. 1. Autonomous LLM-based inspection and maintenance planning diagram.

2.1. *Damage Detection & Geolocalisation*

For data acquisition, the inspection hardware shown in Fig. 2 is employed. The system is designed to be carried onboard railway vehicles, such as maintenance draisines and trains, enabling automatic data collection. It consists of a LiDAR sensor and global-shutter cameras for measuring ballast morphology and detecting track surface defects, together with an onboard computer for data synchronisation and navigation.

For defect detection (Fig. 3), a multi-modal analysis is conducted by processing data from the onboard sensors. Camera images are analysed in two stages. First, a DeepLab V3+ neural network is applied to characterise the different track elements and verify the presence of fasteners. Subsequently, YOLO-based object detectors are used to detect and estimate the size of smaller defects, such as rust or cracks in the concrete sleepers. Further details on the implementation can be found in (Aldao et al. (2023, 2025)). LiDAR data are processed by reconstructing point clouds and comparing their geometry against a reference railway track model, which enables the detection of voids and deformations in the ballast bed.

Once these pathologies are identified, their geographical position is computed. To this end, a multi-modal navigation solution is adopted, combining tachometer data, visual navigation, and GNSS RTK, as introduced in previous work (Aldao et al. (2025)). This approach provides increased robustness in areas with degraded GNSS reception, such as tunnels or mountainous regions, enabling accurate georeferencing within a Geographic Information System (GIS).

Fig. 2. AIRIS autonomous inspection hardware.

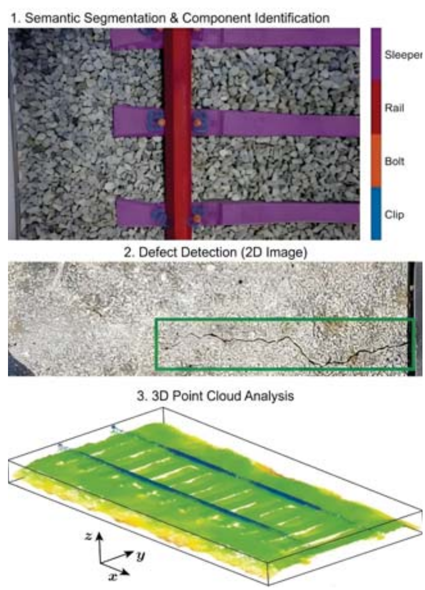

Fig. 3. Multi-modal analysis for autonomous defect detection Aldao et al. (2023, 2025)

2.2. *Damage Severity Assessment*

Once defects are detected, their severity is characterised according to Table 1. Each pathology is assigned a severity value ranging from 5 to 90, representing its expected impact on the infrastructure. In this work, due to the lack of available data, severity values are generated randomly within the specified ranges. In a real-world application, however, this characterisation could be refined using a decision-logic approach based on damage size, location, and anticipated impact, informed by expert judgement.

Table 1. Infrastructure defects and severity ranges

Defect	Severity Range	Repair time
Cracked sleeper	30–60	25 min
Missing bolt	50–80	2 min
Missing fastener	40–70	4 min
Rust	10–40	15 min
Stones over sleeper	5–20	1 min
Ballast void	50–90	3 min
Ballast deformation	40–85	10 min

2.3. *LLM-Based Maintenance Planning*

As a practical case study, the Galicia high-speed railway network is considered, a regional network comprising three main lines: Santiago–A Coruña, Santiago–Vigo, and Santiago–Ourense. For the LLM-based maintenance planning, a total of 251 defects were generated, distributed across the network, with randomised severities and categories as illustrated in Fig. 4.

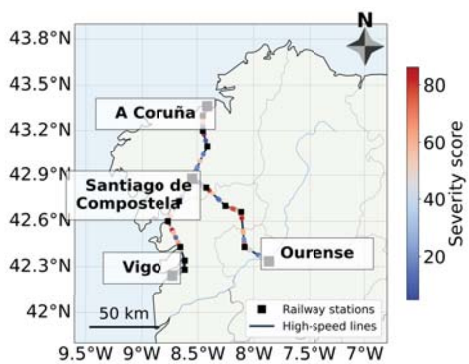

Fig. 4. Case study of 251 defects with random severities in Galicia's high-speed railway network.

To automate the scheduling of these inspection tasks, the problem was formulated as a multi-objective constraint satisfaction problem and solved using the Gemini 3.0 Pro LLM. Defect information was provided to the model in a JSON file, including details such as pathology ID, internal network location (kilometre position and associated line), maintenance duration, and priority score. Additionally, the following constraints were imposed:

(1) **Kinematic Constraints:** The vehicle can only travel along the rail network shown in Fig. 4. Distances between defect points are defined by their kilometre positions along the line. For nodes belonging to different track sections, the vehicle must first pass through Santiago de Compostela station to switch sections. Travel time between nodes is determined by distance and the vehicle's typical speed, assumed to be 80 km/h on average.

(2) **Temporal Windows:** Operations are strictly limited to the maintenance window from 23:30 to 05:30. A soft constraint allows minor exceedances past 05:30 only if justified by proximity to a station or the need to complete high-priority tasks. Total operation time is computed as the sum of maintenance durations and travel times between nodes.

(3) **Return to Logistic Stations:** Upon completing each working journey, vehicles are required to return to a logistic station as indicated in Fig. 4. The network includes four primary stations—A Coruña, Santiago, Ourense, and Vigo—while additional smaller service stations are available for vehicle storage as needed.

Under these assumptions, the model was tasked with generating inspection plans for different levels of maintenance coverage. Three levels—low, medium, and high—were selected to balance repair time and expected defect detection, resulting in a maintenance schedule specifying the operations to be performed. The number of operating days was left for the model to determine, allowing it flexibility to adjust based on the total required repair time. All instructions were provided to the model via the prompt described in Appendix A.

3. Results

Figures 5, 6, and 7 show the spatio-temporal scheduling for the three maintenance configurations. In the high-coverage plan, the operations take up to five days to complete all scheduled repairs. On the first day, work begins at the Santiago station, addressing the repairs closest to Santiago and progressing to an intermediate service station, where operations pause until the following day. The second day completes the Santiago–Vigo track repairs, finishing the day's schedule in Vigo. Subsequently, the Santiago–Ourense track is repaired over two days to complete all planned tasks. On the final day, repairs on the Santiago–Coruña track are carried out, completing all operations in a single journey.

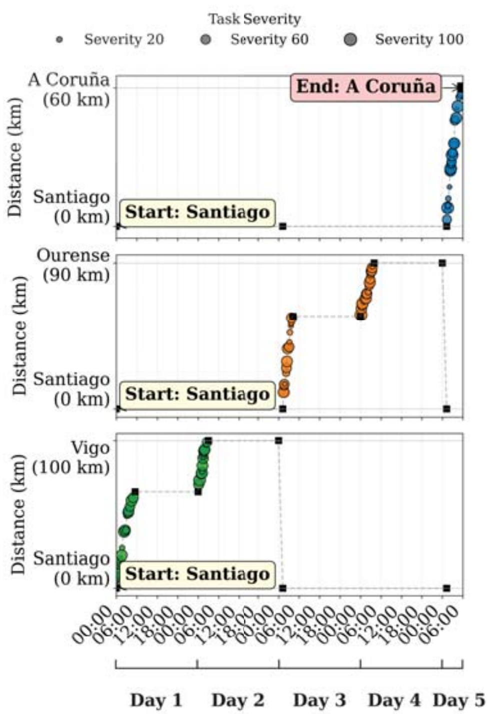

Fig. 5. Spatio-temporal distribution of the high-coverage maintenance plan.

In the intermediate plan, each line is repaired in a single day, without stopping at intermediate stations, only pausing at the main stations—Vigo, Santiago, and Coruña—at the end of each working day. To complete all tasks within three days, the total number of repairs is reduced. In the fast low-coverage configuration, the Coruña and Ourense tracks are repaired in just one day, limiting the number of maintenance actions to allow completion of all tasks in only two days.

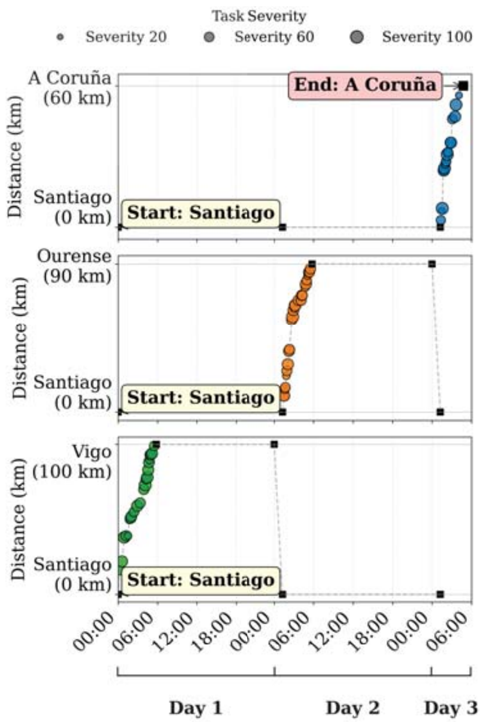

Fig. 6. Spatio-temporal distribution of the medium-coverage maintenance plan.

Figure 8 shows the percentage of performed actions for the three configurations. In the high-coverage plan, almost all maintenance actions (99%) were completed, leaving only 1% unfinished, which would require an additional inspection day. In the medium and low configurations, the completed actions decrease to 67% and 52%, respectively, due to the shorter available schedule. The model prioritises defects with higher severity, focusing primarily on those most critical for infrastructure health. This illustrates a clear trade-off between coverage and completion time.

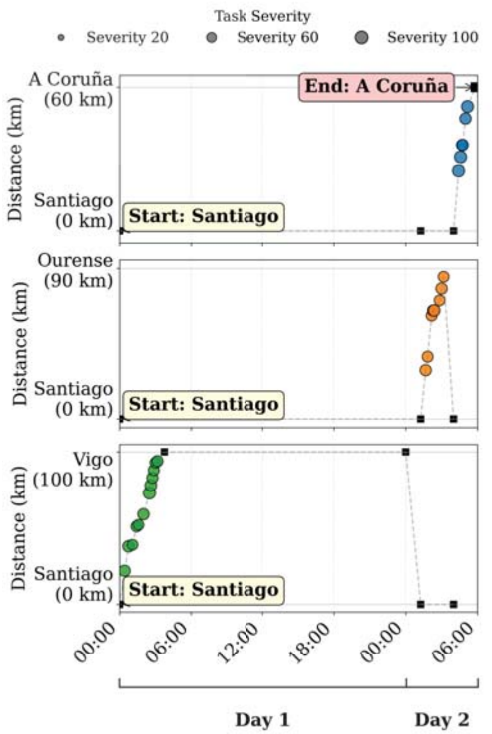

Fig. 7. Spatio-temporal distribution of the low-coverage maintenance plan.

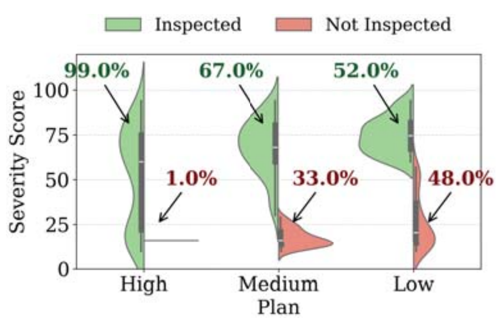

Fig. 8. Percentage of maintenance actions completed under the three maintenance plans.

Figure 9 further highlights this trade-off by showing the number of performed maintenance actions across configurations. For low-priority defects, the low-coverage plan performs no maintenance actions, whereas for critical defects, it closely approaches the number performed in the higher-coverage plans.

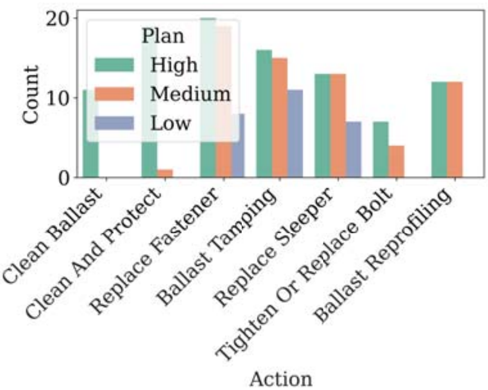

Fig. 9. Distribution of maintenance action categories across the three maintenance plans.

4. Conclusions

In this work, a comprehensive pipeline for autonomous railway maintenance is presented, integrating an autonomous inspection solution for multi-modal data acquisition with a decision-making system based on a Large Language Model (LLM). The approach utilises geolocated defect maps characterised by severity magnitude. This information is processed through a state-of-the-art LLM to generate maintenance plans while accounting for kinematic constraints of vehicle movement, working time windows, and logistic constraints of the railway network. Validation on a simulated case study of the Galicia high-speed network demonstrated the system's ability to generate optimised maintenance plans for varying levels of repair coverage. Results highlighted a trade-off between execution time and thoroughness: a 99% defect resolution rate was achieved in high-coverage scenarios (5 days), while high-severity defects were effectively prioritised in time-restricted, low-coverage configurations (2 days).

Future work will focus on expanding LLM integration to incorporate real-world historical data, moving beyond the synthetic severity values employed in the current study. This development aims to support the transition toward fully predictive maintenance and Structural Health Monitoring (SHM) programs. Additionally, multi-agent LLM architectures will be explored to manage complex, real-time disturbances in the railway network, enabling dynamic rescheduling and more detailed cost–benefit analysis for infrastructure management.

Acknowledgement

This work was funded by the European Union - Horizon Europe, Agencia Estatal de Investigación, Ministerio de Ciencia, Innovación y Universidades, and Xunta de Galicia through the following grants: SARIL A302, STORCITO 101182153, RADIANCE 101235536, PID2021-125060OB-I00 A202, PID2024-156356OB-I00 A507, CPP2022-009679 A308, FPU21/01176, FEADER2024/052B A406, Ciencias Mariñas-MRR C286, GPC-ED431B 2025/31 A505, CO-0091-2024.

References

Aldao, E., L. Fernández-Pardo, L. González-deSantos, and H. González-Jorge (2023). Comparison of deep learning and analytic image processing methods for autonomous inspection of railway bolts and clips. *Construction and Building Materials 384*, 131472.

Aldao, E., L. Fernández-Pardo, F. Veiga-López, L. M. González-deSantos, and H. González-Jorge (2025, 5). Synthetic data generation techniques for enhancing crack detection in railway concrete sleepers. *Journal of Computing in Civil Engineering 39*.

Aldao, E., E. Ríos-Otero, F. Veiga-López, E. Balvís, and H. González-Jorge (2025, 10). Autonomous railway inspection for critical infrastructure security in gnss-denied environments. *Autonomous Systems for Security and Defence II*, 3.

Arain, A., S. Mehran, M. Z. Shaikh, D. Kumar, B. S. Chowdhry, and T. Hussain (2024). Railway track surface faults dataset. *Data in Brief 52*.

Bai, G., X. Luo, Y. Wang, J. Zhou, J. Shen, and Z. Sun (2026). Uav dual-agent: A framework integrating llm and knowledge graphs for low - altitude uav operation and maintenance. In *Lecture Notes in Networks and Systems*, Volume 1698 LNNS.

Binder, M., V. Mezhuyev, and M. Tschandl (2023). Predictive maintenance for railway domain: A systematic literature review. *IEEE Engineering Management Review 51*.

Buurman, B., K. Gkiotsalitis, and E. C. van Berkum (2023). Railway maintenance reservation scheduling considering detouring delays and maintenance demand. *Journal of Rail Transport Planning and Management 25*.

Chen, A., Y. Tian, J. Zhang, C. Li, and H. Zhang (2025). Llm-based intelligent q&a system for railway locomotive maintenance standardization. *Scientific Reports 15*.

Consilvio, A., A. Di Febbraro, R. Meo, and N. Sacco (2019). Risk-based optimal scheduling of maintenance activities in a railway network. *EURO Journal on Transportation and Logistics 8*(5), 435–465.

Costa, M. A., T. P. Alves, A. R. Andrade, and F. Corman (2024). A hybrid maintenance approach for key components of the train bogie to optimize fleet availability. *Engineering Failure Analysis 165*.

de Weert, Y. R., K. Gkiotsalitis, and E. C. van Berkum (2024). Improving the scheduling of railway maintenance projects by minimizing passenger delays subject to event requests of railway operators. *Computers and Operations Research 165*.

Figueroa, A. J., R. Poler, and B. Andres (2024). Adaptive Production Rescheduling System for Managing Unforeseen Disruptions. *Mathematics 12*(22).

Gao, Y., G. Xiong, H. Li, and J. Richards (2024). Exploring bridge maintenance knowledge graph by leveraging grapshsage and text encoding. *Automation in Construction 166*.

He, J., S. Khebbache, M. F. Anjos, and M. Hadji (2023). Optimization of maintenance for complex manufacturing systems using stochastic Remaining Useful Life prognostics. *Computers & Industrial Engineering 182*, 109348.

Jeong, D., D. Kim, K. Lee, and C. Park (2025). A cost-efficient error measurement system for catenary construction in railway applications. *Journal of Electrical Engineering and Technology 20*.

Li, J., E. Elahi, P. Cheng, A. Wu, F. Cao, W. Jian, M. I. Abro, and Z. Khalid (2023). The Opening of High-Speed Railway and Coordinated Development of the Core–Periphery Urban Economy in China. *Sustainability (Switzerland) 15*.

Maggio, L. G. D. (2025). Toward autonomous llm-based ai agents for predictive maintenance: State of the art, challenges, and future perspectives.

Odolinski, K. and H. E. Boysen (2019). Railway line capacity utilisation and its impact on maintenance costs. *Journal of Rail Transport Planning and Management 9*.

Olivier, B., F. Guo, Y. Qian, and D. P. Connolly (2025). A review of computer vision for railways.

Sedghi, M., O. Kauppila, B. Bergquist, E. Vanhatalo, and M. Kulahci (2021, 11). A taxonomy of railway track maintenance planning and scheduling: A review and research trends. *Reliability Engineering & System Safety 215*, 107827.

van Hövell, M. E., R. M. Goverde, N. Bešinović, and M. M. de Weerdt (2022). Increasing the effectiveness of the capacity usage at rolling stock service locations. *Journal of Rail Transport Planning and Management 21*.

Ye, C., Y. Zheng, S. Lin, and Z. Zhao (2022). The Impact of High-Speed Railway Opening on Regional Economic Growth: The Case of the Wuhan–Guangzhou High-Speed Railway Line. *Sustainability (Switzerland) 14*.

Zhan, H.-Y., K.-F. Guo, Y.-J. Xie, W. Shan, and G. Zhao (2022, 1). Design of multifunctional railway track inspection vehicle. *Journal of Physics: Conference Series 2174*, 012004.

Appendix A. Prompt for Inspection Plan Generation

The following prompt was submitted to Gemini 3.0 Pro to automatically generate structured railway inspection plans:

LLM Prompt

Develop a railway inspection plan using the file `maintenance_tasks.json`. The output must be structured with the fields: Day, Time, Location, Required action, Pathology ID.

File contents: Each entry includes Maintenance ID, Pathology ID, line/station, exact location, maintenance type, estimated duration (min), night task indicator, and priority score.

Network Topology and Distances: The network is a star topology centered at **Santiago**. Distances between consecutive stations are approximate rail distances based on the provided coordinates:

- **Route 1: Santiago → A Coruña (North)**
  - Santiago → Ordes: 28 km
  - Ordes → Meirama: 12 km
  - Meirama → Uxes: 11 km
  - Uxes → A Coruña: 9 km
  - *Total Line Length: 60 km*
- **Route 2: Santiago → Vigo (South-West)**
  - Santiago → Padrón: 20 km
  - Padrón → Vilagarcía: 25 km
  - Vilagarcía → Pontevedra: 25 km
  - Pontevedra → Arcade: 12 km
  - Arcade → Redondela: 8 km
  - Redondela → Vigo: 10 km
  - *Total Line Length: 100 km*
- **Route 3: Santiago → Ourense (South-East)**
  - Santiago → Boqueixón: 12 km
  - Boqueixón → Silleda: 20 km
  - Silleda → Lalín: 13 km
  - Lalín → Carballiño: 25 km
  - Carballiño → Ourense: 20 km
  - *Total Line Length: 90 km*

Operational constraints:

- Vehicle operates only between 23:30 and 05:30.
- Maximum speed 80 km/h; travel times must be calculated based on the distances provided above.
- Vehicle may start at any station of the previously mentioned waypoints.
- Service must end at a valid station of the previously mentioned waypoints.
- If tasks are incomplete, continue next day from the last station.
- Minor exceedances past 05:30 are allowed only if justified and minimal (close to station, short remaining task, or high-priority task).

Objective: Generate a plan covering all tasks realistically according to travel times, time window, task durations, maximum speed, routes and stations, and any other necessary parameters.

Required plans: Generate three separate plans: (1) Fastest possible, (2) Intermediate solution, (3) Priority-based (Try to repair as much as possible). Save each plan as a separate CSV.