Thursday, April 30, 2026

TReC 2026 Project Proposal Submission Form

Submit your project proposal for the 7th TRAIL Research Camp (August 24th - September 4th, 2026, Lausanne, Switzerland). Please complete all required sections and submit your proposal before April 30th, 01:00 PM (CET).

Administrative Data

Full Name of Team Leader Melih Taki

Contact Email melih.taki@dtsc.be

Project Information

Project Title Improvement of Railway Incident Management using Temporal Causal Analysis and Simulation

Profile of the Team Leader(s) & Expected Team Composition

Profiles:
Melih Taki - applied simulation/transportation engineering
Gian Marco Paldino - Theoretical causal machine learning

Melih Taki is a research engineer specialized in complex system modeling (digital twins, graph-based approaches, optimization) for decision support in constrained environments. He graduated in civil engineering from UMONS and works in R&D within an AI lab (BRAIN), where he structures technical architectures and develops data-driven frameworks. He builds data pipelines and prototypes, especially in railway systems, while coordinating academic and industrial stakeholders and leading small teams. He has strong skills in programming (Python, JS, C++, Go, SQL), AI (PyTorch, TensorFlow, ML, LLMs), data architecture, cloud/DevOps (AWS, Azure, Docker), and cybersecurity (PKI, OpenSSL). His experience also includes railway security, IoT systems, and optimization, combined with soft skills such as autonomy, proactivity, communication, and Agile methodologies

Gian Marco Paldino is a Postdoctoral Researcher at the Machine Learning Group of ULB, specializing in Multivariate Time Series analysis, forecasting, and causal inference. With a background in Computer Science and Engineering, his work spans digital twins, MLOps, and fraud detection, resulting in numerous Q1 publications. He combines deep technical expertise in Python/PyTorch with practical experience in industrial research, system deployment, and academic teaching/mentoring.

This complementarity is strictly necessary to bridge the gap between abstract causal graphs and concrete simulation environments.

Have you already identified potential team members for your project? No

Domain of Application Mobility & Logistics

Scientific Theme Physics-based Models & Digital Twins

Proposal Content

Abstract

This project addresses the critical issue of delay propagation in the North–South railway junction in Belgium, a highly saturated corridor where local perturbations can cascade into nationwide disruptions. Current approaches primarily rely on statistical correlations or predictive models, which fail to capture the underlying causal mechanisms driving these failures and therefore provide limited support for actionable decision-making.
The objective of this project is to identify and validate causal patterns of delay propagation using time-series data and causal discovery methods such as Granger causality and PCMCI. By focusing on a restricted and well-defined subnetwork, the project aims to uncover robust causal chains linking local disturbances to large-scale impacts. These insights are then leveraged to design and evaluate counterfactual operational strategies, enabling the assessment of targeted interventions (e.g., delay mitigation at critical nodes or sequencing adjustments) and their effect on overall network performance. To ensure methodological reliability, the approach combines real railway data from Infrabel with controlled synthetic scenarios, allowing the benchmarking and validation of causal inference methods under known conditions. The final outcome is a proof-of-concept decision-support module that translates causal analysis into actionable recommendations, including quantified improvements in delay propagation metrics under simulated scenarios.
In addition, the project adopts a socio-technical perspective by providing an operational guideline and audit framework to support the interpretation, validation, and responsible use of causal results in railway operations. The deliverable includes a reproducible and reusable “brick” integrating causal analysis and simulation, designed for extension to other transportation and infrastructure systems.

Background Information & Problem Statement

The North–South junction represents a real, structural problem because it concentrates a disproportionate share of Belgium’s rail traffic into a single, already saturated corridor, creating a system with very low resilience. With around 1,200 trains per day, which 300 of those in peak time [INF], even minor local disruptions—such as rolling stock delays, signaling issues, or dwell time variations—quickly cascade into nationwide delays due to the absence of equivalent alternative routes. This makes it a textbook single point of failure at a national scale.

Current approaches, including timetable optimization, statistical correlation, and conventional machine learning models [HAN14], fall short because they primarily model observable outcomes (delays, congestion patterns) rather than the underlying causal mechanisms. They are effective at predicting *when* disruptions may occur, but not *why* they emerge specifically in this critical segment, as they struggle to capture high-dimensional, context-dependent interactions between infrastructure constraints, traffic density, operational decisions, and real-time perturbations.

As a result, their predictive power does not translate into actionable prevention. A causality-driven approach, focused on identifying and modeling the root causes of failures within this specific environment, brings real added value: by disentangling the interactions between signaling behavior, train sequencing, infrastructure bottlenecks, and local operational dynamics, it becomes possible to isolate the actual triggers of disruption and act directly on them. This shifts the paradigm from reactive prediction to targeted intervention in the most critical part of the network.

Project Objectives & Concrete Implementation

A. Objectives

The project focuses on identifying and validating causal patterns of delay propagation within the North–South junction, a highly constrained railway segment where local perturbations can trigger large-scale network disruptions. Unlike traditional approaches that rely on statistical correlations or predictive models, this project explicitly aims to uncover causal mechanisms driving delay propagation.

More specifically, the objective is to detect and quantify causal relationships in time series data describing train movements and delays, using established causal discovery methods such as Granger causality, transfer entropy, and modern causal network reconstruction techniques. These methods enable the identification of domino effects, where delays propagate through sequences of trains and infrastructure constraints.

The project also aims to reconstruct simplified causal chains linking upstream perturbations to downstream impacts, thereby enabling the identification of root causes of disruptions within this critical bottleneck. This aligns with research in complex socio-technical systems, where failures emerge from interactions rather than isolated events

As a secondary objective, if there is still time, the project explores counterfactual simulation, allowing the evaluation of alternative operational scenarios (e.g., mitigation of a specific delay source) and their impact on overall system performance. This step bridges the gap between causal analysis and decision support.

Beyond methodological contributions, the project aims to demonstrate how causal insights can support operational decision-making for railway traffic management stakeholders (e.g., infrastructure managers such as Infrabel). In particular, the project evaluates whether causal-informed interventions can reduce delay propagation, improve network robustness, and support more resilient incident management strategies.

To ensure practical relevance, the project defines measurable outcomes, including:

- reduction in delay propagation length and intensity,
- comparison of baseline vs. causal-informed rerouting strategies,
- robustness of the network under simulated disruption scenarios.

The project adopts a socio-technical approach to make causal analysis interpretable, reliable, and actionable for railway stakeholders. In complex systems, decisions depend not only on data but also on operational practices, safety constraints, and regulations. It therefore proposes a lightweight SSH deliverable (operational guideline + audit framework) defining how to interpret causal relations, identify risks (e.g., spurious causality, misinterpretation), and apply validation procedures aligned with railway standards. It also considers human factors and adoption to ensure integration into incident management workflows, aiming to bridge advanced analytics and responsible real-world use.

B. Concrete Implementation

First, the scope is explicitly limited: the project does not attempt to reconstruct a full causal graph of the railway network, but instead focuses on a small number of predefined causal patterns, such as sequential delay propagation, congestion amplification at bottlenecks, and dwell-time or signaling-induced delay spread. This restriction is critical, as full causal discovery in large-scale networks remains computationally challenging and sensitive to noise.

The first stage consists of generating synthetic time series data with known causal structures and calibrated SUMO simulator [SUMO]. Controlled delay propagation scenarios are injected to create a ground truth against which different causal discovery methods can be evaluated. This step is essential to validate the reliability of the methods, as purely observational data can lead to spurious correlations and incorrect causal inference. The output of this stage is a benchmark comparing methods such as Granger causality, transfer entropy, and PCMCI, based on their ability to recover known causal links [SCH23].

In the second stage, the selected methods are applied to real railway punctuality data, focusing on a reduced subset of the North–South axis (e.g., Brussels North–Central–Midi). Time series representing delays and operational events are analyzed to extract causal relationships between trains and nodes. This enables the identification of dominant delay propagation chains, as well as candidate root causes. The results are expressed through interpretable representations such as causal graphs or propagation chains, providing a structured view of the system dynamics.

In the third stage, the project introduces counterfactual simulation. Based on the identified causal relationships, alternative scenarios are constructed by modifying one or two key causal drivers (e.g., reducing delays at a critical node or changing train sequencing), depending on the time left for the project. These scenarios are simulated in SUMO and compared to the observed baseline, allowing the quantification of the impact of specific interventions. This approach transforms causal analysis into actionable insights, supporting operational decision-making [GUA25].

B. Final outputs

The final outputs include the identification of a limited number of causal delay patterns within the North–South junction, supported by quantitative evidence of domino effects and root causes. In addition, a prototype counterfactual simulation tool enables the comparison between observed delays and simulated improvements under alternative scenarios. These results provide a clear and actionable understanding of how local disruptions propagate and how they could be mitigated.

Do you plan to deliver, as an outcome of your project, a reusable “brick” for the TRAIL Factory (https://factory.trail.ac/en/home_page) that could later be transferred and converted into a company process?

Yes

Briefly describe what the brick would be and its intended users.

The project aims to deliver a reusable module combining:

- causal analysis pipeline for time-series data,
- and a simulation-based evaluation framework for decision strategies.

This “brick” can be extended to other transportation or infrastructure systems.

To ensure immediate reusability by the TRAIL Factory, the code will be delivered as a well-documented Git repository. It will include:

Containerization: A Dockerfile to guarantee a reproducible environment (especially for the SUMO dependencies).

Testing: Unit tests for the causal graph generation and the routing logic.

Documentation: A comprehensive README with tutorials on how to plug in new datasets (e.g., from other infrastructure networks like tramways or logistics chains).

Project Dataset

The core structural dataset is the Infrabel track segments dataset, which provides a highly detailed geometric representation of the railway network, including all track segments, switches, and connections, with a precision ranging from 10 cm to 1 meter. This dataset is used to extract and isolate the North–South junction and reconstruct a faithful micro-network topology. Unlike higher-level representations based on stations or operational points, this track-level view captures the real physical constraints of the network, such as merging flows, switches, and local bottlenecks, which are essential to understanding delay propagation mechanisms.
Source: https://opendata.infrabel.be/explore/dataset/geografische-positie-van-alle-spoorsegmenten

This structural layer is combined with the punctuality dataset, which contains real train circulation records. Each entry corresponds to a train passing at a given point, with planned and actual arrival and departure times, allowing delays to be computed precisely. This dataset provides the temporal dimension of the system, enabling the reconstruction of time series of delays for each train as it moves through the junction. By aligning successive passages, it becomes possible to observe how delays propagate across trains and track sections in real conditions.
Source: https://opendata.infrabel.be/explore/dataset/stiptheid-gegevens-maandelijksebestanden

To obtain precise platform-level information and real-time operational data, the project will leverage the iRail API (https://api.irail.be), which provides structured access to Belgian railway data including stations, platforms, live departures/arrivals, and delay information. This enables the reconstruction of a more granular representation of the network, including track/platform assignments and temporal train movements.

Additional datasets can complement the core analysis: Infrabel provides simplified track and station-to-station data (useful for prototyping, visualization, validation without full micro-topology), enabling faster experimentation when precision is not critical. External sources (GTFS for timetables, OpenStreetMap for geographic context) can enrich infrastructure models or support simulations. These datasets remain optional, offering flexibility (routing, simulation, visualization) without creating dependencies, since the core pipeline relies entirely on primary Infrabel data.

The datasets are combined to build a spatio-temporal model of the North–South junction (track segments = physical network, train passages = dynamic signals). The junction is extracted by geographic filtering and mapping passages onto tracks, creating a micro-network (nodes = key zones, edges = connections, signals = delay time series).

Detailed Work Plan

DAY 1:
Data ingestion and environment setup. Download and structure Infrabel datasets (track segments + punctuality logs). Extract a restricted micro-network around the North–South junction and build initial time-series (delays per train per point).
Output: clean, usable dataset + first time-series.

DAY 2:
Data alignment and preprocessing. Synchronize train passages across nodes, handle missing values, and construct consistent temporal sequences for causal analysis.
Output: finalized time-series dataset ready for modeling.

DAY 3:
Synthetic data generation. Create controlled delay propagation scenarios with known causal structures (ground truth).
Output: synthetic dataset with injected causal patterns.

DAY 4:
Causal method implementation and evaluation. Test a small set of methods (Granger, PCMCI) on synthetic data and compare performance.
Output: benchmark of methods + selected approach.

DAY 5:
Application on real data. Apply the selected causal method to the junction dataset and extract first causal relationships and propagation chains.
Output: initial causal graph / delay propagation patterns.

DAY 6:
Refinement and filtering. Clean causal results, remove noise, and focus on 2–3 robust patterns (e.g. sequential propagation, bottleneck effects).
Output: validated causal patterns.

DAY 7:
Synthetic delay injection on real structure. Inject controlled delays into real trajectories to validate causal robustness under different scenarios.
Output: stress-tested causal relationships.

DAY 8:
Counterfactual simulation. Modify key causal drivers (e.g. reduce delay at a node) and simulate

alternative scenarios.
Output: comparison between baseline and counterfactual outcomes.

DAY 9:

Packaging of the TRAIL Brick: code refactoring, pipeline structuring, and creation of a reusable repository integrating causal analysis and simulation. Reproducibility ensured via Docker, with testing and documentation. An Operational Guideline & Audit Framework defines causal interpretation, known limitations, and simple validation procedures.

Output: clean reusable prototype with operational guideline and audit framework.

DAY 10: Prototype finalization, visualization of causal chains and delay propagation, and delivery to the scientific committee. A short decision-support document for non-technical users explains causal results, usage conditions, limits (data sparsity, spurious causality), and includes audit/validation procedures with railway examples.

Output: finalized prototype, presentation-ready results, and decision-support/audit document for operational use.

Bibliographic References

[INF] Infrabel, "La jonction Nord-Midi," [Online]. Available: https://infrabel.be/fr/mediakit/jonction-nord-midi. [Accessed: Apr. 30, 2026].

[HAN14] . A. Hansen and J. Pachl, *Railway Timetabling & Operations: Analysis, Modelling, Optimisation, Simulation, Performance Evaluation*. Hamburg, Germany: Eurailpress, 2014. ISBN: 978-3962450892

[SUMO] DLR - German Aerospace Center,"Railway Simulation in SUMO", 2024. Available at: https://sumo.dlr.de/docs/Simulation/Railways.html.

[GUA25] Guastella, D., Silva-Muñoz, M., Montero-Porras, E., & Bontempi, G. (2025). A Simulation Tool to Assess the Impact of Deviation Plans on Disruptive Events of Urban Traffic. In Proceedings of the 15th International Conference on Simulation and Modeling Methodologies, Technologies and Applications SIMULTECH - Volume 1, 51-61, 2025 , Bilbao, Spain https://www.scitepress.org/Link.aspx?doi=10.5220/0013518900003970

[SCH23] V. Schoonderwoerd, *"Causal discovery from train network data with background knowledge,"* Master's thesis, Utrecht University, Faculty of Science, Graduate School of Natural Sciences, Utrecht, Netherlands, Jun. 2023.

Eligibility & Evaluation

Does the project include multidisciplinarity between STEM & SSH? Yes

How?

The project integrates multidisciplinarity between STEM and SSH by combining advanced data-driven methods (causal inference, time-series analysis, and simulation) with a socio-technical perspective on their use in real-world railway operations. Beyond the development of analytical models, the project explicitly addresses how causal insights are interpreted, validated, and applied by operational stakeholders.

To support this integration, the project delivers an Operational Guideline and Audit Framework that defines rules for interpreting causal relationships, identifies potential risks such as spurious causality or misinterpretation, and proposes simple validation procedures aligned with railway operational constraints. It also considers human factors, decision-making practices, and adoption challenges to ensure that the proposed methods can be effectively integrated into incident management workflows.

This approach ensures that the project does not only produce technically sound results, but also delivers interpretable, auditable, and actionable outcomes, bridging the gap between advanced analytics and responsible, real-world use in critical infrastructure systems .

We confirm that the Team Leader will be present for the full duration of TReC'26 if the project is selected (August 24th - September 4th, 2026, Lausanne, Switzerland)

I/We agree and confirm

Additional Comment

Risk 1:
Building a detailed micro-level representation of the Brussels North–South junction may be time-consuming and complex, especially if all tracks, switches, and routing configurations are modeled with full precision in SUMO. This level of detail could slow down the implementation phase and delay the causal analysis.
Mitigation: We will begin Week 1 with a simplified representation of the North–South junction, limited to the most critical nodes, tracks, and switches involved in delay propagation. The objective is to preserve the main operational bottlenecks while avoiding unnecessary modeling complexity. If the simplified model works reliably, additional track-level details can be progressively integrated, and scaling to a larger network will only be considered if time allows.

Risk 2: Lack of granularity in open dataset. Temporal causal discovery (e.g., PCMCI) requires high-resolution time-series data, whereas some public data is aggregated.
Mitigation: We have already verified that Infrabel Open Data provides minute-level event logs via API. As a fallback, we have a Python script ready to generate synthetic, causally-linked delay data to ensure the optimization layer can be developed without data bottlenecks.

Risk 3 : A main issue is low disruption variability (rare, uneven incidents), making causal detection difficult;
Mitigation : this is mitigated by selecting worst-performing months and injecting synthetic delays (controlled scenarios, known causality, more pattern diversity, fewer spurious correlations).