Transport and Communications, 2025; Vol. II.
ISSN: 2730-0854
DOI: 10.26552/tac.C.2025.2.1

Synthetic Data for Resilient Urban Traffic Systems: A Methodological Framework

Maroš Jakubec[1], Radovan Madleňák[2], Michal Palčák[1], Pavol Kudela[1], Eva Jakubcová[1], Daniel Gachulinec[2], Viktória Cvacho[2]

[1]University Science Park UNIZA, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia
[2] Faculty of Operations and Economics of Transport and Communications, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia

Abstract Urban traffic models often struggle with rare and disruptive events because real-world data for such situations are limited. This article presents the SynTraffic project, which explores the use of synthetic traffic data to support more robust and resilient traffic modelling. The proposed approach combines real-world observations with artificially created traffic scenarios to expand the range of conditions available for model development. The methodology is demonstrated using the city of Žilina as a representative urban case with complex traffic patterns and long-term monitoring infrastructure. The article focuses on the theoretical background, methodological design, and expected benefits of synthetic data in intelligent transportation systems, highlighting its potential to address data scarcity, support privacy-aware analysis, and improve the handling of unusual traffic conditions.

Keywords synthetic traffic data, urban traffic modelling, intelligent transportation systems, traffic resilience

JEL R41, L91

1. Introduction

Urban mobility systems face persistent challenges arising from traffic congestion, inefficiencies, and operational disruptions. In the European Union, urban traffic jams are estimated to cost €110 billion annually, reflecting substantial losses in productivity and environmental quality [1]. Cities like Žilina (a mid-sized transport hub in Slovakia) grapple with these pressures as travel demand continues to grow while infrastructure capacity remains limited.

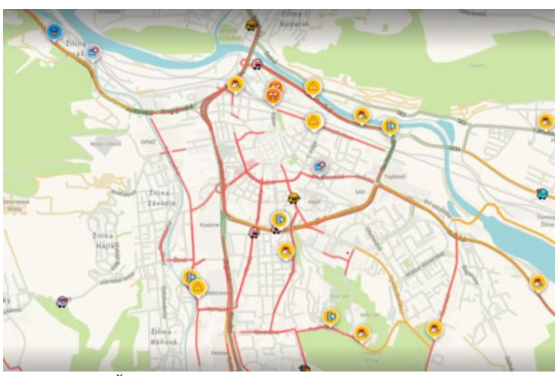

Figure 1. The Žilina traffic volume during extreme traffic conditions.

As illustrated in Fig. 1, extreme traffic conditions in Žilina can escalate rapidly, revealing how sensitive the network is to adverse or unexpected situations. Traditional traffic management strategies (fixed-timed signals, static models) often struggle to perform under such dynamic or extreme circumstances, including accidents or severe weather events. These limitations highlight the need for more adaptive, responsive, and data-informed approaches.

These challenges explain why modern traffic systems increasingly rely on flexible and predictive decision-support tools capable of handling unexpected disruptions. The field has seen a marked rise in the adoption of Intelligent Transportation Systems (ITS) supported by Artificial Intelligence (AI). Deep learning techniques are now widely explored for traffic prediction, adaptive control, and incident detection [2]. Recent research has shown that AI models, particularly those using generative techniques, can maintain strong performance even under adverse or unusual conditions. For instance, Liu et al. [3] proposed a two-stage deep model that delivers robust scene segmentation in heavy fog and snow, while Lee et al. [4] demonstrated that a Generative Adversarial Network can translate LIDAR data from clear weather into rainy or foggy conditions, , thereby improving model generalization and enhancing perception in adverse environments. Such findings underscore the potential of AI to enhance traffic safety and operational efficiency.

A fundamental limitation of these methods is the need for large, diverse, and representative training datasets. Real traffic data often captures only standard conditions, making it difficult to model rare but high-impact situations, such as a sudden citywide jam during a blizzard. Collecting such data is not only costly but also constrained by privacy and legal regulations. Camera-based and connected-vehicle data may include personal identifiers, restricting how they can be shared or processed under frameworks like GDPR [5]. These constraints create a data bottleneck, where the scenarios most relevant for improving system resilience are precisely the ones with the least available data.

Synthetic traffic data provides a practical solution to this challenge. It consists of artificially generated datasets that retain key statistical patterns of real traffic without exposing personal information. Modern generative models make it possible to create realistic artificial samples that supplement real-world training data [6]. This approach helps fill gaps related to rare events, increases overall data volume, and inherently supports privacy protection. Forecasts from the European Data Protection Supervisor suggest that synthetic data may become the dominant source for AI model training by 2030 [7]. In ITS, synthetic data allows traffic models to learn from accidents, extreme weather, or infrastructure failures without needing these events to happen in real life.

The Synthetic Traffic Data for Mobility and Resilience (SynTraffic) project builds directly on this shift toward synthetic data in AI-based mobility research. Developed at the University of Žilina, the project aims to integrate generative AI and simulation techniques to improve the modelling of urban traffic dynamics. Its methodology is based on combining real and synthetic datasets so that AI-based prediction and control models can learn to operate effectively under both routine and atypical conditions that may arise in a city such as Žilina. The objective is to support the development of traffic management strategies that remain robust, maintaining flow and safety even when the system is exposed to unexpected disturbances. From a theoretical perspective, SynTraffic operates at the intersection of transportation engineering, machine learning, and data science.

The following sections present the theoretical foundations of the SynTraffic methodology and the context in which it is applied. They outline the key concepts that guide the project, including the use of generative models and the integration of real and synthetic data in the modeling process. They also describe the expected improvements in model performance and system resilience and explain how this approach supports both scientific understanding and practical development in smart urban mobility.

2. Materials and Methods

This section outlines the methodological foundation of the SynTraffic project, which combines generative modelling, data integration, and resilience-oriented design to enhance the performance of urban traffic prediction and control systems. It first presents the generative approaches used to create synthetic traffic data, including data-driven models and microscopic simulation tools. It then describes the framework for integrating synthetic and real datasets to ensure consistent, balanced, and reliable training conditions for machine learning models. Finally, it introduces the theoretical principles that guide the project's focus on resilience, explaining how diverse data sources and systematic validation contribute to robust behaviour under both typical and stress conditions.

2.1. Generative Modelling for Synthetic Data

Generative modelling forms a central component of the SynTraffic project because it provides a way to create synthetic traffic data that reflects the patterns found in real measurements from the transport system. The project focuses on developing and evaluating generative models for creating synthetic traffic data, in particular Generative Adversarial Networks (GANs) [8], [9] and Variational Autoencoders (VAEs) [10]. In parallel, microscopic traffic simulation tools are used as a separate source of scenario-based data. Together, these approaches support the generation of synthetic images, time series, and traffic scenarios that represent a broad range of conditions, including situations that are difficult to capture through real-world observations.

GANs will be developed and tested in SynTraffic to learn how traffic situations appear in real datasets and to generate new samples that follow the same structure. A GAN consists of two neural networks trained together. The generator produces synthetic data, while the discriminator evaluates whether the data looks real or artificial. Through repeated updates, the generator gradually improves its ability to produce samples that the discriminator cannot reliably differentiate from the real data. This makes GANs useful for producing high-fidelity synthetic traffic images and time series, especially in situations where traffic patterns vary strongly with environmental context. The project uses established techniques to improve training stability and to encourage the model to produce a broad variety of scenarios, including during different times of day or under different weather conditions. The resulting synthetic datasets can extend the coverage of the real dataset, thereby supporting the development of more robust prediction and detection models.

VAEs provide a complementary approach. A VAE learns a latent space that captures the essential structure of real traffic data, and from this space it can sample new variations. The model consists of an encoder, which compresses the data into a latent representation, and a decoder, which reconstructs the data from this representation. This probabilistic framework allows the model to generate smooth variations of typical traffic behaviour and to provide plausible values in regions of the dataset where real measurements may be missing. In SynTraffic, VAEs will be used primarily for numerical data, such as traffic flows or speeds, where learning a structured latent space can support the generation of consistent synthetic sequences and help describe how different traffic states relate to one another.

Alongside data-driven generative models, SynTraffic uses microscopic traffic simulation tools. Simulation provides a controlled virtual environment in which traffic demand, roadway layout, weather, and incidents can be introduced with precision. This enables the generation of physically consistent data for rare or safety-critical situations that cannot easily be collected in real operations. Simulation also provides complete ground truth for all elements of the system, which is valuable when training or evaluating machine learning models [11]. In the project, simulation serves two purposes. First, it produces synthetic datasets for underrepresented scenarios. Second, it acts as a reference for checking whether GAN- or VAE-generated data remains consistent with the expected physical behaviour of the transport network.

Together, GANs, VAEs, and microscopic simulation provide a diverse toolbox for generating synthetic data. Each method contributes different strengths: GANs for realistic detail, VAEs for structured variation and stability, and simulation for physical consistency. Table 1 summarises the roles of these approaches in the project.

Table 1. Approaches for Synthetic Traffic Data Generation

Approach	Strengths	Challenges
GAN e.g., to create realistic traffic images or time-series	– High realism in outputs; captures complex correlations in data *implicitly*. – Can produce detailed, high-fidelity synthetic examples (e.g., sharp images of traffic scenes).	– Difficult to train (adversarial instability). – May suffer from mode collapse (missing some scenario types). – Requires large training dataset and careful tuning to get broad coverage.
VAE e.g., to model distribution of traffic flows and generate new variations	– Stable training with explicit likelihood optimization. – Learns an organized latent space of traffic patterns, enabling controlled sampling and interpolation. – Good for data augmentation and imputing missing data with uncertainty.	– Generated data can be blurred or less precise (tendency to average outputs). – Lower detail for visual data; might miss fine nuances without further refinement.
Traffic Simulation e.g., to simulate accidents, roadworks, or new infrastructure virtually	– Ensures physical and logical consistency (traffic obeys rules of the road). – Can produce rare dangerous events safely and provide complete ground truth labels. – Flexible scenario design (any "what-if" can be tested by setting initial conditions in the sim).	– Relies on quality of simulation models (driver behaviour, vehicle dynamics); if those are imperfect, data may not reflect real human behaviour. – Computationally intensive for large networks or many repeated runs. – Needs expert input to set up realistic scenarios (calibration to real city conditions).

2.2. Integrating Synthetic and Real Data in Traffic Models

The value of synthetic data depends on how effectively it can be combined with real measurements for model training and analysis. SynTraffic therefore develops a structured framework for integrating synthetic and real traffic data so that machine learning models can treat them as a unified information source. This includes harmonizing data formats, balancing the contribution of different data types during training, and validating the integrated dataset to ensure that synthetic data supports rather than distorts model behaviour.

The first step is data harmonization. All synthetic samples generated by GANs, VAEs, or simulation tools are transformed to match the structure and feature space of the real dataset. If the real data contains fields such as timestamp, location, vehicle count, and average speed, the synthetic data is required to follow the same schema. Units and measurement scales are aligned, and multi-source data is placed on a common timeline. This reduces the risk that the learning algorithm identifies accidental cues that distinguish synthetic from real data. Scenario metadata may be used during training when appropriate, either to test model sensitivity or to enrich the variability of training conditions.

A central methodological consideration is avoiding bias when mixing synthetic and real data. If the datasets were simply concatenated, one type of data could dominate the training process. SynTraffic will therefore use controlled sampling and weighting strategies. For example, early training may include more synthetic samples to expose the model to diverse conditions, while later refinements may rely more heavily on real data. This reflects standard ideas from domain adaptation, where synthetic data is treated as a related source domain. After training on mixed data, a fine-tuning step on purely real data can correct small differences between the two domains. This strategy aligns with findings from related work in advanced driver assistance systems, where combining synthetic and real data and then fine-tuning on real examples produced the best performance [12].

SynTraffic also performs statistical checks to validate the combined dataset. Model outputs are compared across real, synthetic, and mixed inputs to detect whether the model reacts differently depending on the data source. If a model consistently predicts unrealistic patterns when fed synthetic samples that should match real-world conditions, this signals a domain shift that must be corrected. Validation is carried out both on holdout sets of real data and on synthetic scenarios with known ground truth from simulation. Any discrepancies feed back into refining the generative models or adjusting the synthetic-to-real data ratio.

Data quality assurance is included throughout the integration process. Real traffic data may contain noise, outliers, or missing values, and synthetic data can occasionally produce artifacts. SynTraffic applies filtering and outlier detection to remove or correct values that violate expected physical constraints, such as negative traffic counts or sensor malfunctions. The goal is to produce a coherent dataset that is both diverse and reliable. Conceptually, this approach addresses the bias-variance trade-off by reducing variance through data cleaning and reducing bias by incorporating a wider variety of synthetic scenarios.

After integration and training, the project performs resilience testing of the resulting models. The models are evaluated under stress conditions generated through simulation or synthetic data to confirm that the integration process has achieved the intended improvements in robustness. If performance does not meet expectations, the integration loop is repeated with revised weighting or updated generative models. This iterative testing reflects the project's focus on robust optimization and resilience assessment for urban traffic systems.

2.3. Theoretical Basis for Resilience in Traffic Systems

A central goal of SynTraffic is to strengthen the resilience of urban traffic systems by improving the quality and diversity of data used to train predictive and control models. In theory, resilience refers to the ability of a system to absorb disturbances, adapt to changing conditions, and recover quickly from disruptions. In the context of traffic networks, this means that during events such as accidents, extreme weather, or sudden increases in demand, the system should continue to operate at an acceptable level and restore normal flow without cascading failures.

Synthetic data contributes to resilience by filling knowledge gaps that make traffic models fragile. Many operational models perform poorly during extreme or unusual situations simply because they were never exposed to them during development. By incorporating synthetic examples of severe weather, infrastructure failures, or unexpected demand surges into training, SynTraffic broadens the range of conditions that the AI system can recognize and respond to. This follows the same logic used in other safety-critical fields, where simulation is used to prepare systems for low-frequency, high-impact events. The theoretical basis is expanded scenario coverage: models trained with synthetic data effectively gain experience with a wider portion of the system's state space.

A predictive model that remains reliable under stress allows traffic management strategies to be implemented proactively. If the model can forecast that a specific combination of conditions, such as a major event combined with heavy rain, will result in congestion within a short time frame, operators can adjust signal timings, reroute vehicles, or issue traveller information in advance. Without this predictive capability, intervention would occur only after congestion fully develops, which is often too late to prevent large delays. The principle is straightforward: better and more diverse information directly enhances the system's ability to act resiliently.

Resilience also involves the ability to handle situations that were not explicitly seen during training. No dataset can include all possible future scenarios, so SynTraffic emphasizes generalization. By exposing the models to a wide range of synthetic variations, the approach reduces dependence on narrow correlations and encourages learning of broader patterns, such as the effect of weather on vehicle speeds or how traffic shifts when a major link becomes unavailable. This aligns with established theory in machine learning that adding diverse samples near the edges of the input domain can significantly improve a model's performance in novel or partially unfamiliar contexts [13].

Another source of resilience in SynTraffic comes from redundancy in data and modelling approaches. Because the project uses synthetic and real data together, along with multiple generative models and simulation tools, it avoids reliance on any single data stream or method. If one model produces unrealistic outputs for a particular scenario, others can provide a corrective reference. This ensemble of data sources creates a safety net and supports the idea of hybrid modelling, where physics-based and data-driven components can complement each other to improve robustness in complex systems.

Ethical and legal considerations also form part of the theoretical basis for resilience. Traffic data often includes sensitive information, and its use is constrained by privacy regulations. Synthetic data offers a practical way to reduce dependence on personal data while still enabling detailed analysis and model development. This aligns with principles of data protection by design. Synthetic data can also be used to correct imbalances in the real dataset and improve fairness. If certain regions or conditions are underrepresented in real data, additional synthetic samples can be generated to balance the distribution. The project monitors model performance across different subsets of the dataset to detect potential biases and address them through targeted synthetic data generation. In this way, synthetic data supports not only technical resilience but also compliance and fairness in the modelling process.

The theoretical foundation of SynTraffic rests on expanding scenario knowledge, improving generalization, incorporating redundancy, and embedding ethical safeguards. These elements together form a coherent strategy for building urban traffic models that can withstand disruptions and continue to operate reliably under a wide range of conditions.

3. Study Area and Data

This section introduces the study area selected for the SynTraffic project and describes the types of real-world data that form the basis for the methodological development. The aim is to provide a clear contextual background of the urban transport environment and to explain how available data sources support research focused on traffic behaviour under both normal and disrupted conditions.

3.1. Study Area: The City of Žilina

The study area of the SynTraffic project is the city of Žilina, a medium-sized urban area located in northwestern Slovakia. Owing to its geographical position, Žilina represents an important node within both the national and regional transport system. Several major transport corridors intersect in the city, linking domestic routes with cross-border connections towards the Czech Republic and Poland. Žilina also functions as a key road and rail hub, which is directly reflected in the intensity and diversity of traffic flows within its territory.

The transport system of the city is characterized by the simultaneous presence of multiple traffic types. Local urban traffic, suburban commuting, freight transport, and transit flows all overlap within a relatively compact road network.

In transport research, such an environment is typically described as having a high level of demand heterogeneity, where local and supra-local traffic relations intersect and interact. This structure makes the system sensitive to changes in traffic organization and external disturbances.

From a spatial perspective, Žilina has a compact urban core connected by a system of radial access roads and collector routes linking the city center with surrounding residential areas and the regional road network. This configuration is typical for Central European cities, where historically developed urban structures must accommodate modern traffic demand, including increasing regional and transit movements.

The city's road network includes several capacity-sensitive sections and signalized intersections that act as natural bottlenecks. Major road corridors such as the D1 and D3 motorways, which are part of the TEN-T network, pass through or near the city and are complemented by first-class roads including I/11, I/18, I/60, and I/61. Road I/18, connecting north and south, plays a particularly critical role, as transit traffic is frequently redirected through the city during motorway maintenance or incidents, especially in the Strečno section. The city ring road, known as "Veľká okružná" (I/60), concentrates a large share of both urban and transit traffic and represents a structurally vulnerable element of the network.

Traffic conditions in Žilina exhibit strong temporal and spatial variability. Daily traffic patterns change with commuting cycles, while weather conditions and extraordinary events such as accidents or infrastructure restrictions can rapidly degrade traffic performance. Long-term observations indicate growing levels of individual car traffic, which, combined with limited urban space, result in recurring congestion and increased sensitivity to disruptions. These characteristics create favourable conditions for studying unbalanced and dynamic traffic states.

Another important factor is the ongoing evolution of the city's transport system. Changes in traffic organization, interventions on higher-level road infrastructure, and regulatory measures related to parking or access management continually reshape traffic behaviour. Such processes increase the need for monitoring, prediction, and evaluation of transport system resilience.

3.2. Justification of the Site Selection

The selection of Žilina as the study area for the SynTraffic project is based on a combination of transport characteristics, data availability, and continuity of research activities. From a traffic engineering perspective, the city offers a compact yet complex system in which urban, suburban, and transit flows naturally overlap. This allows traffic behaviour to be observed and analysed at multiple scales within a single territory.

The structural complexity of the road network, including major entry radials, capacity-limited sections, and signal-controlled junctions, creates situations in which traffic conditions can deteriorate quickly due to increased demand, adverse weather, or partial infrastructure failures. These properties make Žilina a suitable environment for examining system behaviour under stress and non-standard conditions.

A key reason for choosing Žilina is the availability of long-term, consistent real-world traffic data. Previous research and implementation projects have established a monitoring infrastructure that enables continuous data collection with comparable structure over time. Initiatives such as CleverNet and EnCLOD laid the foundation for what is often referred to as a Living Urban Laboratory, where traffic data is used not only for research but also to support decision-making at the city level.

The fact that Žilina has already served as a reference area in traffic modelling and simulation studies further confirms its suitability for methodologically oriented research. Within SynTraffic, the city provides a realistic framework for exploring approaches based on the integration of real and synthetic data in the context of an operating urban traffic system.

3.3. Data Sources and Data Types

The SynTraffic project is based on a heterogeneous set of real-world data sources that together provide a comprehensive view of urban traffic behaviour. Rather than relying on a single data stream, the project integrates multiple data types capturing both traffic dynamics and external influencing factors. This multi-layered data structure forms the empirical foundation for subsequent analytical and modelling tasks.

The primary data layer consists of high-frequency traffic sensor data collected from the existing monitoring infrastructure in the city. These data include vehicle counts, traffic intensities, speed measurements, and basic flow characteristics. High temporal resolution is essential for capturing short-term fluctuations and transient phenomena typical of real traffic, such as sudden speed changes or gradual congestion formation.

Camera-based data represent an important complementary source. Video recordings and still images from selected road sections and intersections provide a visual representation of traffic conditions. This data supports the interpretation of sensor measurements, allows observation of interactions between road users, and helps identify non-standard situations that may not be fully reflected in numerical data alone.

Environmental and microclimatic data form another essential component. These data include meteorological variables such as precipitation, temperature, visibility, lighting conditions, and road surface state. Such factors have a well-documented influence on driver behaviour, network capacity, and safety, and are therefore critical for understanding variability in traffic performance.

The project also uses historical traffic records collected in previous research and operational projects. These include long-term traffic counts, accident records, and information on infrastructure failures or maintenance activities. Historical data provide context for current observations and are particularly important for characterizing rare but impactful events, which cannot be reliably analysed without empirical reference.

4. Expected Theoretical Outcomes

SynTraffic is designed as a methodological project, so the expected outcomes focus on theoretical advances supported by previous studies and early experiments. By training traffic models on a combined set of real and synthetic data, the project anticipates measurable improvements in predictive accuracy and operational robustness across a wide range of traffic conditions.

One expected outcome is improved performance in forecasting key traffic variables such as flow, speed, and travel time. Models trained with synthetic augmentation should achieve lower error rates than models trained only on real data, especially in scenarios that are rare or difficult to capture. Prior work supports this expectation. Zhu et al. [13] demonstrated that adding synthetic samples to real traffic volume data reduced forecasting error in a graph neural network model, particularly during peak and off-peak transitions. In SynTraffic, a deep learning model exposed to synthetic incident scenarios should predict the effects of real incidents more accurately than a model lacking such exposure, with anticipated error reductions consistent with the 10–20% improvements reported in related studies.

For vision-based tasks, such as incident detection or vehicle classification in camera streams, synthetic images of rare events are expected to enhance recognition performance. Dewi et al. [8] showed that generating additional images of rare traffic signs using a DCGAN improved a CNN's classification accuracy. SynTraffic aims to achieve similar benefits for detecting anomalies in road scenes. Early tests with YOLO-based models already indicate fewer false negatives when synthetic incident scenes are included in training, suggesting that exposure to a broader spectrum of examples helps the model recognize events that seldom appear in real footage.

Resilience is another key expected outcome. The models developed within SynTraffic should remain stable when confronted with scenarios they have not seen before, because the synthetic data broadens the domain of conditions the model has learned to interpret. In simulation-based stress tests, models trained with synthetic augmentation remain reliable under combinations of events that would typically cause baseline models to fail. For example, during a combined snowstorm and partial signal blackout, a conventional model might behave unpredictably, whereas a SynTraffic-trained model can still provide meaningful predictions by drawing on related synthetic examples seen during training. Similar patterns were observed by Jelić et al. [12] in the ADAS domain, where hybrid models trained with both synthetic and real data-maintained performance under domain shifts that caused real-only models to degrade.

The project envisions improvements in system-level behaviour once such models are applied in traffic control contexts. More accurate and robust predictions are expected to support proactive interventions, such as early rerouting or adaptive signal adjustments. At this stage, these effects are conceptual and will be explored in future simulation-based studies. The expectation is that control strategies informed by SynTraffic-enhanced models could help reduce congestion duration and limit queue formation during high-impact events when compared to traditional approaches.

A further anticipated outcome is the transferability of SynTraffic's methodology. By demonstrating in Žilina that synthetic data improves model accuracy and resilience, the project contributes a generalizable framework that can be adapted to other cities and transport domains. This includes applications such as public transport modelling, active mobility planning, or emergency response analysis. The framework shows how synthetic data can systematically fill gaps in real datasets and support more robust decision-support tools.

These outcomes will be substantiated through analytical evaluation and comparison with prior literature. Studies such [3], [8], [13] provide external references for expected improvements, while SynTraffic's internal evaluations will document specific error reductions, stability measures, and performance across incident scenarios. Results will include quantitative indicators, such as forecast accuracy under stress conditions, and qualitative assessments of model robustness in novel or unexpected environments. Together, these outcomes represent the theoretical advances expected from combining real and synthetic data in urban traffic modelling.

5. Discussion

The SynTraffic project offers several important insights for the field of urban mobility, particularly regarding how synthetic data and generative modelling can support more resilient traffic systems. One of the clearest contributions is demonstrating a practical approach to overcoming data scarcity in complex system modelling. Traditional traffic models have always been limited by the range of events captured in historical data or by simplifying assumptions needed for analytical models. SynTraffic expands this space by generating new, domain-consistent scenarios that enrich the learning process. This represents a shift in how transportation researchers can think about data: the dataset is no longer a fixed constraint but something that can be actively shaped to support better modelling. Similar shifts have already reshaped domains like robotics and autonomous driving, where simulation and synthetic data are routinely used to expose algorithms to conditions unlikely to be observed in practice. SynTraffic shows how the same principle can extend to traffic modelling by incorporating both empirical observations and hypothetical scenarios into a unified learning framework.

The project also highlights how data-driven intelligence can enhance urban resilience. Historically, resilience in traffic networks has depended on physical redundancy and incident response strategies. While these remain important, SynTraffic demonstrates that informational resilience can offer comparable benefits. A city may not be able to expand its infrastructure, but with predictive tools that have been trained on a broad range of synthetic scenarios, it can operate the existing network more effectively and prevent gridlock in cases where previous systems would fail. This has broader

DOI: 10.26552/tac.C.2025.2.1

implications for planning and policy. Investments in data, analytics, and predictive modelling can complement traditional infrastructure projects and, in some contexts, provide more cost-effective improvements in system reliability. The methodological approach developed in SynTraffic could integrate naturally with emergency response planning, climate adaptation strategies, and long-term mobility development by revealing which parts of the network are most vulnerable under synthetic stress scenarios and therefore require attention.

A distinctive feature of SynTraffic will be the coordinated use of generative models and traffic simulation within a single methodological framework. While many studies rely on individual techniques such as GAN-based data augmentation or standalone simulations, SynTraffic will combine these tools in a structured manner. Simulation will be used to create physically consistent traffic scenarios, while generative models will be explored to capture and extend patterns observed in data. This combined approach is expected to support a more flexible research workflow in which data-driven and model-based methods complement each other. It also suggests that advanced ITS methodologies can be developed for mid-sized cities using open-source tools and focused research efforts, without requiring large-scale or complex infrastructure.

The discussion must also recognize limitations of the approach. Synthetic data is only as comprehensive as the assumptions built into the generative models and simulations. There will always be events that fall outside the imagined scenario space. Sudden societal disruptions, such as pandemic lockdowns, show how quickly mobility patterns can deviate from established norms. SynTraffic reduces this risk by generating a wide variety of scenarios, but the models will still require continuous updates as real conditions evolve. Computational cost is another consideration. Training generative models and running extensive simulations require substantial resources. SynTraffic's use of high-performance computing at the University of Žilina provides insight into these trade-offs and raises questions about scalability. Transfer learning may help address this challenge by allowing models trained in one city to be adapted to another city with similar characteristics, but whether this can be done reliably remains an open question.

Ethical considerations also accompany the use of synthetic data. While synthetic data reduces reliance on personal information and aligns with data protection principles, it introduces new responsibilities. The assumptions encoded in simulated or generated scenarios must be transparent to avoid embedding unintended biases. If certain neighbourhoods, behaviours, or conditions are underrepresented or misrepresented, model performance may degrade in real-world use. SynTraffic addresses this by documenting scenario generation procedures and evaluating model performance across different spatial and temporal subsets. This helps identify gaps and guide the targeted generation of additional synthetic samples. In this way, synthetic data not only supports technical resilience but also strengthens transparency and fairness in model development.

6. Conclusion

The SynTraffic project demonstrates how combining artificial intelligence with simulation can open new possibilities for traffic management. By integrating the adaptive learning capabilities of AI with the structured scenario generation offered by simulation, and supporting both with a coordinated data management framework, the project shows that this hybrid approach is both theoretically robust and practically valuable. The expected improvements in prediction accuracy and system resilience illustrate that investing in data synthesis can be as important as innovating new model architectures. This challenges the long-standing assumption that better AI systems mainly require collecting more real-world data. Instead, SynTraffic shows that intelligently generated synthetic data can expand the operational range of traffic models in meaningful ways.

For practitioners and city planners, the project suggests what future traffic management systems may look like: tools that continually learn from a broad space of possible conditions, simulate upcoming risks, and prepare for emerging challenges. As cities face growing uncertainties related to climate events, changing mobility habits, and infrastructure constraints, such systems can help maintain reliable and efficient operations even under stress.

The findings also reinforce an important theoretical point: the data used for training strongly determines what an AI system can understand and predict. By expanding the dataset with realistic synthetic scenarios, SynTraffic improves model generalization and stability while preserving privacy and reducing dependence on scarce or sensitive real-world measurements. The project contributes to the theoretical development of intelligent transportation systems by demonstrating how synthetic data can be integrated into modelling frameworks and how it influences the behaviour of predictive models across a wider domain of conditions. It also provides a practical pipeline that other cities or research teams can adapt and extend.

More generally, SynTraffic provides an example of how AI-driven systems can be made more resilient by using synthetic data to fill in the gaps in empirical data collection. The methodological principles demonstrated here can extend beyond road traffic to other areas of smart cities where data is limited, expensive, or incomplete. For Žilina and similar cities, approaches like SynTraffic offer a path toward transport networks that operate with greater foresight, adaptability, and robustness. The theoretical insights and methodological tools developed through the project provide a foundation for future work and represent an important step toward building resilient and intelligent urban mobility systems.

ACKNOWLEDGEMENTS

The article is supported by the project "APVV-24-0282 - Synthetic Traffic Data for Improving the Resilience of Transport Systems".

REFERENCES

[1] European Court of Auditors. Audit preview: Urban mobility in the EU. Luxembourg, Apr. 2019. http://www.eca.europa.eu/en/Pages/Report.aspx?did=49865 (accessed Dec. 10, 2025).

[2] A. K. Haghighat, V. Ravichandra-Mouli, P. Chakraborty, Y. Esfandiari, S. Arabi, and A. Sharma, 'Applications of Deep Learning in Intelligent Transportation Systems', J. Big Data Anal. Transp., vol. 2, no. 2, pp. 115–145, Aug. 2020, doi: 10.1007/s42421-020-00020-1.

[3] P. Liu, Y. Chen, F. Yu, and Q. Zhang, 'Mastering adverse weather: a two-stage approach for robust semantic segmentation in autonomous driving', Vis. Comput., Oct. 2024, doi: 10.1007/s00371-024-03663-1.

[4] J. Lee, D. Shiotsuka, T. Nishimori, K. Nakao, and S. Kamijo, 'GAN-Based LiDAR Translation between Sunny and Adverse Weather for Autonomous Driving and Driving Simulation', Sensors, vol. 22, no. 14, Art. no. 14, Jan. 2022, doi: 10.3390/s22145287.

[5] E. Abdessater et al., 'A Novel Method for ECG-Free Heart Sound Segmentation in Patients with Severe Aortic Valve Disease', Sensors, vol. 25, no. 11, Art. no. 11, Jan. 2025, doi: 10.3390/s25113360.

[6] S. I. Nikolenko, 'Synthetic Data for Deep Learning', Sep. 25, 2019, arXiv: arXiv:1909.11512. doi: 10.48550/arXiv.1909.11512.

[7] European Data Protection Supervisor. TechSonar: Synthetic Data. Brussels, 2022. https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en (accessed Dec. 10, 2025).

[8] C. Dewi, R.-C. Chen, Y.-T. Liu, and S.-K. Tai, 'Synthetic Data generation using DCGAN for improved traffic sign recognition', Neural Comput. Appl., vol. 34, no. 24, pp. 21465–21480, Dec. 2022, doi: 10.1007/s00521-021-05982-z.

[9] D. Tamayo-Urgilés et al., 'GAN-Based Generation of Synthetic Data for Vehicle Driving Events', Appl. Sci., vol. 14, no. 20, Oct. 2024, doi: 10.3390/app14209269.

[10] D. P. Kingma and M. Welling, 'An Introduction to Variational Autoencoders', Found. Trends® Mach. Learn., vol. 12, no. 4, pp. 307–392, Nov. 2019, doi: 10.1561/2200000056.

[11] P. A. Lopez et al., 'Microscopic Traffic Simulation using SUMO', in 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI: IEEE, Nov. 2018, pp. 2575–2582. doi: 10.1109/ITSC.2018.8569938.

[12] B. Jelić, R. Grbić, M. Vranješ, and D. Mijić, 'Can We Replace Real-World With Synthetic Data in Deep Learning-Based ADAS Algorithm Development?', IEEE Consum. Electron. Mag., vol. 12, no. 5, pp. 32–38, Sep. 2023, doi: 10.1109/MCE.2021.3083206.

[13] K. Zhu, S. Zhang, J. Li, D. Zhou, H. Dai, and Z. Hu, 'Spatio-temporal multi-graph convolutional networks with synthetic data for traffic volume forecasting', Expert Syst. Appl., vol. 187, p. 115992, Jan. 2022, doi: 10.1016/j.eswa.2021.115992.