An Integrated Digital Twin Architecture for Real-Time Urban Air Quality Management

Manoj Herath, Hrishikesh Dutta, Roberto Minerva, Noel Crespi, Maira Alvi, Syed Mohsan Raza
Samovar, Telecom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France.
Email: {manoj.herath, hrishikesh.dutta, roberto.minerva, noel.crespi, maira.alvi, syed-mohsan_raza}@telecom-sudparis.eu

***Abstract*—The growing demand for smart cities necessitates advanced solutions to promote urban sustainability. Digital Twin (DT) technology offers transformative potential by synchronizing virtual and physical environments in real time and enabling capabilities such as monitoring, forecasting, simulation, and prescription. However, existing smart city approaches often emphasize virtual representation while lacking core DT functionalities and a unified, extensible framework. This paper introduces a comprehensive DT framework designed to support a broad range of urban applications with enhanced operational capabilities. Validated using real-world air quality data, the proposed system demonstrates significant improvements in emissions reduction, air quality monitoring, and pollutant management. Key contributions of this work include a function-rich and generalizable DT framework, AI-driven adaptability, and extensive validation through predictive analytics. This study establishes a replicable blueprint for metropolitan-scale DTs, balancing comprehensive functionality with responsive, data-driven urban analytics.**

***Index Terms*—Digital Twin, Smart City, Air Quality Management, Artificial Intelligence**

I. Introduction

The rapid growth of urban environments and the increasing demand for sustainable development have spurred global interest in innovative solutions for smart cities. Within this evolving landscape, Digital Twin (DT) technology has emerged as a transformative tool, offering a dynamic, virtual representation of physical urban systems that continuously synchronize with real-time data to reflect the city's behavior, conditions, and attributes [1]. By enabling real-time monitoring, predictive analytics, and scenario simulations, DTs support improved connectivity, optimized resource utilization, enhanced infrastructure efficiency, and stronger environmental sustainability initiatives.

Effective air quality management, an essential component of smart city development, requires accurate, real-time monitoring and future insights. These capabilities empower urban planners to reduce vehicle emissions, improve air quality, and support broader sustainability goals. Digital Twin technology offers a data-driven foundation for modeling pollution dynamics, including emissions and air quality indices. However, while existing smart city DTs address environmental monitoring, they often lack advanced functionalities and a unified, extensible framework for comprehensive pollution control across diverse urban landscapes.

Existing smart city DTs are application-specific and predominantly emphasize virtual 3D visualization [2], [3], rather than incorporating advanced digital twin functionalities such as predictive forecasting, "what-if" scenario analysis, and diagnostic capabilities. This limits their broader applicability and impact.

This paper addresses these limitations by introducing a smart city digital twin framework. Focusing on real-time air quality management, the proposed architecture is validated using data from IoT-based air quality sensors deployed across Madrid. Our framework supports advanced DT functionalities by integrating dynamic data processing and deep learning models to enable forecasting, simulation, and operational insights. Through comprehensive experimentation, we evaluate the effectiveness of this approach in managing city-scale applications, thereby establishing a robust foundation for next-generation smart city digital twins.

The key contributions of this paper are as follows.

1) A comprehensive framework is introduced for Urban Digital Twins that allows implementing advanced digital twin functionalities.
2) An AI-driven framework is incorporated into the architecture that adjusts prediction models based on real-time urban dynamics and changing conditions.
3) The developed DT architecture supports generalizability, enabling its application across various smart city domains.
4) The feasibility of the proposed methodology is validated through a case study of real-time air quality monitoring using an IoT network deployed in Madrid.

II. Related Work

In recent years, urban Digital Twin implementations have gained global attention, with notable examples in Helsinki, Finland [4], Rennes, France [5], Berlin, Germany [6], Florence, Italy [7], Singapore[1] and Victoria, Australia[2]. However, these solutions are typically designed for specific use cases, limiting their adaptability to broader smart city applications, and most of these DT implementations primarily focus on 3D virtual modeling rather than leveraging key digital twin functionalities, such as real-time prediction, dynamic simulation, and prescriptive analytics.

While the visualization of city characteristics enhances urban connectivity, the prediction and simulation functionalities of

[1] https://oecd-opsi.org/innovations/virtual-twin-singapore/
[2] https://www.land.vic.gov.au/maps-and-spatial/digital-twin-victoria

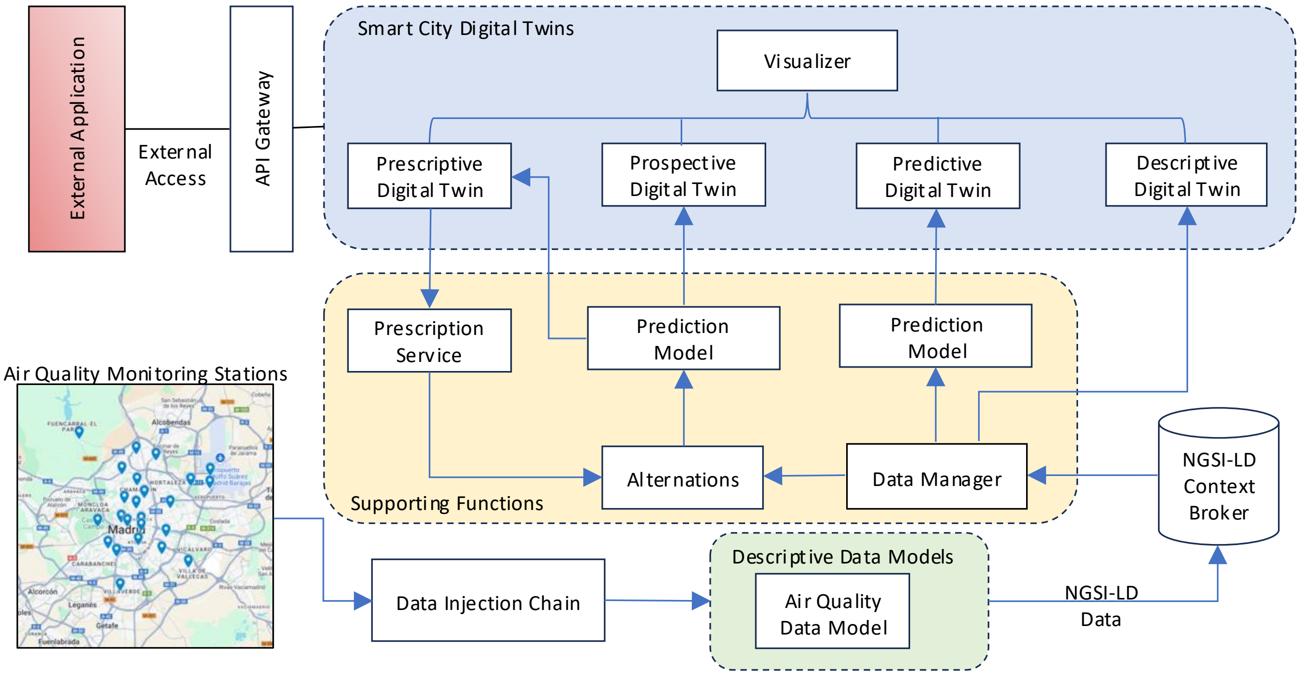

Fig. 1: Overview of the System Architecture

a digital twin further optimize city operations through proactive management. Both prediction and simulation rely heavily on time-series data. In this context, Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have emerged as state-of-the-art models for time-series prediction. Although Gated Recurrent Units (GRUs) are also employed, LSTM networks often demonstrate superior predictive performance [8]. Prior studies have validated the effectiveness of LSTM models in urban prediction tasks. For instance, combining traffic and noise pollution data has been shown to improve prediction accuracy [9], while incorporating surrounding traffic flow data can further enhance model performance [10].

In contrast, existing solutions face several key limitations, including limited functionality stemming from a primary focus on virtual 3D visualizations, and restricted generalizability due to application-specific designs. To address these challenges, we propose a comprehensive urban digital twin framework capable of supporting diverse digital twin functionalities, demonstrated through an air quality monitoring use case in the city of Madrid.

III. System Architecture

In this section, we describe the proposed framework for urban Digital Twins. As shown in Figure 1, the proposed architecture implements four types of digital twins: descriptive, predictive, prospective, and prescriptive, each with unique functionalities. An architectural view of the proposal and the components involved in the different types of digital twins for smart city air quality management is detailed below.

A. *Descriptive Digital Twin*

The Descriptive DT captures the smart city's current and historical states, encompassing static and dynamic characteristics. It provides a real-time snapshot by accessing air quality data collected from IoT sensors deployed across Madrid[3]. At the data injection chain, incoming data are preprocessed to validate the time-series consistency and ensure sufficient information before being transformed into the Next Generation Service Interface-Linked Data (NGSI-LD) format[4]. The data are then injected into the context broker using the "Air Quality Observed" data model from Smart Data Models[5]. It is important to note that the framework is designed so that it could support various types of smart city applications, such as traffic management. These data can be connected to the proposed architecture by leveraging appropriate smart data models. For instance, for traffic information "Traffic Flow Observed" can be utilized at the Descriptive Data Models module. All descriptive data are stored in Stellio[6], an open-source, NGSI–LD–compliant context broker built on linked-data principles, which provides a standardized API. The Descriptive DT also supports time-based filtering, aggregation, and other augmentation functions on time-series data.

B. *Predictive Digital Twin*

The Predictive DT extends the descriptive DT by estimating future states based on historical and current data of the smart

[3]https://datos.madrid.es/portal/site/egob/
[4]https://www.etsi.org/committee/cim
[5]https://smartdatamodels.org/
[6]https://stellio.readthedocs.io/en/latest/

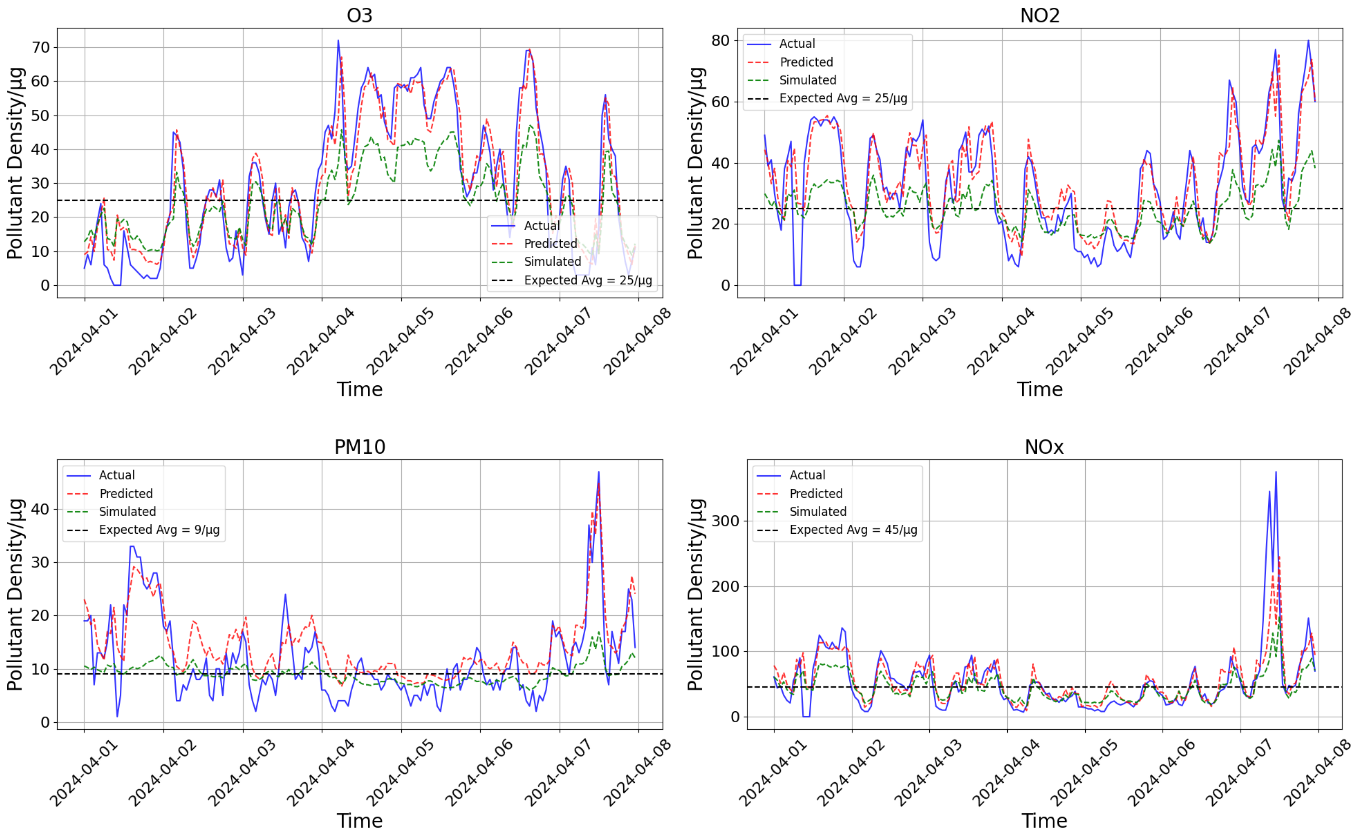

Fig. 2: Actual, Predicted and Simulated Air Quality Data for 28079008 Air Monitoring Station Different Pollutants

city. These predicted values are generated using prediction models in the supporting functions module. In our implementation, predictive AI models are dynamically trained using the data from air quality monitoring stations. The predictions are generated on demand, and they can be either stored in the context broker or provided to external applications.

C. *Prospective Digital Twin*

The Prospective DT extends predictive capabilities by enabling "what-if" analysis. This is done by modifying the current state to simulate intended actions, and then executing predictions on this altered state. It allows alterations to the air quality data, and simulations are performed using the prediction AI models.

D. *Prescriptive Digital Twin*

The Prescriptive DT determines the actions required to achieve a desired target state. It performs iterative "what-if" analyses on air quality data to evaluate the impact of potential alterations. This twin leverages a prescription service, which proposes successive adjustments, and utilizes AI-based prediction models to assess and guide the system toward the intended outcomes.

Moreover, the API gateway enables external applications and users to access the smart city DTs. We adopted the open-source FastAPI library[7] to implement a RESTful interface for seamless interaction with the DTs. The data manager plays a critical role in ensuring smooth data flow and facilitating control information exchange between the DTs and the context broker. It communicates with the context broker via its REST API to access and manage stored data. During real-time monitoring, prediction, simulation, and prescriptive analysis processes, the data manager retrieves the necessary air concentration data to support accurate forecasting and analysis. It also handles data retrieval tasks required for LSTM model training.

The proposed system offers a holistic framework that significantly enhances both functionality and generalizability. In contrast to existing architectures, it streamlines development processes and places greater emphasis on digital twin modeling, making it particularly well-suited for smart city applications.

IV. Experiments and Results

The urban digital twin implementation was validated using hourly air quality data from Madrid[8], collected between January and April 2024. The learning models used in predictive DT and prospective DT has the specifications of layer 1 units = 150, layer 2 units = 100, dropout rate =0.03, learning rate = 0.01, time steps = 24, epochs = 100 with early stopping, batch

[7]https://fastapi.tiangolo.com/

[8]https://datos.madrid.es/portal/site/egob/

size = 32, activation function ReLU, Optimizer = Adam, Loss function = MSE. The experiments evaluate the performance of descriptive, predictive, prospective, and prescriptive DTs for air quality data. To achieve this, we conducted comprehensive testing using LSTM models trained on pollutant concentration readings from air quality monitoring stations and recorded their outputs and error metrics.

TABLE I: Prescriptive DT Recommendations for Air Pollutant Reduction to Achieve Target Weekly Averages (April 1–8)

Pollutant	Alternation Required for the Current State	Expected Average ug/m^3
O3	Reduction of 35%	25
NO2	Reduction of 39.5%	25
PM10	Reduction of 49.5 %	9
NOx	Reduction of 31.5%	45

TABLE II: Performance of the Prediction Model based on MAE, MSE, and RMSE for different Air Monitoring Stations

Air Monitoring Station	Pollutant	MAE	MSE	RMSE
28079004	NOx	0.0307	0.0022	0.0472
	CO	0.0270	0.0013	0.0355
	NO	0.0304	0.0026	0.0506
	NO2	0.0503	0.0046	0.0675
28079008	NOx	0.0145	0.0008	0.0276
	CO	0.0084	0.0002	0.0132
	NO	0.0100	0.0006	0.0241
	NO2	0.0628	0.0067	0.0817
	PM10	0.0235	0.0009	0.0299
	O3	0.0707	0.0079	0.0891
	SO2	0.0019	0.0000	0.0028
	TOL	0.0230	0.0014	0.0375
	EBE	0.0092	0.0036	0.0597
	PM25	0.0459	0.0033	0.0577
28079017	NOx	0.0260	0.0018	0.0418
	CO	0.0260	0.0016	0.0403
	NO	0.0260	0.0052	0.0724
	NO2	0.0621	0.0068	0.0824
28079035	NOx	0.0688	0.0129	0.1138
	CO	0.0504	0.0046	0.0678
	NO	0.0432	0.0055	0.0739
	NO2	0.1072	0.0168	0.1295
	O3	0.0765	0.0118	0.1085
	SO2	0.0063	0.0001	0.0078

Figure 2 presents the actual, predicted, and simulated pollutant concentrations—corresponding to the outputs of the Descriptive, Predictive, Prospective, and Prescriptive DTs, respectively—for O_3, NO_2, PM_10, and NO_x at the air quality monitoring station 28079008. The results demonstrate that the predicted values are closer to the observed data across all pollutants. This is because the fact that the temporal dependency-capturing capability of LSTM networks effectively models pollutant dispersion patterns.

Table I outlines the recommendations provided by the prescriptive DT to reduce pollutant levels to meet target average values for the week of April 1 to April 8. According to the results, to achieve a target of 25µg/m³ for O3, a reduction of 35% should be done on the data. For NO2, reaching 25µg/m³ necessitates a 39.5% decrease. PM10 must be reduced by 49.5% to meet a 9µg/m³ target, and NOx requires a 31.5% reduction to reach an average of 45µg/m³. The simulated values in Figure 2 show the Prospective DT's output when implementing the Prescriptive DT's recommendations. This same prescriptive approach can be applied to determine actionable changes required to achieve other target environmental states.

Table II presents the prediction errors, quantified using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The observed trend of improved Predictive DT performance across the air quality monitoring stations is further validated by these error metrics, reinforcing the efficacy of the approach.

V. Conclusions

In this paper, we propose a methodology to implement different types of smart city digital twins. Our framework addresses key limitations of existing implementations, including application specificity and functionality constraints. The methodology has been validated through the use case of air quality management for the city of Madrid. The results demonstrate the advanced functionality of the integrated digital twins for air quality management. Future work will focus on integrating additional smart city applications, incorporating data from varied sensing modalities, and refining model generalization to further enhance the broad applicability of digital twins in smart city environments.

References

[1] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the iot context: A survey on technical features, scenarios, and architectural models," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020.

[2] X. He, Q. Ai, J. Wang, F. Tao, B. Pan, R. Qiu, and B. Yang, "Situation awareness of energy internet of things in smart city based on digital twin: From digitization to informatization," *IEEE Internet of Things Journal*, vol. 10, no. 9, pp. 7439–7458, 2022.

[3] M. Jafari, A. Kavousi-Fard, T. Chen, and M. Karimi, "A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future," *IEEE Access*, vol. 11, pp. 17 471–17 484, 2023.

[4] Helsinki digital twin. [Online]. Available: https://kartta.hel.fi/?setlanguage=en\#,andhttps://kartta.hel.fi/3d/\#/

[5] Rennes en 3d: Rennes 2030 la vill à votre idée. [Online]. Available: https://rennes2030.fr/rennes-en-3d/

[6] 3d city db. [Online]. Available: https://www.3dcitydb.org/3dcitydb/visualizationberlin/

[7] L. Adreani, P. Bellini, M. Fanfani, P. Nesi, and G. Pantaleo, "Smart city digital twin framework for real-time multi-data integration and wide public distribution," *IEEE Access*, 2024.

[8] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*. IEEE, 2016, pp. 324–328.

[9] F. M. Awan, R. Minerva, and N. Crespi, "Using noise pollution data for traffic prediction in smart cities: experiments based on lstm recurrent neural networks," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20 722–20 729, 2021.

[10] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, D. Alahakoon, and D. Pothuhera, "Online incremental machine learning platform for big data-driven smart traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679–4690, 2019.