OUTFIT: Crowdsourced Data Feeding Noise Maps in Digital Twins

Elena Ascari
Institute for Chemical-Phisical Processes, CNR
Italy
elena.ascari@cnr.it

Mauro Cerchiai
Institute for Chemical-Phisical Processes, CNR
Italy
m.cerchiai@arpat.toscana.it

Pasquale Gorrasi
Institute for Chemical-Phisical Processes, CNR
Italy
pasquale.gorrasi@pi.ipcf.cnr.it

Antonella Longo
Dept. of Engineering for Innovation
University of Salento
Italy
antonella.longo@unisalento.it
Italian Cen. on Supercomputing(ICSC)
Casalecchio di Reno, Italy

Gabriele Mencagli
Dept. of Computer Science
University of Pisa
Italy
gabriele.mencagli@unipi.it

Francesca Miccoli
Dept. of Engineering for Innovation
University of Salento
Italy
francesca.miccoli@unisalento.it

Marco Zappatore
Dept. of Engineering for Innovation
University of Salento
Italy
marcosalvatore.zappatore@unisalento.it
Italian Cen. on Supercomputing(ICSC)
Casalecchio di Reno, Italy

Abstract

In the realm of urban management, Digital Twins (DTs) have recently shown their potential to improve planning and sustainability. The OUTFIT PRIN 2022 project aims to optimize data streams to dynamically render Road Traffic Noise (RTN) in an urban DT model, incorporating both noise levels and citizens' perceptions. In this paper, we introduce OUTFIT and propose the methodology aimed at providing a set of tools to assist policymakers in addressing noise issues and promoting actions for improving the well-being of the citizenship. The project, aligned with Mission 1 of the Italian National Recovery and Resilience Plan (PNRR) and Horizon Europe priorities, focuses on mobility, energy, urban infrastructure, circular economy, and behavioral change. OUTFIT also aims at building a traffic-related database from crowdsourced data for developing a reliable RTN model input to start. This follows an optimization of the data streams for dynamic traffic data processing and 3D noise rendering in order to create a DT model which integrates noise, traffic, and complaints data, with a particular focus on efficient monitoring and orchestration of edge resources. In addition to the definition of a validated method for deriving traffic flows and RTN, a 3D DT with dynamic noise rendering and a set of APIs to enable the interoperability of OUTFIT system's open data will be developed as well.

CCS Concepts: • **Information systems → Decision support systems**; **Online analytical processing**.

Keywords: Digital Twins, Crowd-sourced Data, Noise Maps

ACM Reference Format:
Elena Ascari, Mauro Cerchiai, Pasquale Gorrasi, Antonella Longo, Gabriele Mencagli, Francesca Miccoli, and Marco Zappatore. 2024. OUTFIT: Crowdsourced Data Feeding Noise Maps in Digital Twins. In *Workshop on Flexible Resource and Application Management on the Edge (FRAME '24), June 3–7, 2024, Pisa, Italy*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3659994.3660312

1 Introduction

A *digital twin* [16] (DT) is a complete virtual description of a physical entity. Nowadays, it represents an innovative and powerful technology to support decision-making processes in different application domains such as industry or Smart Cities. Specifically, urban DTs (UDTs) represent an increasingly effective way of supporting crucial urban management

FRAME '24, June 3–7, 2024, Pisa, Italy

ACM ISBN 979-8-4007-0641-7/24/06
https://doi.org/10.1145/3659994.3660312

activities in a Smart City, such as asset control optimization, city planning, and sustainability-boosting interventions [10]. A relevant input when building and managing a DT is given by crowd-sourced big data streams, which day by day are becoming more and more available thanks to the pervasiveness of IoT sensing devices. In addition, a scenario that has experienced a significant expansion over the recent years is represented by highly-decentralized environments exhibiting distributed computational capabilities provided by edge devices that need to communicate effectively with cloud resources across the so-called Cloud-Edge Continuum, by dynamically depending on the resources, applications, and constraints of the distributed cloud-edge infrastructure [33]. For instance, digital buildings equipped with multiple IoT sensors represents a typical edge-cloud continuum scenario, where smart sensors at the edge layer gather data and interact with the cloud layer to exploit considerably higher computational capabilities [31].

Moreover, there sources can be further combined with social networks, which allow collecting citizens' perceptions about urban areas, so to drive policies and then improve the promptness and effectiveness of political and regulatory actions, including even the crowd's feedback.

Despite the huge potential impact, however, crowd-sourced data useful to build and update DTs need careful collection and processing. The OUTFIT approach, which is the main contribution of this paper, aims at understanding how crowd-sourced data can be profitably used to build a DT of a city area in order to continuously report the level of noise pollution perceived by citizens due to the actual traffic volume.

The traditional approach to deliver such an analysis is to derive traffic data required by EU noise pollution models [15] implemented in commercial software, which are mandatory for a large-scale representation of cities' complex geometries. Numeric simulations are run at a given value (e.g., mean, peak) without evaluating time evolution. Such algorithms are usually fed by expensive measurements taken by long-lasting campaigns based on manually installed measurement devices. Our ambition is to do that without such expensive measurement processes, but rather by relying on continuous data streams of crowd-sourced information only.

OUTFIT also aims at implementing an UDT to enable dynamic *Road Traffic Noise* (RTN) estimation based on crowd-sourced data, thus developing a methodology able to rely on live-timing changes thanks to Data Stream Processing [2] (DSP) and High-Performance Computing (HPC) approaches. This allows the study of changes in pollution models and hence of RTN maps to be updated in terms of small time intervals (e.g., a few minutes).

More specifically, OUTFIT builds a database of crowd-sourced traffic data, derives its flows by equations, and gives them as input to the RTN model. The RTN 3D model is developed by means of an attenuation matrix from commercial software and correction factors to be applied to real-time input data derived from Data Stream Processing pipelines. The RTN model derives the noise in the 3D space based on European shared method (CNOSSOS-EU) and noise levels can be estimated at different points and heights of the selected area. The output 3D map is then presented in a 3D UDT, providing a dynamic view of RTN based on real-time crowd-sourced data. DSP and HPC approaches are applied to the collection and processing of the crowd-sourced input data stream and the rendering of the environmental model output. The UDT does not only visualize noise levels but also citizens' complaints retrieved from social networks and from a specific OUTFIT mobile app. Feedback data are delivered to the UDT to drive policies, i.e., offering further perspectives other than quantitative levels.

Last but not least, OUTFIT allows the optimization of traffic data and social data in an UDT that could fit various environmental models, e.g., it can be even used for air pollution models and a site-independent method for RTN, allowing maps based on crowd-sourced data and derived correction factors. Therefore, the potential implication of the approach is large and with high value.

2 Background and Motivation

Mitigating climate change at a global scale is a huge challenge. In Europe, it is estimated that by 2050 almost 85% of Europeans will be living in cities. Among several aspects of urban life that may contribute to the increase in health-related issues, RTN is a growing concern and it has been recognized by the World Health Organization as the second most significant environmental stressor, only after air pollution. Nearly 113 million Europeans are affected by RTN levels of at least 55 dBA [28], which is risky to health. Thus, the reduction of the urban RTN has become mandatory.

The EU provided the Directive 2002/49/EC [7] to guide a noise policy in member states, so offering a common framework for noise mapping. It obliged road owners and public administrations to update noise maps every 5 years. Twenty years later, the involvement of citizens is pivotal for the success of noise reduction. Indeed, experts usually estimate RTN by long-lasting measurement campaigns, official RTN models implemented in commercial software applications, and requiring a huge number of inputs, particularly traffic flows. Thus, collecting data with low-cost sensors (LCS) might increase the number of points for estimating RTN. Recently, LCS networks in several cities allowed estimating noise in a capillary way [3, 6]. Some projects provided cities with a real-time update of noise maps based on LCS [5].

Other approaches adopt citizens-owned smartphones as moving LCS [12, 29]. In this way, citizens are involved directly in the collection of acoustic measurements aimed at modelling the noise of the urban areas where they reside and live. However, since RTN is too largely distributed to be mapped only by smartphones, other studies took advantage of Web-accessible API services providing travel times along routes [34, 35]. These data might gather traffic flows and noise but need to be carefully processed to build reliable flows as input of standard noise models, especially to develop a dynamic time-scale model to be incrementally updated with data streams [19].

Together with traffic data, crowd-sourced data related to noise issues are also available (e.g., popular times, traffic jams, complaints) but they have been typically considered as a separate source for noise studies [4, 17]. In other studies, noise levels have been collected together with the perception of noise in mobile crowd sensing projects [20], sometimes using citizen science approaches [21].

Starting from such premises, our approach with OUTFIT develops a novel model to collect and process heterogeneous crowd-sourced data to combine quantitative and measurable aspects with citizens' perception. The resulting dynamic noise maps and social data streams will be presented in an UDT in order to enhance Geographical Information System (GIS) capabilities, as they are virtual replicas of real-world assets or natural systems integrating real-time data to make it usable for city stakeholders. Therefore, DTs and UDTs might capture statuses, monitor performance, predict future outcomes, and also allow a representation of current, past, or even future asset conditions. Overall, DTs are promising planning solutions to show scenarios as well as to collect and visualize the user's feedback, and to provide immersive views of different data sources [14, 37]. In particular, urban observatories are nowadays investigating UDTs as valuable enablers for supporting effective and sustainable urban planning, fostering and improving environmental awareness positive behavioral patterns in citizens, monitoring natural and anthropic events in the urban landscape.

3 Overview of the Approach

OUTFIT addresses the processing of crowd-sourced streams to provide a dynamic rendering in an UDT model of RTN obtained from both the noise levels and citizens' perception to assist decision-makers to help them in optimizing citizen well-being. We consider noise as another layer of the UDT in order to reduce the gap between citizens and policymakers. These general goals will be put into practice through the realization of the following objectives: (1) creating a database of noise generated by vehicular traffic from crowd-sourced data complemented with social data about how the noise is perceived by citizens; (2) setting up of a new approach to prepare crowd-sourced big data into reliable inputs of the RTN model; (3) optimizing data streams for dynamic traffic data processing and rendering of noise; (4) setting up of an UDT model with noise, traffic, and complaints data.

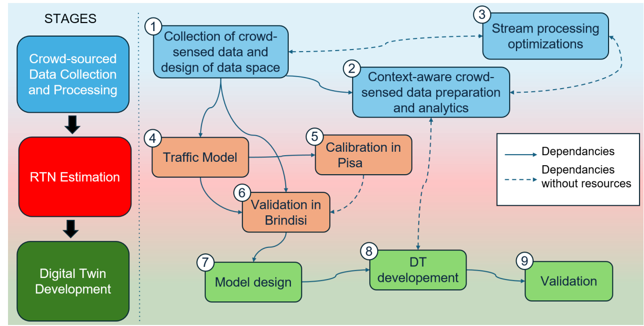

Figure 1. Abstract overview of OUTFIT and its methodological stages.

To accomplish these objectives, OUTFIT is organized into three methodological stages comprising a set of computational techniques and proper software tools. The first is related to crowd-sourced data collection and processing, along with effective edge resource monitoring; the second stage is aimed at providing a continuous RTN estimation; the third stage encompasses the UDT development, which is expected to be enriched with an adaptive management of applications in the edge-cloud continuum. Figure 1 depicts such methodological stages. In Figure 2, the system architecture in the large is proposed. In the upper layer, the addressed stakeholders (i.e., organizations, public authorities, citizens, enterprises, schools, etc.) are listed. The central layer comprises the functional components that will cover the data-driven processing pipeline up to the final presentation of the system output (also via Open Data APIs). In the bottom layer, all the envisioned data sources are encompassed, ranging from traffic data coming from sensing stations to institutional data and open data repositories.

4 Crowd-sourced Data Collection and Processing

Collecting data from traffic-analysis techniques and processing methodologies is crucial in order to provide a coherent and rich set of data streams conveying crowd-sensed information and high-level knowledge extracted in real time from such a continuous flow. We analyze in this section the specific aspects covered by this stage.

4.1 Details of the data sources, collection of crowd-sensed data, and design of the data space

We conducted a comprehensive study of the state-of-the-art related to different data sources and data models available for monitoring vehicular traffic, also in order to collect available data about complaints by citizens for excessive noise pollution [30]. Consequently, in OUTFIT we propose a data space [27] designed to gather data from other available sources, as well as from previous projects and open data repositories,

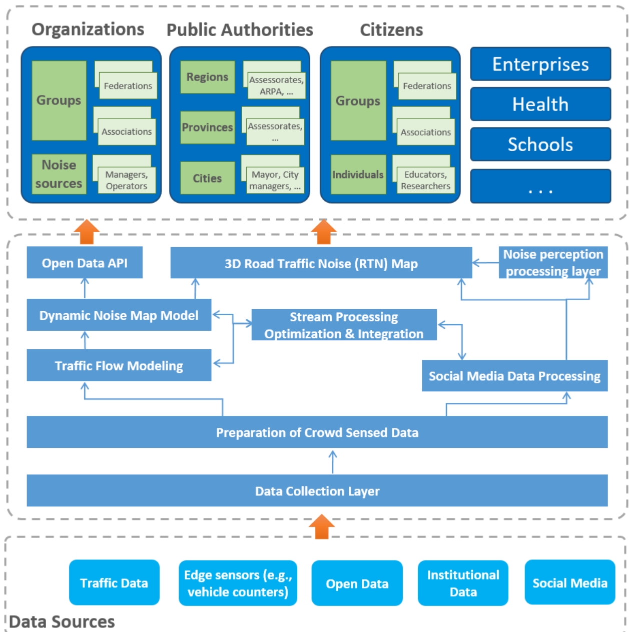

Figure 2. OUTFIT: overall architecture.

such as the Newcastle Urban Observatory (NUO) [13]. Similarly, existing cloud-services and apps (e.g., Waze or Google Directions APIs) will be also considered for gathering traffic monitoring information in near-/real-time. In addition, social data about the perception of noise from citizens will be collected from various social media (e.g., X, Waze, etc.). Overall, the following data sources will be included in OUTFIT (as also indicated in Figure 2):

- Real-time traffic data: mainly provided by traffic counters and other IoT sensors deployed across the city.
- Near-time traffic data: mainly collected via Google Directions APIs and other similar cloud services, according to the sampling frequency specified by the service provider.
- Social media data: posts and comments related to noise-related complaints from citizens, crawled from API-accessible social media.
- Legacy, open, and/or institutional datasets: any other noise-related and traffic-related accessible dataset made available by open data providers and by public institutions, associated with the urban area considered. Also historical and legacy datasets about previous traffic monitoring campaigns are considered.

A dedicated mobile app will be also designed and released, based on Telegram, for collecting sentiments and opinions of users. Overall, developing a data space is crucial for two reasons: first, it makes the system capable of integrating crowd-sourced data with social data to enhance the accuracy of traffic-related data; second, it supports real-time and near-/real-time management of traffic-noise models.

4.2 Context-aware crowd-sensed data preparation and analytics

Due to the intrinsic nature of the heterogeneous data sources listed in the previous section, a data lake architecture [32]is needed, as it behaves as a centralized repository allowing the storage of structured and unstructured data at any scale. Therefore, we conducted an exploratory data analysis to investigate the main features, the available data and metadata, and the potential sources of data enrichment, in order to properly contextualize the data lake. This was performed by using both visual and analytical methods, based on Python libraries. According to the provenance and velocity of traffic and social data, different pipelines for improving each source's quality and enriching them will then be designed and implemented also considering that traffic data is usually contextualized thanks to the collection of travel times, popular times, and road characteristics. For instance, the pipelines will range from data preprocessing (e.g., ensuring the normalization of timestamp formats and geographic coordinates between the data collected from vehicle counting devices and traffic density data fetched via Google APIs) to data integration (e.g., aligning all datasets to a common schema suitable to be used for analytics purposes), and quality checks (e.g., traffic volumes do not exceed road estimated capacity, estimated noise values are within expected ranges), In addition, sentiment analysis will consider comments about noise complaints, traffic jams, and formal complaints from the city's police department. Social media data will be treated in order to comply with privacy and data protection regulations. Proper processes will be also designed and developed to prepare data for input in the traffic model system and for sentiment analytics. A dedicated database will be created to store enriched data, containing both traffic and noise data streams, as well as sentiment data. Finally, OUTFIT data will be made available in a FAIR (Findable, Accessible, Interoperable, reusable) format [38], as required by the open science guidelines [8] and made available through an API interface.

4.3 Stream processing optimizations

The third phase is devoted to the study of DSP and HPC approaches to process data streams of crowd-sensed information, both raw streams from the first stage and from the knowledge extraction logical pipelines within the second stage. In such a scenario, parallelization techniques are required to leverage the underlying parallel hardware (e.g., multicores, GPUs, and FPGAs) in a profitable way to meet real-time requirements. This implies not only the adoption of costly and powerful traditional HPC resources (e.g., clusters or Clouds) but rather the adoption of devices with limited computational capabilities like Edge resources, which are still

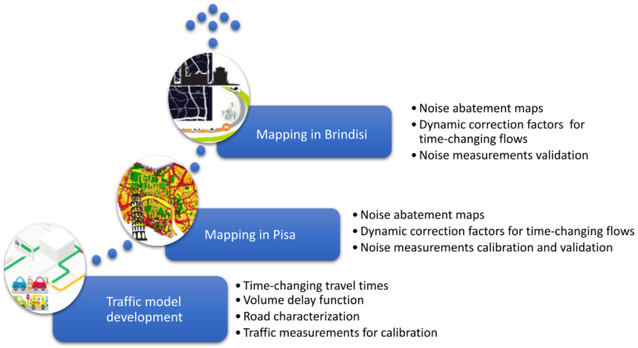

Figure 3. Towards a 3D noise map with time-changing input.

equipped with CPUs and hardware accelerators with moderate parallelism. To that purpose, we have already developed a set of tools for analyzing streams by running efficiently on low-end devices streaming tasks such as filtering, and aggregation with large and frequent temporal windows, joins, and correlations, and stream mining algorithms on multicores, GPUs, and FPGAs. The library will be adopted to develop the data processing pipelines of OUTFIT, to enhance the quality of raw data gathered from the data sources in Sect. 4.1. Special emphasis should be taken in enabling such kind of analysis in Edge-Cloud continuum resources by leveraging efficient sketching of data streams to reduce the memory footprint and using Key-Value Stores efficiently to cope with large-than-memory state scenarios, where the amount of historical data to be kept in the pipelines for stateful processing exceeds the available device memory. The complete tool chain is available as an open-source library on GitHub [23].

5 RTN Estimation

According to the methodological stages introduced in Figure 1, the first stage provides a collection of crowd-sourced data, structured into databases that are the input of the second stage aimed at elaborating a 3D noise time-changing information. This section details the steps needed to develop the RTN in an urban context: *1)* traffic model development; *2)* mapping and calibration in a first urban area (i.e., Pisa municipality, Central Italy; *3)* mapping test and validation in a second urban area (i.e., Brindisi municipality, Southern Italy. Figure 3 explains achievements expected for each step.

5.1 Traffic model development

At first, a new method for traffic flow estimation from crowd-sourced data is to be obtained. The innovative algorithm must use different and accessible Big Data to predict traffic volumes. In particular, the new idea combines infrastructure-specific characteristics extracted from Open Street Map [26] with data extracted from social media and web services (such as travel times of road links and popular times of attraction poles). This step is crucial because geometric characteristics and road intersection types influence relationships between travel times and on-site traffic volumes [19]. In light of this fact, an initial clustering of the different types of road links based on their peculiarities and the type of intersections (one-lane, two-lane roads, presence or not of traffic light systems, and roundabouts) is carried out. Subsequently, crowd-sourced data (e.g., travel times and popular times) are used to estimate traffic volumes through appropriate link delay functions. These models express the travel time or the speed on a road link as a function of traffic volumes [1],[25],[22],[24]. Each function is particularly suitable to a particular type of road link. The BPR (Bureau of Public Roads) model [36], characterized by a simple form, is suitable for links not subject to congestion conditions and without traffic lights. Instead, the Akçelik model [1] is appropriate for modeling traffic conditions with intersections and signalized links.Thus, calibration will also be dedicated to the analysis of roads and intersections to verify which algorithm should be applied.

Crowd-sourced data enables calculating traffic flows at the time which data refers to, so potentially at every time interval for which data are extracted (ideally 15 minutes). The use of popular time might allow the evaluation of noise time evolution during the day even for the periods in which travel times are not reliable. In fact, when traffic flow is too low, people tend to drive at their free speed and the travel times are no longer correlated to the flows. Traffic models based on transportation theories tend to fail for small roads: popular times can contribute to modulating noise estimation, especially for areas with low traffic. Therefore, firstly an approach based on road and areas' characteristics which do not depend on the testing area will be established. Then, this approach will be tested in the case study in step 2 and validated in step 3. An iterative method will be then used in defining and testing the approach in steps 1 and 2.

5.2 Mapping and calibration

The second step focuses on the acquisition and processing of traffic flows data in a testing area in Pisa (Central Italy) to calibrate the traffic noise model. The iterative mapping in Pisa allows the definition of the traffic flow deriving method and the assessment of noise on building floors. The area will be modeled within commercial software for noise mapping only to derive abatement of noise due to obstacles and environment which do not change as a function of the traffic flows. Then, the time evolution of estimated noise can be calculated by means of corrective factors based on noise emission model equations [9]. Practically corrective equations will be established as a function of vehicle flows and road type in order to establish changes in noise emission.

This preliminary test is necessary for the calibration of the method to derive flows from Web data since an initial setup has to be performed using measurements as a reference. Thus, the parameters of transportation equations need to be refined by minimizing differences in terms of estimated total

flows and measured ones. To this aim a specific measurement campaign is performed in Pisa, including on-site traffic flows by means of traffic counters. In addition to traditional methods, on-site traffic monitoring is detected with innovative systems for automated vehicle recognition based on video recordings with low-cost digital camera sensors [11]. The system acquires data with low-cost sensors and by using Artificial Intelligence techniques based on Neural Networks can detect, track, and classify the vehicle flows. It can also evaluate the speed and distinguish between different vehicle categories as required by CNOSSOS-EU.
In this calibration phase also RTN will be monitored and overall model will be calibrated and adjusted against measurements such that accuracy will be within 3 dB(A).

5.3 Mapping test and validation

In the third step, the proposed method is applied in another testing area in Brindisi (Southern Italy). In this second test area, no iterative process will be needed since the new method intends to apply to the new areas only using crowd-sourced data (road characteristics, travel times, popular times) without direct traffic measurements. Therefore, a new time-changing 3D noise map is set up. A further noise measurement campaign will be carried out in the Brindisi municipality to validate results comparing measured and estimated levels (dB(A)) and demonstrate the replicability of the method for noise estimation, based only on crowd-sourced data. In addition, collected measurements will be also considered in conjunction with: *a)* feedback coming from the second validation step provided by the developed social mobile app, and *b)* (possibly) from additional data streams coming from smart sensors deployed as edge devices.

In Brindisi, the noise will be estimated for different time intervals and at different heights (floors). These layers of noise will be used as input to the UDT.

6 Digital Twin Development

The conclusive stage refers to the design and develpoment of the UDT related to the Brindisi testing area. This digital representation will visualize dynamic noise level maps at different heights, enriched with collected analytics about citizens' perceptions through social media. Dedicated indicators to support mitigation actions will also be visualized to support policy-makers and involved stakeholders in the decision-making process. HPC methods will be used to improve the performance of data stream visualization (see subsection 6.2). This section briefly outlines the steps needed to design, build, and validate the UDT.

6.1 Design the data model

Since DTs virtually represent physical entities along with their relationships, processes, and behaviors, they must rely on appropriate data models which data collected from their physical counterparts must comply with. In addition, reports, analytics, user experiences, and feedback make a DT truly effective. Therefore, at first, the definition of a data model suitable for the scenario addressed in OUTFIT is addressed. Starting from the state of the art related to data representation of DTs for urban spaces (or UDTs), the corresponding challenges are addressed as well.

The UDT data model manages several types of data. First, data coming from the twinned physical counterpart: they provide spatial and temporal snapshots and refer to traffic flows, acoustic measurements, and user-collected and sensor-gathered measurements. Second, data providing the UDT domain knowledge in terms of building-related and noise-related standards and guidelines, as well as of noise abatement rules and interventions, e.g., noise level limits approved by local authorities, plans, and interventions already programmed or in force. The third kind of data comes from the virtual twin in the form of simulation parameters, settings, outputs, and predictions, e.g., the 3D noise dynamic map of RTN. Data from physical measurements and data from digital predictions are also to be correlated, thus enabling the final stage of data fusion between them (e.g., a spatial-temporal data fusion where traffic data are integrated with geo-referenced social media data so that spatial analysis can be used to map these data onto the city geographical layout and temporal analysis can be used to align data by their timestamps., which makes the UDT capable of predicting new statuses for the entire system as well as to provide dynamic responses to specific emerging noise issues.

6.2 Application building

From the data obtained in the RTN estimation, a 3D noise dynamic map on the UDT model will be developed for the Brindisi area. A dedicated model is being defined, including output layers of RTN estimation in the Salento geographical region. The rendering of the environmental model output will leverage HPC computing frameworks (e.g., openMP, Intel TBB, and CUDA) that are used to develop specific components accelerating the computing tasks that appends value-added information of the model output into the 3D representation of the UDT. The main purpose of these software components is to represent different outputs from the environmental model, not concerning noise pollution only.

The UDT developed in OUTFIT is based on a recent cutting-edge approach, according to which a massive multiplayer game engine is fed with data sources and enriched with additional features . More specifically, the Unreal Engine v5 (UE5) will be used. First, the 3D model of the urban area considered will be imported via the UE5's Datasmith feature. Then, the live data coming from the sources mentioned in the previous sections will be retrieved as JSON files via RESTful APIs. Finally, a dedicated UI, tailored to the requirements elicited from the envisioned stakeholders will be developed, in order to enable a smooth integration with the UDT and an effective visualization of metrics and KPIs, without requiring the stakeholders any prior knowledge on UE5.

6.3 Noise perception layer development and second validation

Finally, an alert system for possible mitigation actions will be developed and included in the OUTFIT UDT. A noise annoyance-based indicator will be derived from spatially aggregated average noise levels estimated in RTN estimation and visualized as an additional layer [11], [18]. However, noise annoyance is known to be only a part of disturbance since it is not able to take into account noise variability. Therefore, also any social complaints collected are expected to be included in the UDT in order to point out areas where even the population perceives a noise disturbance and already asks for solutions. This will allow policymakers to establish priorities in an easier and faster way, enhancing the implementation of effective mitigation action.

The validation of the developed OUTFIT approach will be carried out in the Brindisi municipality. In the second validation, traffic data collected for feeding the traffic model and stakeholders will be used in providing social media and noise perception measures collected through the OUTFIT app. The envisaged use case to support this scenario involves local schools that will introduce students to act as citizen scientists and perform noise measurements according to ad-hoc guidelines, to provide their perception feedback as well.

6.4 Further considerations about the Edge-Cloud continuum

In the previous sections, we discussed how a proper and fruitful combination of data coming from noise smart sensors and the UDT representing the urban area under test is crucial in OUTFIT. This integration harnesses the strengths of edge computing devices and traditional cloud-computing frameworks. Edge devices are positioned strategically across the city and serve as crucial data-collection points capturing real-time noise data from their locations. Concurrently, the cloud-based UDT aggregates incoming data possibly with additional data sources such as traffic patterns from Google Directions APIs and local weather conditions. This is a typical edge-cloud continuum scenario, where smart sensors equipped with local processing capabilities can be deployed at the edge layer, in order to decentralize preliminary data filtering and basic analytics on noise-level measurements, so to minimize latency and speed up the responsiveness of the system to noise-related events, which could not be handled as efficiently as in fully cloud-centric model.

Further development will be to investigate strategies of dynamic resource allocation and scalability at the edge-cloud continuum with the aim of ensuring efficient data processing. Moreover, we plan to leverage the loop between edge devices and the cloud-based UDT so that once the data are received from the edge devices, the corresponding data-driven insights derived at the cloud layer are back-propagated to the edge devices. This could allow us to expand the possibilities of the proposed approach with specific features for localized responses and more rapid adaptive noise management strategies. For instance, during rush hours, edge devices can detect traffic congestion levels in a given area, preliminary estimate corresponding noise levels, send pre-processed data to the cloud-based UDT, where more complex interventions tailored to the area monitored (e.g., traffic light sequences adjustment, real-time notifications to drivers for alternative routes via the OUTFIT mobile app) can be triggered.

7 Conclusions

This work outlined the main activities and goals of the OUTFIT. It explores the use of crowd-sensing resources and elaborates them with a general approach that can be useful not only for road traffic noise estimates. It proposes a method for dynamic modeling which is not based on noise acquisitions but on source modeling, being more useful for mitigation planning. In the field of HPC, it applies data stream processing to smart city applications, enhancing calculation and rendering capabilities. In the field of UDT, it develops an integration between different data sources and visualizes source data, noise elaboration, and planning indications.

Acknowledgments

The paper is part of PRIN 2022 "OUTFIT" n.2022BAL2F3, funded by European Union, PNRR and Next Generation EU. It has also been supported by the Spoke 9 "Digital Society & SMart Cities" of ICSC - Centro Nazionale di Ricerca in High Performance-Computing, Big Data and Quantum Computing, funded by the European Union - NextGenerationEU (PNRR-HPC, CUP:C83C22000560007).

References

[1] Rahmi Akçelik. 1991. Travel time functions for transport planning purposes: Davidson's function, its time dependent form and alternative travel time function. *Australian Road Research* 21, 3 (1991).

[2] Henrique C. M. Andrade, Buğra Gedik, and Deepak S. Turaga. 2014. *Fundamentals of Stream Processing: Application Design, Systems, and Analytics.* Cambridge University Press. https://doi.org/10.1017/CBO9781139058940

[3] Pierre Aumond, Catherine Lavandier, Carlos Ribeiro, Elisa Gonzalez Boix, Kennedy Kambona, Ellie D'Hondt, and Pauline Delaitre. 2017. A study of the accuracy of mobile technology for measuring urban noise pollution in large scale participatory sensing campaigns. *Applied Acoustics* 117 (2017), 219–226.

[4] Jorge M Bandeira, Pavlos Tafidis, Eloísa Macedo, João Teixeira, Behnam Bahmankhah, Cláudio Guarnaccia, and Margarida C Coelho. 2020. Exploring the potential of web based information of business popularity for supporting sustainable traffic management. *Transport and Telecommunication Journal* 21, 1 (2020), 47–60.

[5] Roberto Benocci, H Eduardo Roman, and Giovanni Zambon. 2021. Optimized sensors network and dynamical maps for monitoring traffic noise in a large urban zone. *Applied Sciences* 11, 18 (2021), 8363.

[6] Arnaud Can, Luc Dekoninck, and Dick Botteldooren. 2014. Measurement network for urban noise assessment: Comparison of mobile measurements and spatial interpolation approaches. *Applied acoustics* 83 (2014), 32–39.

[7] European Commission and Directorate-General for Environment. 2016. *Evaluation of directive 2002/49/EC relating to the assessment and management of environmental noise – Final report.* Publications Office. https://doi.org/doi/10.2779/171432

[8] Koenraad De Smedt, Dimitris Koureas, and Peter Wittenburg. 2020. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. *Publications* 8, 2 (April 2020), 21. https://doi.org/10.3390/publications8020021

[9] Commission Delegated. 2020. *Directive (EU) 2021/1226 of 21 December 2020 amending, for the purposes of adapting to scientific and technical progress, Annex II to Directive 2002/49/EC of the European Parliament and of the Council as regards common noise assessment methods C/2020/9101.* Technical Report. EUROPEAN COMMISSION.

[10] Jaume Ferré-Bigorra, Miquel Casals, and Marta Gangolells. 2022. The adoption of urban digital twins. *Cities* 131 (Dec. 2022), 103905. https://doi.org/10.1016/j.cities.2022.103905

[11] Luca Fredianelli, Stefano Carpita, Marco Bernardini, Lara Ginevra Del Pizzo, Fabio Brocchi, Francesco Bianco, and Gaetano Licitra. 2022. Traffic flow detection using camera images and machine learning methods in ITS for noise map and action plan optimization. *Sensors* 22, 5 (2022), 1929.

[12] Gabriella Graziuso, Simona Mancini, Antonella Bianca Francavilla, Michele Grimaldi, and Claudio Guarnaccia. 2021. Geo-Crowdsourced sound level data in support of the community facilities planning. A Methodological Proposal. *Sustainability* 13, 10 (2021), 5486.

[13] Philip James, Luke Smith, Jennine Jonczyk, Neil Harris, Tom Komar, Aare Puussaar, Miles Clement, and Richard J. Dawson. 2020. Urban Observatory Data Newcastle. https://doi.org/10.25405/DATA.NCL.C.5059913.V4

[14] Feng Jiang, Ling Ma, Tim Broyd, Weiya Chen, and Hanbin Luo. 2022. Digital twin enabled sustainable urban road planning. *Sustainable Cities and Society* 78 (2022), 103645.

[15] Anfosso-Lédée F. Kephalopoulos S, Paviotti M. 2012. *Assessment Methods in Europe (CNOSSOS-EU).* Technical Report. Publications Office of the European Union; JRC72550. https://doi.org/10.2788/32029

[16] Bernd Ketzler, Vasilis Naserentin, Fabio Latino, Christopher Zangelidis, Liane Thuvander, and Anders Logg. 2020. Digital twins for cities: A state of the art review. *Built Environment* 46, 4 (2020), 547–573.

[17] Yuan Liao, Sonia Yeh, and Jorge Gil. 2021. Feasibility of estimating travel demand using geolocations of social media data. *Transportation* (2021), 1–25.

[18] Gaetano Licitra, Elena Ascari, and Luca Fredianelli. 2017. Prioritizing process in action plans: A review of approaches. *Current Pollution Reports* 3 (2017), 151–161.

[19] Gaetano Licitra, Antonino Moro, Luca Teti, and Lara Ginevra Del Pizzo. 2021. The influence of link characteristics on road traffic noise mapping by using Big Data. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 263. Institute of Noise Control Engineering, 5225–5232.

[20] Antonella Longo, Marco Zappatore, Mario Bochicchio, and Shamkant B Navathe. 2017. Crowd-sourced data collection for urban monitoring via mobile sensors. *ACM Transactions on Internet Technology (TOIT)* 18, 1 (2017), 1–21.

[21] Antonella Longo, Marco Zappatore, and Mario A Bochicchio. 2020. Apollon: Towards a citizen science methodology for urban environmental monitoring. *Future Generation Computer Systems* 112 (2020), 899–912.

[22] Stefano Manzo, Otto Anker Nielsen, and Carlo Giacomo Prato. 2014. Effects of Uncertainty in Speed–Flow Curve Parameters on a Large-Scale Model: Case Study of the Danish National Model. *Transportation Research Record* 2429, 1 (2014), 30–37.

[23] Gabriele Mencagli, Massimo Torquati, Andrea Cardaci, Alessandra Fais, Luca Rinaldi, and Marco Danelutto. 2021. WindFlow: High-Speed Continuous Stream Processing With Parallel Building Blocks. *IEEE Transactions on Parallel and Distributed Systems* 32, 11 (2021), 2748–2763. https://doi.org/10.1109/TPDS.2021.3073970

[24] Enock T Mtoi and Ren Moses. 2014. Calibration and evaluation of link congestion functions: applying intrinsic sensitivity of link speed as a practical consideration to heterogeneous facility types within urban network. *Journal of Transportation Technologies* (2014).

[25] Otto Anker Nielsen and Rene Munk Jørgensen. 2008. Estimation of speed–flow and flow–density relations on the motorway network in the greater Copenhagen region. *IET Intelligent Transport Systems* 2, 2 (2008), 120–131.

[26] OpenStreetMap contributors. 2017. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

[27] Boris Otto. 2022. *The Evolution of Data Spaces.* Springer International Publishing, 3–15. https://doi.org/10.1007/978-3-030-93975-5_1

[28] Eulalia Peris et al. 2020. Environmental noise in Europe: 2020. *Eur. Environ. Agency* 1 (2020), 104.

[29] Judicaël Picaut, Nicolas Fortin, Erwan Bocher, Gwendall Petit, Pierre Aumond, and Gwenaël Guillaume. 2019. An open-science crowdsourcing approach for producing community noise maps using smartphones. *Building and Environment* 148 (2019), 20–33.

[30] Zhaokun Qin and Yanmin Zhu. 2016. NoiseSense: A Crowd Sensing System for Urban Noise Mapping Service. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS).* IEEE. https://doi.org/10.1109/icpads.2016.0020

[31] Prapaporn Rattanatamrong, Jarunchai Srisawat, Thapana Boonchoo, and Jason Haga. 2022. Development of a Digital Twin for Smart Building over Edge-Cloud Continuum. In *2022 IEEE 8th World Forum on Internet of Things (WF-IoT).* IEEE. https://doi.org/10.1109/wf-iot54382.2022.10152153

[32] Franck Ravat and Yan Zhao. 2019. *Data Lakes: Trends and Perspectives.* Springer International Publishing, 304–313. https://doi.org/10.1007/978-3-030-27615-7_23

[33] Christian Sicari, Alessio Catalfamo, Lorenzo Carnevale, Antonino Galletta, Antonio Celesti, Maria Fazio, and Massimo Villari. 2023. *Toward the Edge-Cloud Continuum Through the Serverless Workflows.* Springer Nature Switzerland, 1–18. https://doi.org/10.1007/978-3-031-42194-5_1

[34] Pavlos Tafidis, João Teixeira, Behnam Bahmankhah, Eloísa Macedo, Margarida C Coelho, and Jorge Bandeira. 2017. Exploring crowdsourcing information to predict traffic-related impacts. In *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe).* IEEE, 1–6.

[35] Anna Izabel J Tostes, Fátima de LP Duarte-Figueiredo, Renato Assunção, Juliana Salles, and Antonio AF Loureiro. 2013. From data to knowledge: City-wide traffic flows analysis and prediction using bing maps. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing.* 1–8.

[36] Washington DC US Dept.of Commerce Bureau of Public Roads Office of Planning Urban Planning Division. 1964. *Traffic Assignment Manual for Application with a Large, High Speed Computer.* Bureau of Public Roads.

[37] Gary White, Anna Zink, Lara Codecá, and Siobhán Clarke. 2021. A digital twin smart city for citizen feedback. *Cities* 110 (2021), 103064.

[38] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, and et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (March 2016). https://doi.org/10.1038/sdata.2016.18