5G-QoERA: An Integrated Dataset for QoE Assessment in 5G NR Based on User Mobility, Radio Map, Scheduling Decisions, and Application Details

Enrico Boffetti, Arcangela Rago, Giuseppe Piro, Gennaro Boggia
Dept. of Electrical and Information Engineering, Politecnico di Bari, Italy
CNIT, Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Italy
Email: e.boffetti@phd.poliba.it, arcangela.rago@poliba.it, giuseppe.piro@poliba.it, gennaro.boggia@poliba.it

***Abstract*—Today, an unprecedented number of researchers and companies are interested in exploring advanced and optimized protocols and algorithms making the 5G New Radio technology (and of course its evolutions) able to support different services under heterogeneous scenarios. In most cases, their studies leverage computer simulations and simple datasets describing isolated functionalities of the overall complex mobile communication system. Accordingly, despite all these valuable and available tools, there remains a lack of comprehensive, high-quality datasets that can support in-depth analysis, modeling, and testing of the 5G New Radio in real-world deployments. To bridge this gap, this paper presents a 5G dataset for Quality of Experience assessment in the new RAdio (5G-QoERA). Focusing the attention on a video streaming use case, the goal of the dataset is to indicate the level of quality experienced by mobile end users in terms of Mean Opinion Score, while jointly considering several influencing aspects such as real user mobility traces, real-world base station deployments and related radio map environments, variable scheduling decisions, and application-level details. After detailing the generation process of the dataset, a preliminary analysis of its features is conducted to underline the possible research activities that can take advantage of its usage.**

***Index Terms*—5G, NR, QoE, QoS, Mobility, Multimedia.**

I. Introduction

As mobile networks evolve Beyond 5th Generation (5G) and 6th Generation (6G), the increasing complexity of wireless communication systems and related new services needs novel approaches to improve performance by facing up the requirements of end users [1]. Next-generation networks aim to provide not only greater throughput but also better Quality of Service (QoS) and quality perceived by end users, which is defined as Quality of Experience (QoE) [2]. In fact, QoE is a central concept for measuring end-user satisfaction in using services and applications over wireless networks, considering both technical (i.e., objective) and perceptual (i.e., subjective) factors, or solely technical or perceptual factors [3].

After the introduction of new technologies such as massive MIMO, beamforming, and millimeter frequencies in 5G, Beyond 5G and 6G networks promise to further expand network capabilities, by also considering THz communications, the integration of Terrestrial and Non-Terrestrial Networks, the design based on Artificial Intelligence (AI)/Machine Learning (ML), and Network Digital Twin (NDT) paradigm [4]. In particular, the integration of AI and NDT are two key concepts of the new future networks as they allow for a range of useful information to be available for scheduling, anticipatory allocating, and redistributing resources, and preventively maintaining network nodes. Thus, it is possible to ensure the continuous availability of the networks with an increasingly high QoS and improvements of QoE.

Over the years, several QoE-oriented solutions have been proposed to improve the user experience. Such approaches often integrate AI/ML, edge computing, and network slicing techniques to dynamically allocate network resources [5]–[7], considering network conditions and user preferences. However, there is a shortage of open and public datasets to study the direct correlation between the physical layer, the application layer, and perceived QoE. This gap limits the research and development of more efficient solutions to optimize resource allocation in current and future wireless communication systems.

To contribute to the existing literature and to support in-depth analysis, modeling, and testing of the 5G New Radio (5G NR) in real-world deployments, this paper presents a 5G dataset for QoE assessment in new RAdio (5G-QoERA). The purpose of this integrated dataset for evaluating QoE in 5G NR is to advance the study of QoS and QoE in the new era of 5G and Beyond networks, by also considering the user mobility traces from the Taxi/Rome dataset [8], the actual locations of Base Stations (BSs) [9], also using Radio Environment Maps (REMs), by varying the Physical Resource Blocks (PRBs), i.e., the actual set of subcarriers and time intervals assigned for data transmission [10]. For the 5G-QoERA dataset, different values of the number of PRBs are considered to obtain the throughput values of users. Moreover, application details on Packet Loss Rate (PLR) and QoE values based on real multimedia videos are considered to support scheduling decisions and obtain Mean Opinion Score (MOS) values for objective QoE assessment.

The remainder of the paper is as follows. Section II contains a state-of-the-art analysis of various QoE-related datasets, providing a comparative table among them and the 5G-QoERA dataset. Section III describes the source datasets of 5G-QoERA, how they were used to generate the new dataset, and the generation details. Section IV provides an overview of the 5G-QoERA dataset, by presenting its structure, main features, and usage examples. Finally, Section V concludes the paper and outlines future research activities on resource management,

TABLE I
COMPARISON AMONG THIS WORK AND THE OTHER DATASETS ON QUALITY OF EXPERIENCE.

References	QoE Metrics	User Mobility	Physical Layer			Application Layer		
			Radio Technology	Real BS Positions	Variable PRB	Variable PLR	Variable Bitrate	Variable Video Quality
[11]	Subjective						✓	✓
[12]	Subjective					✓		
[6]	Objective					On bottleneck	✓	✓
[13]	Objective		Wi-Fi					✓
[14]	Objective	✓					✓	
[15]	Both						✓	✓
[16]	Both		Wi-Fi			✓	✓	✓
5G-QoERA	**Objective**	✓	**5G-NR**	✓	✓	✓	✓	✓

which can use the presented dataset and benefit from it.

II. RELATED WORK

Recent research efforts have increasingly focused on developing methodologies to maximize the QoE perceived by end users, which remains one of the primary goals of emerging telecommunications networks. To support these endeavors, numerous datasets have been made available to the scientific community, providing real or simulated data on various key network statistics such as jitter, end-to-end delay, PLR, bitrate, and more. These datasets serve as critical resources for understanding and optimizing QoE in different networking scenarios, by adopting subjective metrics [11], [12], objective metrics [6], [13], [14] or both types of categories [15], [16].

For subjective QoE evaluations, the contribution presented in [11] includes 220 video sequences of 5 seconds each, in four different resolutions (from 360p to 1080p), with variable bitrate and variable video quality. The dataset is generated by conducting various tests on more than 30 subjects for several video sequences. In particular, the evaluation is based on the Just Noticeable Difference measurement, representing the point at which a human subject notices a quality difference between compressed videos. The conducted study in [12] introduces WebRTC-QoE, a dataset for Web Real-Time Communications (WebRTC) that focuses on subjective testing under varying conditions of PLR, delay, and jitter. Users were asked to rate their experience using the Absolute Category Rating scale, ranging from 1 (Bad) to 5 (Excellent). The study also uniquely captured data on facial expressions, offering further dimensions to analyzing user QoE.

Regarding the objective QoE analysis, the dataset in [6] is generated by using a specific adaptive multimedia streaming simulation framework to simulate an HyperText Transfer Protocol client and a LibDASH server, a library based on Dynamic Adaptive Streaming over HTTP (DASH). Thus, the dataset captures data like the number of clients, bandwidth, resolution, and delay and the adopted QoE metrics are stalling events and rebuffering ratio. Each entry represents average statistics with variable PLR on the bottleneck, bitrate, and video quality, contributing valuable insights into QoE under different network and client conditions. The dataset in [14] collects real multimedia traffic statistics from a variety of users, both mobile and non-mobile. Specifically, the MONROE dataset, which innovatively contains user mobility information in the context of QoE evaluations, contains application layer metrics such as packet loss, packets received, and end-to-end delay, analyzed across two multimedia streaming protocols, i.e., WebRTC and DASH. This real-world dataset helps understand QoE, through the objective metric on stalling events, under diverse network conditions and protocol implementations. To obtain the dataset presented in [13], experimental tests are carried out using two Access Points (APs) while considering different video quality and network statistics to measure the Peak Signal-to-Noise Ratio (PSNR), which is then converted to an objective QoE assessment, i.e., MOS scale. Moreover, the authors adopt this dataset to build a Deep Reinforcement Learning (DRL) model to optimize QoE by dynamically adjusting AP configurations, such as transmission power and channel selection.

Also, joint subjective and objective QoE evaluations are conducted through the widely adopted Differential Mean Opinion Score (DMOS) [17], a metric that can help to determine how much the differences introduced in test videos degrade subjective picture/video quality. In the work [15], a dataset based on subjective tests in a laboratory environment is presented. Specifically, various participants evaluated distorted video streams with variable bitrate and video quality on mobile devices. The participants' feedback was recorded using DMOS. Similarly, the contribution in [16] provides a dataset containing distorted videos. The dataset simulated various wireless network conditions, including video compression and packet loss, with variable distortions, like frame freezing, over time. Subjective tests were conducted to collect both final mean scores and continuous user ratings during video playback, enabling a detailed analysis of QoE degradation over time.

These previous studies lack a direct correlation between effective resource allocation and QoE metrics of mobile users in a 5G NR environment, particularly when using real BSs. To the best of the authors' knowledge, the existing research also falls short of addressing the impact of real-world mobility on QoE, as highlighted in Table I. To bridge this gap, the 5G-QoERA framework introduces a novel approach by incorporating user mobility traces in 5G NR networks, leveraging real BS REMs. This dataset captures application-level details alongside geographical distributions of actual BSs, enabling an in-depth assessment of QoE across various video quality scenarios. 5 By integrating resource scheduling decisions with user mobility data, 5G-QoERA first offers a comprehensive analysis of how network resource allocation impacts user experience, presenting a groundbreaking advancement in the field.

III. Dataset Generation

The methodology for realizing the 5G-QoERA dataset is described in the following subsections, with a detailed explanation of each block shown in Fig. 1. The latter presents the generation process of the original dataset, which is built upon existing data sources and enhanced through advanced modeling and simulation of 5G and Beyond 5G communication systems. Starting from real user mobility data [8], realistic BS positions [9], and video bitrate information [18], the 5G New Radio (NR) REMs are integrated and combined with application-specific details to obtain throughput values so as to calculate PLR values and consequently MOS values, by creating a comprehensive dataset for evaluating user experience in mobile networks.

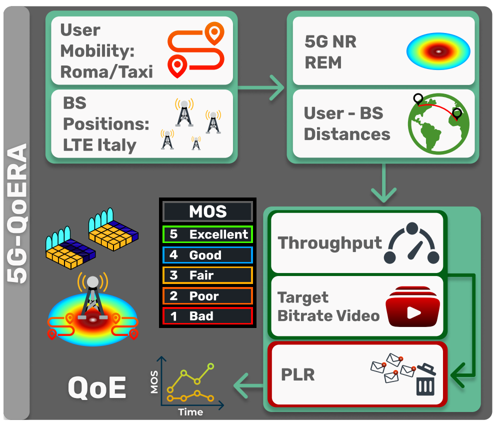

Fig. 1. Generation process of the 5G-QoERA dataset.

A. User Mobility and BS Positions

To consider real user mobility traces (i.e., the top left block in Fig. 1), the Roma/Taxi dataset [8] is used. This dataset includes several mobility traces from taxi drivers in Rome, Italy. Specifically, it provides data on the movements of 320 taxi drivers over 30 days, recorded at approximately 15-second intervals. For each user, identified by an ID y, the dataset provides both the timestamp and the position, expressed in terms of latitude and longitude coordinates.

To obtain verified and accurate BS positions in Rome (i.e. the second block from the top left in Fig. 1), the LTE Italy tool [9] is employed. The purpose of this tool is to share the positions of BSs belonging to various Italian Mobile Network Operators (MNOs), allowing precise estimates of the coverage and speeds that specific MNOs can achieve in a given location. Therefore, by considering the user mobility dataset, the BS positions of an Italian MNO in Rome have been considered.

B. 5G NR REM and User-BS distances

To simulate the behavior of BSs, MATLAB 5G Toolbox is used. It provides advanced tools for modeling, simulating, and analyzing 5G network performance, as demonstrated in recent studies [19]. In particular, user-BS distances and 5G NR REM (i.e., the two blocks at the top right in Fig. 1) have

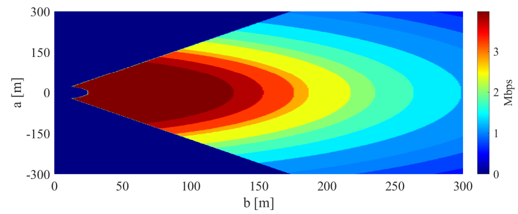

Fig. 2. 5G NR REM on throughput for $N_{\text{PRB}} = 5$ and for each BS.

been taken into account. The 5G NR REM is a spatial representation of the radio environment that considers signal power distribution, interference, and other critical parameters such as throughput, Signal-to-Interference-plus-Noise Ratio (SINR), and Modulation and Coding Scheme (MCS) values. Note that MCS determines the modulation scheme and coding rate (ratio of information bits to total transmitted bits, affecting throughput and error correction) based on channel conditions [10]. Using this kind of maps, it is possible to obtain detailed and accurate throughput information by varying the number of PRBs per user, i.e., N_{PRB}. An example of 5G NR REM on throughput for $N_{PRB}=5$ and for each BS is reported in Fig. 2. For each point in the figure, the exact throughput value is provided and reported in the associated matrix via the indices $(a, b)_{y,\tau}^{x,N_{PRB}}$ where a and b are the matrix indices, x defines the BS ID, N_{PRB} defines the number of PRBs assigned, y defines the User ID, and τ represents the time instant. Thus, in this context, the term REM refers to the precise use of real-world radio maps to obtain the 5G-QoERA dataset, generated with actual BS locations and detailed link-level analysis performed using MATLAB 5G Toolbox.

To obtain the distance between the BS x and the specific user y by knowing their positions, the Great-circle distance [20] is considered:

$$D_{x,y} = R \cdot \cos^{-1}\Big(\sin(\phi_x)\sin(\phi_y) + \cos(\phi_x)\cos(\phi_y)\cos(\lambda_x - \lambda_y)\Big) \quad (1)$$

where $D_{x,y}$ is the distance between the the BS x and the user y, depending on the latitude ϕ_x and the longitude λ_x of the BS x and the latitude ϕ_y and the longitude λ_y of the user y, and $R = 6372.795\ km$ is the Earth radius. Specifically, the Great-circle distance formula provides an accurate estimate of the distance between two geographical coordinates using the latitude and longitude of the points, considering the Earth's curvature. Since the approximation of the Earth to a sphere introduces a negligible error [20], the adoption of (1) is suitable for the calculations of user-BS distances.

C. Throughput, Target Video Bitrate, PLR, and QoE

In this subsection the blocks at the bottom of Fig. 1 are detailed. Note that the 5G-QoERA dataset is based on scheduling decisions because it involves the generation of different QoE levels for end-users, based on the varying allocation of PRBs by the BS within specific time intervals τ. In addition, application-level details encompass key performance metrics such as PLR, bitrate, and video quality information at the application layer.

After determining the distance of each user from each BS, the throughput is calculated using Algorithm 1. The throughput is only computed for SINR values greater than -20, as the noise level for values below -20 is so high compared to the useful signal that transmission would become impractical. The evaluation is conducted for N_{PRB} values ranging from 5 to 40 (i.e., values in line with 3GPP specifications [10]). By utilizing the distance information between BSs and users, as well as REMs, the calculation of $SINR$, MCS, and $Throughput\_N_{PRB}$ is performed based on the specified parameter settings, which also include power values and matrix size.

Algorithm 1 SINR, MCS, and Throughput Calculation

$N_{\text{PRB}} \leftarrow 5$
while $N_{\text{PRB}} \leq 40$ **do**
  Parameter settings
  for each point (a, b) in the Grid **do**
    Compute $distance$
    Compute $SINR$
    Determine MCS based on $SINR$
    if $SINR >$ -20 **then**
      Compute $Throughput$
    end if
  end for
  Save *Throughput_N_PRB.mat*
  $N_{\text{PRB}} \leftarrow N_{\text{PRB}} + 5$
end while

Then, for PLR and MOS calculation, Algorithm 2 takes as input the matrices generated by the Algorithm 1, which contains the throughput values in each matrix cell. For each value of N_{PRB}, the complete throughput matrix $Throughput\_N_{PRB}$ and the related timestamp are saved in mat and τ, respectively. For each user y, the distance $D_{x,y}$ from the BS x is calculated. If the REM limits are not exceeded (see Fig. 2), a mapping between the distance and the complete throughput matrix is performed by using the *map2table* function to extract the specific cell indices a and b, obtaining the related $Throughput$ value. To calculate the PLR, analysis on specific Media Presentation Description (MPD), which are summarized in Table II, have been conducted. Specifically, authors in [18] developed a dataset to specify the target bitrate for each resolution level using DASH, facilitated by the MPD. This target bitrate was subsequently employed to calculate the PLR for each record. In fact, for each bitrate value, the PLR is calculated according to the throughput and the target video bitrate of the various resolutions (Table II). Note that to calculate how many packets are transmitted and received based on the various target video bitrates, the standard Maximum Transmission Unit (MTU) equal to 1500 bytes has been considered, including IP and UDP headers.

Consequently, to have an estimation on the QoE, perceived by users experiencing the specific and identified throughput value, the MOS metric has been calculated for each PLR value through:

$$MOS_{x,y} = e^{1.576-(4.188\times10^{-4}\times Delay)-(5.766\times10^{-2}\times \text{PLR})}. \quad (2)$$

The formula in (2) refers to the model adopted in [21] and inspired by the ITU E-model standard [22], which is particularly suitable for evaluating QoE in the case of multimedia streaming over the network. The above formulation considers the delay and the PLR. As can be seen, by having multiplying factors that differ by two orders of magnitude (i.e., 10^{-4} and 10^{-2} for the delay and PLR, respectively), the PLR has a much higher impact than the delay on the QoE assessment, causing it to decrease exponentially proportional to the coefficient. To this end, for the dataset analysis, a fixed and standard value of delay has been adopted ($0.1\ ms$).

TABLE II
VIDEO RESOLUTIONS AND CORRESPONDING TARGET BITRATE.

Resolution [px]	**Target Bitrate [kbps]**
240p	50, 100, 150
360p	200, 250, 300, 400, 500, 600, 700
720p	900, 1200, 1500, 2000
1080p	2500, 3000, 4000, 5000, 6000, 8000

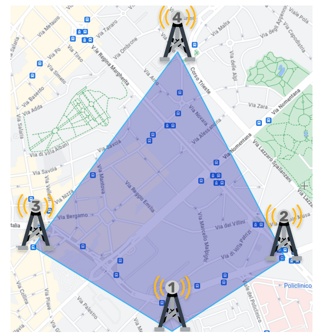

Fig. 3. Selected area of 608707 m^2 delimited by the four BSs.

Algorithm 2 PLR and MOS Calculation

for $N_{PRB} \in \{5, 10, \ldots, 40\}$ **do**
  $mat \leftarrow$ loadmat(*Throughput_N_PRB.mat*)
  $\tau \leftarrow$ timestamp
  for each τ in $user\ y$ **do**
    $D_{x,y} \leftarrow$ calc_distance$(\phi_x, \lambda_x, \phi_y, \lambda_y)$
    if $D_{x,y} \leq$ *REM_limits* **then**
      $(a, b) \leftarrow$ map2table$(D_{x,y})$
      $Throughput \leftarrow mat[a, b]$
      for $bitrate \in \{50, \ldots, 8000\}$ **do**
        PLR $\leftarrow$ calc_plr$(Throughput, bitrate)$
        MOS $\leftarrow$ calc_mos(PLR)
        Store PLR and MOS values for each τ
      end for
    end if
  end for
end for

IV. Dataset Overview

To analyze the 5G-QoERA dataset, a preliminary investigation is conducted. The following subsections provide an overview of the dataset with its main features, structure, and usage examples. Note that the 5G-QoERA dataset can be accessed through the GitHub repository reported here[1].

[1] https://github.com/telematics-lab/5G-QoERA

TABLE III
SAMPLE RECORDS OF THE 5G-QOERA DATASET FOR THE BS $x = 1$ WITH $N_{PRB} = 30$.

User features					PLR [%]				QoE Metric
y [#]	τ [time]	ϕ_y[°]	λ_y[°]	Throughput [Mbps]	240p	360p	720p	1080p	MOS [#] for [240p, 1080p]
2	2014-02-01 00:00:10.16	41.9081300	12.5043668	14.344	0.0 - 0.0	0.0 - 0.0	0.0 - 0.0	0.0 - 0.0	[4.834, 4.834]
2	2014-02-01 00:00:25.77	41.9086630	12.5066009	4.352	0.0 - 0.0	0.0 - 0.0	0.0 - 0.0	0.0 - 45.60	[4.834, 1.000]
315	2014-02-27 17:28:14.48	41.9071726	12.5012530	2.152	0.0 - 0.0	0.0 - 0.0	0.0 - 0.0	13.92 - 73.10	[4.834, 1.000]
315	2014-02-27 17:28:22.05	41.9073797	12.5014076	3.240	0.0 - 0.0	0.0 - 0.0	0.0 - 0.0	0.0 - 59.50	[4.834, 1.000]

A. Parameter Settings and Main Features

Only four BSs and users within a specified area of $608707m^2$ are considered, as represented in Fig. 3. The locations of the four considered BSs, by using (ϕ_x, λ_x) for the Global Positioning System (GPS) coordinates, are: (41.907008, 12.5048333) for $x=1$; (41.9097288, 12.5103417) for $x=2$; (41.9101547, 12.4978939) for $x=3$; (41.9172276, 12.5049841) for $x=4$.

The distance between users and BSs does not exceed 300 m (i.e., REM limits); otherwise, it denotes that the user is not served by that specific BS. Additionally, the BSs transmits in all directions under the Line-of-Sight (LoS) conditions, neglecting the effect of Doppler shifts for the purposes of the contribution.

The throughput range varies depending on the chosen number of PRBs. For example, when $N_{PRB} = 5$ is assigned, the maximum value is 3.968 Mbps, while the minimum value is 0.0 Mbps. Similarly, for $N_{PRB}=20$, the maximum value is 15.88 Mbps, and the minimum value remains 0.0 Mbps. To obtain target bitrate values for different video resolutions, the information provided in Table II is used.

B. Dataset Structure and Analysis

By considering four BSs, the 5G-QoERA dataset presents 32 Tab-Separated Values (TSV) files, 8 for each BS, whose total weight is 1.05 GB. Each file has two varied indices: the first indicates the ID of the BS, i.e., x, while the second shows the number of the assigned PRBs N_{PRB}. Table III is a simplified extract of the generated dataset for the BS $x=1$ with $N_{PRB}=30$. Note that within each generated file there are several records by considering the user-BS distance calculated through (1), the target bitrate values, and then the PLR based on Table II. As reported in Table III, each record consists of the user features and information on PLR and QoE. Specifically, each record has a user ID, i.e., y, unique for each user, τ representing the exact timestamp (i.e., date and hours) when the latitude ϕ_y and the longitude λ_y coordinates of the user were detected, and the *Throughput*, i.e., the throughput value for the specific user. Depending on the considered PRBs, the value of *Throughput* varies, also changing the values of *PLR* and consequently the values of *MOS*. In particular, by considering 30 PRBs as in the sample records, different values of PLR, expressed in percentage, are reported in Table III for different resolutions (i.e., 20 columns in total by considering the different target bitrate for the four values of resolutions, as in Table II). Due to lack of space, QoE information is expressed through the range of MOS values for the resolution 240p, 360p, 720p, and 1080p (i.e., 20 columns in total of MOS values as PLR varies).

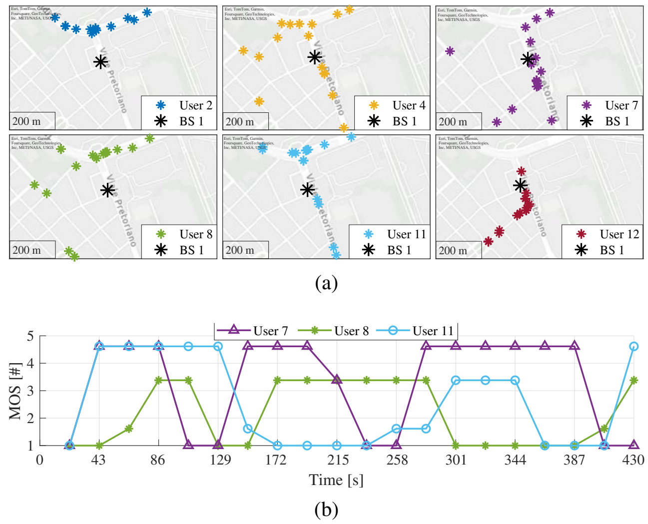

Fig. 4. (a) Mobility of 6 example users during $430s$, served by the BS x=1, and (b) MOS over time with N_{PRB}=5 for an example 4 Mbps video (1080p).

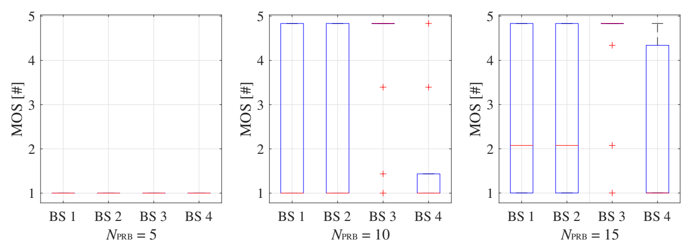

Fig. 5. Cumulative MOS values for each BS for an example video segment (1080p) requiring 6 Mbps for N_{PRB}=5, 10, 15.

To give an insight, Fig. 4 (a) represents the mobility behavior for six example users served for a certain period (i.e., $430s$) by the BS $x=1$. Fig. 4 (b) shows the MOS trends in the same period for three of the sample users for an example 4Mbps video. It can be seen that MOS values are different and they vary over time. Furthermore, Fig. 5 highlights the median value (i.e., the red line), the 25th and the 75th percentile (i.e., the bottom line and the top line of the blue rectangle), as well as the minimum and the maximum MOS values (i.e., the edges of the vertical black line) for each BS for N_{PRB}=5, 10, 15. It can be noted that the median MOS values for $N_{PRB}=5$ and $N_{PRB}=10$ is generally equal to 1 for most BSs. By increasing N_{PRB} to 15, the latter raises, by also reaching the maximum MOS value (i.e., 4.8). Note that for the BS $x=3$, starting from $N_{PRB}=10$, high values of MOS and some outliers (i.e. the plus sign in red) are obtained.

C. Usage Examples

The developed dataset, based on real mobile user traces and enriched with throughput and QoE information via the MOS values, offers numerous application opportunities in the area of AI for Beyond 5G and 6G networks. Given its temporal nature, it is particularly suitable for implementing advanced predictive models and Deep Learning (DL) techniques, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memorys (LSTMs), by predicting future patterns related to user behaviors and network performance. This approach could facilitate dynamic optimization of network resources and QoS management by anticipating changes in throughput and QoE. In addition, the dataset lends itself to the development of Reinforcement Learning (RL) and DRL algorithms, which can be leveraged to create adaptive and intelligent solutions in complex network scenarios. In particular, such algorithms could improve resource provisioning, traffic management, and spectrum resource allocation in real-time, maximizing QoE for mobile users, in conjunction with the spreading NDT paradigm.

Thus, the use of real and detailed information makes this dataset a valuable asset for research and deployment of AI-based and NDT solutions in the networking domain, helping to innovate the traditional approach to efficiently and proactively managing Beyond 5G and 6G networks by analyzing, modeling, and testing the 5G NR in real-world deployments.

V. Conclusions

This paper introduced a 5G dataset for Quality of Experience assessment in the new RAdio (5G-QoERA). It is a powerful instrument to facilitate a comprehensive evaluation of the Quality of Experience in 5G New Radio. It incorporates real user mobility traces, actual Base Station deployments with related Radio Environment Maps, variable scheduling decisions, and application-level details for accurate objective Mean Opinion Score calculation in a video streaming use case. The preliminary analysis of the dataset demonstrates its potential for advancing research in the field of 5G and Beyond 5G networks, particularly in resource management and network optimization. In fact, future work will focus on the the expansion of the 5G-QoERA by considering more BSs and user data, other real mobility effects (e.g. Doppler shifts), and its usage in conjunction with cutting-edge Artificial Intelligence techniques and the Network Digital Twin paradigm to model and test dynamic resource provisioning adjustments, such as selecting the optimal number of Physical Resource Blocks, to enhance the predictive and adaptive network capabilities for improving end-user experience and avoiding over-provisioning.

Acknowledgment

This work was supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, Mission 4, Component 2, in the context of partnership on "Telecommunications of the Future" (PE00000001-program "RESTART", CUP: D93C22000910001), national center on "Sustainable Mobility" (CN00000023 - program "MOST", CUP: D93C22000410001), and partnership on "Cybersecurity" (PE00000007-program "SERICS", CUP: D33C22001300002, project ISP5G+). It was also supported by the PRIN 2022 projects INSPIRE (grant no. 2022BEXMXN 01) and HORUS (grant no. 2022P44KA8) funded by the Italian MUR, by the HORIZON MSCA project BRIDGITISE (grant no. 101119554), and by "The house of emerging technologies of Matera (CTEMT)" project funded by the Italian MIMIT.

References

[1] M. Z. Chowdhury *et al.*, "6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[2] H. H. H. Mahmoud, A. A. Amer, and T. Ismail, "6G: A comprehensive survey on technologies, applications, challenges, and research problems," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 4, p. e4233, 2021.

[3] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A Tutorial on Video Quality Assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[4] W. Jiang *et al.*, "The Road Towards 6G: A Comprehensive Survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[5] J.-B. Wang *et al.*, "A Machine Learning Framework for Resource Allocation Assisted by Cloud Computing," *IEEE Network*, vol. 32, no. 2, pp. 144–151, 2018.

[6] V. Vasilev *et al.*, "Predicting QoE Factors with Machine Learning," in *2018 IEEE International Conference on Communications (ICC)*, 2018.

[7] W. Wu, C. Zhou *et al.*, "AI-Native Network Slicing for 6G Networks," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 96–103, 2022.

[8] L. Bracciale, M. Bonola *et al.* (2022) Crawdad roma/taxi. [Online]. Available: https://dx.doi.org/10.15783/C7QC7M

[9] LTE Italy. Accessed on July 9, 2024. [Online]. Available: https://lteitaly.it/it/

[10] 3GPP, "5G; NR; Physical channels and modulation," 3rd Generation Partnership Project, Tech. Specification (TS) 38.211, July 2023, V17.5.0.

[11] H. Wang *et al.*, "VideoSet: A Large-Scale Compressed Video Quality Dataset Based on JND Measurement," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, 2017.

[12] G. Bingöl, S. Porcu, A. Floris, and L. Atzori, "WebRTC-QoE: A dataset of QoE assessment of subjective scores, network impairments, and facial & speech features," *Computer Networks*, vol. 244, p. 110356, 2024.

[13] H. D. Moura *et al.*, "Improved Video QoE in Wireless Networks Using Deep Reinforcement Learning," in *2023 19th International Conference on Network and Service Management (CNSM)*, 2023.

[14] C. Midoglu *et al.*, "Open video datasets over operational mobile networks with MONROE," in *Proceedings of the 9th ACM Multimedia Systems Conference*, ser. MMSys '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 426–431.

[15] C. G. Bampis *et al.*, "Study of Temporal Effects on Subjective Video Quality of Experience," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5217–5231, 2017.

[16] A. K. Moorthy *et al.*, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, 2012.

[17] Recommendation ITU-T G.800.1, "Mean opinion score (MOS) terminology ," ITU-T, Tech. Rep., 2016.

[18] S. Lederer, C. Müller, and C. Timmerer, "Dynamic adaptive streaming over HTTP dataset," in *Proceedings of the 3rd Multimedia Systems Conference*, ser. MMSys '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 89–94.

[19] L. Ballotta *et al.*, "VREM-FL: Mobility-Aware Computation-Scheduling Co-Design for Vehicular Federated Learning," *IEEE Transactions on Vehicular Technology*, pp. 1–16, 2024.

[20] G. B. N. Department, *Admiralty Manual of Navigation: BR 45(1)*, ser. BR Series. Stationery Office, 1997, no. v. 1.

[21] M. S. Mushtaq *et al.*, "QoE in 5G cloud networks using multimedia services," in *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–6.

[22] Recommendation ITU-T G.107, "The EModel: a computational model for use in transmission planning," ITU-T, Tech. Rep., 2015.