IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 August 2022, accepted 5 September 2022, date of publication 12 September 2022, date of current version 21 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3206371

RESEARCH ARTICLE

Microwave Tomography Data Deconstruct of Spatially Diverse C-Band Scatter Components Using Clustering Algorithms

G. A. SHANMUGHA SUNDARAM[1,2], (Member, IEEE), R. GANDHIRAJ[1,2], B. N. BINOY[1,2], S. I. HARUN[3], AND S. N. SURYA[4]

[1]Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
[2]SIERS Research Laboratory, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
[3]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 637457
[4]Microwave Sensing, Signals and Systems (MS3) Group, Department of Microelectronics, Delft University of Technology (TU Delft), 2628 Delft, The Netherlands

Corresponding author: G. A. Shanmugha Sundaram (ga_ssundaram@cb.amrita.edu)

The work was supported by the National Instruments Corporation, Austin, TX, USA, for providing all of the equipment, with the life cycle support, through the Academic Research Grant, in 2016.

ABSTRACT Communication signals that propagate through free space are subject to multi-path interference due to scattering by various objects in the propagation channel. The effect is especially severe in complex situations in dense urban environments. To investigate the problem, a typical multi-static detection scenario is reconstructed under controlled laboratory conditions, from which suitable data sets are created. Data-driven models are then employed in EDGE computing platforms to profile the scatter centers based on the subjective manner in which they affect the signals. These have been interpreted primarily based on clustering algorithm (CA) operations – using a select suite of pre-processing models that effectively tame the variations in the C-band spatial-temporal data. A subset of the data of interest could then be subjected to an optional, compute-intensive machine learning (ML) approach. The relative advantages of the proposed method *vis-a-vis* an array of conventional schemes are highlighted, while also considering its carbon friendly attribute. Given the more significant association of the data to antenna radiation patterns, estimation of the latter can now be performed free of any anechoic chamber set up in a time and cost agnostic manner. The benefit of this work would lie in the realm of mid-band 5G-NR (and the future 6G) cellular communication systems deployment, where optimizing the distributed antenna location attributes on time and cost-constrained scales becomes imperative before any large-scale deployment.

INDEX TERMS Microwave radio communication, multi-path channels, pattern clustering methods, radio propagation terrain factors, scattering, signal processing antennas.

I. INTRODUCTION

The performance of any radio access network (RAN) determines how customers experience fifth- and sixth-generation (5-6G) radio services and drive the economics of mobile data networks. The 5G RAN architecture and air interface support diverse, high-performance services and can be configured to scale in multiple dimensions (data rate, latency, reliability, etc.) depending on the spectrum, technology implementation, and deployment choices. The use of modern innovative technology in mission-critical operations, industrial automation and manufacturing revolve around efficient connectivity solutions among large-scale machine-to-machine (M2M) communication devices, pervasive computing and the internet of things (IoT) infrastructure for an effective realization of the Industry 4.0 agenda [1], [2], [3].

In various situations that range from terrestrial peer-to-peer transmission to interstellar satellite communications, the primary objective is to transfer information through a propagating channel [4], [5], [6]. A propagation channel has an unpredictable response to radio waves traveling across it. It incorporates undesirable factors that would result in

The associate editor coordinating the review of this manuscript and approving it for publication was Junaid Nawaz Syed.

distortion and scattering in multi-path propagation contributed by line of sight (LoS) and non-LoS objects [7]. The presence of scatter centers in the propagation channel would result in distorted reflect- or refract-type radio signals along several directions as defined by the type of interaction and physical properties. Such random events lead to severe deterioration in the signal quality on the receiver side.

Effective and reliable communication shall factor against the significant amount of additive noise acquired by the receiver antenna. Hence, it is paramount to characterize the channel quantitatively and qualitatively. Defining a robust model for the scatterers has been a considerable challenge. Typical schemes that have been adopted in such a context are the following:

1) A model is defined using analytical formulations or conducting channel measurement campaigns. These approaches have inherent advantages as they rely on a deep understanding of the system based on well-established scientific principles [4], [5], [6], [7].
2) Jie Zhou *et al*. [8] used a generalized three-dimensional (3-D) scattering model that considered the transmitter placed at the center of a 3-D spheroid-shaped region. The scatterers are assumed to be uniformly distributed within the radius of this region. The work led to the development of a model that performs significantly better when compared to previously developed 2-D models.
3) A geometrical model was developed by Lu *et al*. [9], and simulations were carried out using different artificial terrain databases.
4) Various model-driven approaches have been proposed [10], [11], [12] for radio signal communication in indoor and outdoor environments to characterize the propagation channel.
5) The scatterers are assumed to be uniformly distributed along the path of the communication channel. Given the inherently empirical nature of such schemes, finding a suitable model and refining it until it produces the desired results is often tedious [13].
6) The scattered signal magnitude from various types of absorber materials in the propagation medium depends on the morphology and incident angle [14].

Diffuse scattering models have been extensively invoked in the C-band and mmWave regions to explain signal distortions such as path loss, delay spread and cross-polarization ratio in indoor propagation environments [15], [16]. The practical interpretations are given based on ray tracing (RT) approaches. The signal interacts with various objects in the medium, each with a unique surface scattering coefficient. An effective combined, and hence diffuse, representation of the various such incidents is measured at the receive side. It is shown to agree with the simulation results based on the RT schemes. The scattered signal magnitude from various types of absorber materials in the propagation medium depends on the morphology and incident angle [14].

Scattering-induced multipath signals are closely identified with the concept of indirect far-field measurement, wherein the complexities in RF testing would scale directly with the properties of individual scatterers. All such interactions result in a marked departure from the plane wavefront definition integral to the measurements conducted at a lengthier LoS separation classified as the Fraunhofer distance [17]. Extracting accurate multipath components from channel measurements is an essential and distinguishing factor in simultaneous localization and mapping (SLAM) solutions to multipath-assisted positioning [18], [19], [20], [21] and is based on the operating frequency and bandwidth.

II. SIGNIFICANCE

The frequency band considered here for evaluation is significant in the context of the 5G new radio (5G-NR) wireless technology in digital cellular networks. Of the two bands allocated for this service, frequency range 1 (FR1) constitutes the mid-band 5G (and the future 6G) in the sub-6 GHz band, considered the ''waterfront property'' of the RF spectrum for 5G. FR1 offers the furthest reach in terms of negotiating terrestrial obstacles than FR2 (the mm-wave band) [22], hence the best compromise between RF coverage and RF bandwidth. This mid-band spectrum balances coverage and capacity characteristics, establishing itself as a vital player in the global rollout of 5-6G services.

A 5-6G network will have to incorporate extensive testing that optimizes its performance in an eventual deployment, based on propagation analysis and coverage mapping. This testing process shall also ensure peak performance in the related systems while arraying against other negative attributes such as RF interference, noise and distortions. The 3GPPs vision of pervasive 5G communication includes the usage of space-based satellite and airborne high-altitude platform station (HAPS) network nodes [23]. Site and channel characterizations for the eventual location of antennas are an elaborate exercise in employing an extensive array of indoor and outdoor test range equipment.

According to the 3GPP standard, carrier aggregation can be activated for each cell group [24]. The 5G deployments in FR1 combine multiple LTE carriers with one NR carrier. A vast majority of these networks worldwide use the TDD mode. As most frequency bands worldwide are FDD and are used by LTE, the first 5G NR network deployments would take advantage of these underutilized TDD frequency bands. Hence, the first generation of 5G modems and, subsequently, the first generation of 5G mobile devices (MD) only support the TDD mode for FR1. Not all service providers own spectrum licenses within a TDD band. To take advantage of 5G with optimized quality of service, to lower latencies and to further address the new market verticals (*e.g.*, automotive and industrial), a network operator must transition to standalone (SA) mode [25], in which the 5G RAN is connected to the 5G core network across a series of intermediate steps in the deployment. The optimum path an operator follows is based

on a carefully evaluated 5G (and the future 6G) deployment strategy.

Many sub-6 GHz deployments in 5-6G are aimed in the C-band; all these shall be operating on a maximum bandwidth of 100 MHz in a 4 × 4 MIMO scheme and a TDD modulation type [24]. Each cell group would incorporate carrier aggregation as per the 3GPP standards. Given the prevalence of FDD in 4G LTE, TDD is likely to be utilized under par in the various 5G mobile communication devices in FR1. However, another issue with TDD implementations is the service providers' reduced levels of spectrum ownership (SP). Since peak quality of service (QoS), everything-to-everything (X2X) connectivity and reduced latency are the hallmarks of operating in transitioning to 5G, the SPs would have to deploy a stand-alone arrangement of 5G solutions [25].

Current 5G networks build on decades of experience in mobile network design and operation, resulting in competent networks with well-established tools, processes, and algorithms to optimize RAN performance. These traditional models are the foundation of good network design [26], [27]. Now, however, artificial intelligence (AI) and machine learning (ML) technologies are emerging that have the potential to radically reshape how 5G networks are designed, built, and operated. At the same time, 5G can deliver benefits to AI/ML by helping to enable distributed and federated learning [28].

Recent advances in AI technologies, such as reinforcement learning and deep reinforcement learning (DRL), can establish a feedback loop between the decision-maker and the physical system. The decision-maker can iteratively refine its action based on the system's feedback to reach optimality eventually. DRL can be applied to address several emerging issues in communication and networking, including adaptive modulation, wireless caching, data offloading, and so on [29].

Deployment of mainstream AI or ML algorithms in real-time applications has unique and conflicting needs. It requires high-performance computing (HPC) using optimized CPUs, advanced GPUs, flash memory NVMe drives, and high-bandwidth interconnects. Simultaneously, it also requires that this HPC capability be:

–Compact enough, for easy installation in vehicles, aircraft, ships, trailers, and the like.

–Rugged enough, to endure and survive harsh conditions

–Vibration, shock, temperatures, water, dust, and similar threats.

Nearly all AI and ML compute platforms fail to integrate and optimize high-performance computing with compact, rugged form factors.

Classically, AI is viewed as ''intelligence'' (datasets, learning modes, algorithms, etc.) hosted in a centralized cloud and has excellent ongoing potential. However, privacy, reliability, latency, and power consumption are driving toward on-device and edge AI/ML. When combined with connectivity, this is sometimes known as the ''connected intelligent EDGE.'' For real-time processing in the 5G RAN or to support low latency 5G applications, placing ML deployments close to where ms or μs decisions are implemented will be essential.

Multi-access EDGE computing is a significant development in-network functionality where data is processed locally at the edge of the network—close to users and devices—to circumvent congestion. A key benefit is the significantly reduced inference delay. EDGE computers are a new generation of high-performance, modular compute servers that are:

–Optimized for ML workloads, such as the highly facile PALE models considered in this work.

–Designed throughout for rugged environments.

–Scalable by upgrading internal components, such as advanced NVMe flash memory, PCIe 4.0, and high-bandwidth networking that can potentially add additional individual units.

The telecommunication network disaggregation that eventually aims to provide competent resources at the EDGE has an unintended yet lucrative aside to it in enabling a joint integrated multi-access computing solution at these EDGE locations. With the number of such components requiring deployment and management across geographic areas projected to grow exponentially, the communications services providers are looking to transform their networks as capable infrastructure to monetize 5-6G.

The current work discusses using certain pre-processing, data-driven algorithms in 5-6G radio network design and optimization. It explores possibilities for applying these techniques to current 5G (and the future 6G) air interfaces, mainly in the deployment or test phases, to generate immediate improvements to systems. A forecast has been made on the emerging 3GPP AI/ML framework to shape the development of radio access technologies over the longer term. In particular, the work references federated learning [30] – a process that creates a collective model from data sourced across distributed learning nodes. Results from the study shall demonstrate the ability of lightweight-compute-device-based ML models to contribute unique data and efficiencies to the training process and in on-device inference generation for rapid performance improvements. With these rugged-designed EDGE platforms, quick deployment of ML models for real-time ''transportable'' applications is realized.

Recent advancements in computation and the advent of highly efficient clustering algorithms have made it possible to model the propagation channels using an entirely data-driven approach. This task has been attempted using a select band of pre-processing algorithm-lean engines (PALE) that find immediate acceptance in embedded EDGE computing platforms such as MDs and other portable devices. Such lightweight, compact and rugged computing paradigms bring proximity computation to the data sources to enable trade-offs independent, robust, time-sensitive, and acquire-process ability in the information processing resources. The result is an agile decision-making ability regarding the discrimination of scatterer-defined signatures present in the propagation channel. Based on such time-critical operations, a first-order effectiveness of these on-the-fly type algorithms, when

deployed in the exercise of profiling scatterers, has been convincingly demonstrated in work reported here from just a pre-processing level analysis of C-band spatial-temporal data sets. The PALE algorithm shall be positioned ahead of the rigorous ML computation tools (not a part of the current work) in the process chain. It shall induce the latter to further refine the classification process of scatterers from among the select subset of data samples that are either of interest or concern. The ML models, which thrive in an ecosystem of limitless computing resources [31], would be deployed in an *ad-hoc* and rational manner on data that has already been tamed of any significant variations by the PALE schemes. The thoroughness in the performance by the latter would alone suffice in the current context, while they would also have an extended role in eliminating misleading patterns in the data. Such a possibility could otherwise potentially occur due to biases being introduced in the analyzed outcomes of classification because of the ''learning'' process involved in the ML schemes [32]. The advantage of these algorithms over traditional model-driven approaches is that they are adaptive to changes in the environment and various types of inputs while also enabling a quick estimation of the dynamic channel parameters.

Scattering-induced multipath signals are closely identified with the concept of indirect far-field measurement, wherein the complexities in RF testing would scale directly with the properties of individual scatterers. All such interactions result in a marked departure from the plane wavefront definition integral to the measurements conducted at a lengthier LoS separation classified as the Fraunhofer distance [17]. The work reported here aims to characterize the propagation channel to create a comprehensive profile of as many of the scatterers in a given signal space. Applying the concept of image theory (IT) [33], [34] to the conventional RT calculations, for a predefined pair of endpoints (*e.g.*, TX and RX), the RF signals are aimed in all directions, usually with fixed angular spacing. The RF probes can be used to derive details about the paths between any such pair of points and, hence, determine the exact locations where interactions with smaller objects (thereby incurring changes in the calculation of path length and phase) will occur. The approach is repeated for each subsequent pair of TX-RX points in the calculation as part of a complex scene deconstruction exercise. The data set for this work is processed using the proposed methodology and was acquired in a controlled laboratory environment. The task can be attempted convincingly with the approach suggested in this work, augmented by space-based signal coverage models driven by the need for enhanced situational awareness.

The topical contributions identified in the proposed work are as follows:

1) A comprehensive characterization has been conducted on the channel-defined scatter centers in the mid-band 5G (and the future 6G) spectrum with the deployment of the PALE-on-EDGE computing resource. In the real world, there would be multiple such mobile communication node equivalents to help realize a deconstruction tomography of the propagation channel. The proposed technique has a potential future use case in passive RaDaR detection technology.
2) The work projects the PALE pre-processing pipeline approach as a viable alternative to the mainstream and resources-demanding ML/DL algorithms, given its tremendous ability to tame the variations in C-band spatial-temporal big data, resulting in a much diminished relative carbon footprint for the proposed activity.
3) Identification of distributed EDGE acquisition and computation resources, much-touted in recent times as a ubiquitous technology tool - in a first-of-a-kind quantitative way, as the most suitable solution in place of conventional time- and cost-extensive facilities such as the anechoic chamber for location and operation of mobile and base stations in dense, urban environments.

The article is organized in the following manner: a report on the prior art study and an extensive discussion on the experimental setup and hardware specifications are made, leading to the details about data acquisition and specifications. A description of the topical pre-processing algorithms that adopt various clustering techniques to analyze the spatial-temporal data is given. The salient interpretations of prominent scatter signatures are then obtained based on the progressive understanding gained from deploying the special algorithm on the data set, presented in extensive results and discussions. The benefits of the proposed scheme are laid out based on analyses involving a wider gamut of parameters identified to rate its performance against the conventional methods and include the relative carbon footprint during computation. The paper concludes on a poignant note by identifying potential impacts that this research work shall have across a wide gamut of 5-6G technologies related to antennas and RF system design, deployment and operations.

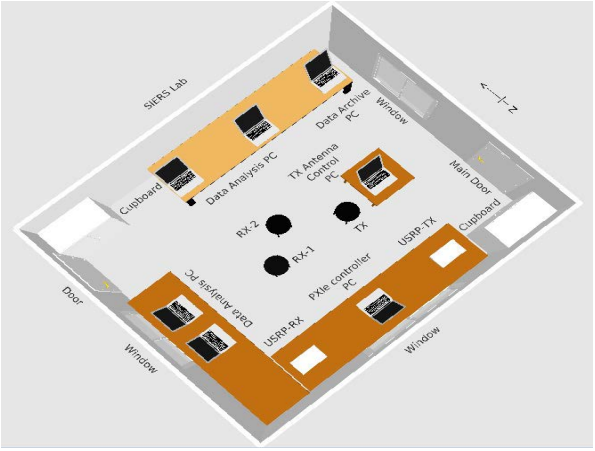

FIGURE 1. Indoor floor map of the SIERS research laboratory.

III. METHODOLOGY

A. EXPERIMENTAL SETUP

The indoor floor map of the laboratory is depicted in Figure 1. It gives an overview and layout of the test

equipment, *vis-a-vis* the diverse range of scattering centers such as furniture and civil infrastructure. Their role as predominant scatter centers is based on results that analyze the experimental data obtained in a generic, representative physical design of the laboratory, akin to a practical mid-band 5-6G signal propagation scenario.

As three deployed units of the type VERT2450, omni-directional linear dipole antennas [35] have dual-band frequency responses (2.4-2.48 GHz and 4.9-5.9 GHz), and a gain of 3 dBi, was used in this experimental setup. One antenna was assigned to the transmit (TX) side, and two more antennas were with the receive (RX) channels. One receiver-side antenna (RX-1) was placed in direct LoS with the transmitter antenna, while RX-2 was placed at variable angles (θ) to the bore-sight alignment.

The IT technique [33] has been applied with varying levels of success in the conventional antenna test and analysis process to estimate the influence of the ground plane reflections. This fundamental idea has been adopted in the current work, with the stark difference being replacing the large number of such transmitters defined along the ground plane with the replaceable off-ground receivers. The radiation scattered from spatially distributed objects in the channel would be effectively detected due to this tomographic procedure. The antenna arrangements in the experimental setup shown in Figure 1 were used to detect and acquire spatially varying signals from preferential scatterers at different locations. The off-boresight angle (θ) was changed from 15° to 60° in increments of 15°.

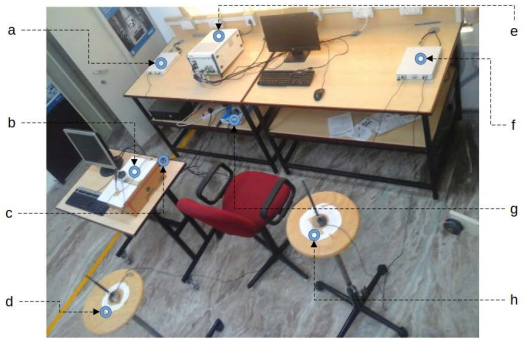

FIGURE 2. Experimental setup for C-band tomography: (a) TX-USRP; (b) TX antenna and stepper motor mount; (c) Arduino driver board; (d) RX-2 antenna and movable platform; (e) NI PXIe chassis with controller and USRP interface; (f) RX-USRP; (g) R-Pi SBC as the EDGE computing platform; (h) RX-1 antenna and the boresight-aligned platform; SMA connectors and cables are used to enable a link across the relevant device pairs.

The TX-RX arrangements used in the experiment, shown in Figure 2 (referenced to the South-East section of the lab as in Figure 1), employ the universal software radio peripheral (USRP) device as a software-defined radio (SDR) platform to provide an integrated hardware and software solution for rapidly prototyping the high-performance wireless communication system. Each of the two National Instruments' (NI) USRP-2954R [36] is a tunable RF transceiver in the frequency range 10 MHz to 6 GHz, with a frequency step less than 1 kHz. The Frequency accuracy is based on a GPS disciplined oven-controlled crystal oscillator (OCXO). The USRP has a total of 2 Tx and 2 Rx channels, and can realize a maximum instantaneous real-time bandwidth of 160 MHz. The maximum I/Q sample rate is 200 MS/s, and the 16-bit DAC has a spurious-free dynamic range (SFDR) of 80 dB. On the Tx side the maximum output power (Pout) can range from 50-100 mW, while the Rx side has a maximum input power (Pin) of -15 dBm and a noise figure of 5-7 dB. The USRP device has a large Xilinx Kintex-7 (410T) FPGA in a half-1U rack-mountable form factor. The Kintex-7 FPGA is a reconfigurable LabVIEW FPGA target that incorporates DSP48 coprocessing for high-rate, low-latency applications, with a flexible hardware architecture and the LabVIEW unified design flow [37].

The NI PXIe-8374 is the PXIe-MXI Express Interface Kit for connecting the two USRP hardware units to the PXIe-1082 chassis. The latter has 8 reconfigurable add-on slots and features a high-bandwidth backplane to meet a wide range of high-performance test and measurement application needs. The overall system is controlled using a PXIe controller that operates in the LabView environment. The NI PXIe-8880 PCI eXtensions for instrumentation (PXI) Express/CompactPCI Express embedded computer [38] is a high bandwidth system controller that integrates standard I/O features in a single unit by using state-of-the-art packaging. Combining an NI PXIe-8880 embedded controller with a PXI Express-compatible chassis, such as the NI PXIe-1082 has resulted in a fully PC-compatible computer in a compact, rugged package. By taking advantage of PCI Express technology in the backplane, PXI Express increases the available PXI bandwidth from up to 132 MB/s to up to 48 GB/s for a more than 60x improvement in bandwidth. The standard I/O on each module includes one DisplayPort 1.2 video port, four high-speed USB 2.0 ports, two high-speed USB 3.0 ports, a PCI-based GPIB controller, two Gigabit Ethernet connectors, a reset button, and PXI Express triggers. The NI PXIe-8880 is a modular PC in a PXI Express 3U-size form factor, and has an Octa-Core Intel® Xeon® E5-2618L v3 processor, triple channel DDR4, 1866 MHz memory, all the standard I/O, and an integrated sold-state drive.

NI's hardware and software work together to facilitate the PXI Express communications system. The LabVIEW application development environment combines with leading hardware drivers such as NI-DAQmx to provide exceptional control of the NI USRP hardware. LabVIEW is a powerful and easy-to-use graphical programming environment for acquiring data from several different instruments interfaced via the various ports in the PXIe-8880, enabling a meaningful convertion of the acquired data into vital results using powerful data analysis routines.

The TX-RX system design has been integrated as an SDR, with the application layer defined in the Communications System Design (CSD) suite [39]. It generates and transmits a single tone 100 kHz sine wave signal over a bandwidth

of 400 kHz in the C-band frequency (5.25 GHz). To obtain comprehensive coverage for the transmit antenna radiation pattern, TX was made to rotate about a vertical axis as part of the acquisition process [40]. This automated scheme employs an Arduino-driven stepper motor [41], controlled asynchronously by a Raspberry Pi-2 (Model B) single-board computer (R-Pi SBC) [42].

Spatial data at enhanced granularity was acquired by spatially sampling in angular steps of 0.698°, resulting in 516 discrete positions (angle steps) over the 360° in the elevation (E) plane. The data acquired at each angle step had eight snapshots comprising 20,000 data points. Similarly, five independent voltage-time series data sets for the various θ were obtained and subjected to further analyses.

For a root mean squared (RMS) value of the transmitter voltage (V_{rms}):

$$V_{\text{rms}}^2 = V^2/2 = 0.25 \text{ V}$$

and with the transmitter voltage V being 0.707 (V), the transmitter power is then given by [32]:

$$P_g = V_{rms}^2/R = 5 \text{ mW} \quad (1)$$

Here, the radiation resistance on the transmit side is R = 50 Ω. The power radiated by the transmitter antenna would then be:

$$P_t = g_t/g_{t(\max)} \times P_g = 315.6\mu\text{W} \quad (2)$$

where g_t = 20 dB is the gain of the transmitting side amplifier, and the maximum gain $g_{t(max)}$ with which the USRP transmits is 32 dB, when there is a full radiated power of 100 mW. The magnitude of measured power across a distance of 0.5 m is typical of the C-band far-field measurements observed in this work. In practical field situations, with mid-band distances spanning 500-2000 m from a 5-6G base station (BS), the received power at the various 5-6G-enabled MDs is constrained by variables such as P_t, the radiation efficiency of the antenna, g_t, g_r, frequency and the channel-driven propagation factor and path loss models [43].

B. SYSTEM-LEVEL SIMULATION

Deconstructing a practical C-band communication system starts with simulations that employ codes developed using the LabView CSD Suite. These software codes have been deployed to generate raw time-series data representing an antenna's radiation pattern. The codes help implement the transmission of a sinusoidal message tone transmission at 100 kHz modulated using a 5.25 GHz carrier over free space and later retrieve the message tone through two receiver antennas with spatial diversity. The LabVIEW software codes enable an effective integration and control of the USRP communication system to realize a robust transmission and reception process.

The data acquisition system consists of the Transmitter and the Receiver-cum-Data processor subsystems as the primary blocks. Besides this, the other functional blocks support the primary ones by deploying different validation metrics such as RX frequency validation, RX bit error rate (BER) validation (derive robust inferences from the eye diagram), and radiation pattern correlation coefficient. A qualitative confirmation is obtained from the radiation pattern plot and the received frequency histogram. The carrier signal frequency corresponds to the resonant frequency of the transmit and receiver antennas. The USRP hardware defines a minimum sampling rate of 390-kilo samples per second (ksps); a more reasonable sampling rate of 400 ksps has been considered during the data acquisition phase.

The sampling rate has to be set per the theoretical definition of the Nyquist criterion. For a message signal bandwidth of 200 kHz to be detected alias-free, a sampling rate of 400 ksps shall suffice. The sampling is done at a much higher rate to improve frequency resolution and signal-to-noise ratio [44]. Oversampling by a factor of 2 has been applied here [45], and this results in a message signal frequency that is a quarter of the sampling rate, *i.e.*, 100 kHz. The USRP hardware allows the transmit and the receive gain to be varied from 0 dB to 31.5 dB, corresponding to a transmit power range of 50–100 mW. The transmit and receive gain parameters were chosen as 10 dB based on careful observations during signal acquisition. The experiment requires 10 GB of storage size without discarding the data, further reduced to 2 GB. Storing 20k samples allows for a frequency resolution as low as 400k/20k = 20 Hz [46]. The read and store process consumes 2 s, referred to as the dwell time – the halt duration in the transmit antenna orientation at a specific angle. After the fixed dwell time lapses, the transmit antenna is rotated through a further 0.697° about a fixed vertical axis to the next angular position. Although it is desirable to decrease the dwell time, optimizing the total duration of data acquisition (to accommodate any potential variability in the ambient parameters), lowering below 2 s impacts the quality of radiation patterns recorded. The quality aspect was inferred from the signal-to-noise ratio (SNR) estimated during raw data processing. The primary data was acquired at $f_c = 5.45$ GHz, and a periodic switching of the receiver to 2.45 GHz was done to obtain data as part of a frequency diversity analysis at angles of 0°, 90°, 180°, and 270°. The read-and-store process was repeated for the remaining angles until the transmit antenna completed a full rotation.

C. PORTING THE DATA

MATLAB R2019a was used for data processing and analysis in this research work. The five independent data sets employed in this analysis each had a size of 1 GB and comprised 4096 rows and 20,000 columns. The EDGE computing hardware for data processing was the R-Pi SBC with a Quad Core 1.2 GHz ARM Cortex-A53 64-bit processor with 1 GB RAM; the Broadcom VideoCore IV at 250 MHz GPU; and Debian 9 version of the Linux kernel 4.19 based R-Pi operating system.

D. DATA PRE-PROCESSING

The received power was determined by extracting the magnitude of power at the frequency of the message signal from a power spectrum of the voltage-time series data.

The reference data regarding antenna radiation patterns were obtained by consulting the manufacturer's datasheet. To match the nulls of all the possible spatial configurations with the reference data set, the angular position at which there was a minimum detected power was identified [47]. The successive data points were then shifted to coincide with the minimum power transfer angle, as depicted for the case of a 60° off-boresight radiation pattern in Figure 3, with 0° assigned as the reference boresight alignment axis. In all further analyses, this angle-shifted data was used.

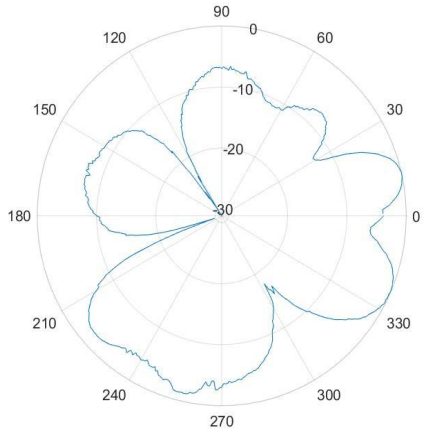

FIGURE 3. Off-boresight detection at $\theta = 60°$ for boresight-aligned antenna radiation pattern.

The theoretical value of received power is determined using the well known Friis radio link (FRL) equation [48]:

$$P_{r\,(\mathrm{th})} = P_t G_r G_t \left(\frac{\lambda}{4\pi r}\right)^2 = 0.26\mu\mathrm{W} \tag{3}$$

Here, P_{t} is determined from equation (2) to be 260 nW, with $G_{\mathrm{r}} = 10$ dB as the gain of the receiver antenna, $G_{\mathrm{t}} = 1$ dB as the gain of the isotropic transmitter antenna, $\lambda = 0.0571$ m being the wavelength of the transmitted signal, and the far-field distance between the transmitter and receiver is determined to be $r = 0.50$ m.

The received power was converted into the decibel scale to obtain $R_{\mathrm{i(exp)}}$ [49]. In effect, this is the difference between the experiment-based value for signal power determined from the power spectrum of the C-band signal ($P_{\mathrm{i(exp)}}$) and the theoretical value of the received power as defined by the FRL equation ($P_{\mathrm{r(th)}}$) [43].

E. FINDING RESIDUALS

The manufacturer's data sheet was consulted for the reference data regarding antenna radiation patterns. The residuals were determined and subjected to further analyses to detect and characterize the scatterers. Residuals indicate the degree of variation in the experimental data to the reference data. Such residuals are computed by:

$$R_i = R_{i(exp)} - D_{\mathrm{ref}}\,(m) \tag{4}$$

The first term on the right is the experimental received power value in dBi, and the second corresponds to the reference power value obtained from the antenna manufacturer's datasheet. The variation observed in the residual values is primarily attributed to the influence of scatterers present in the propagation channel.

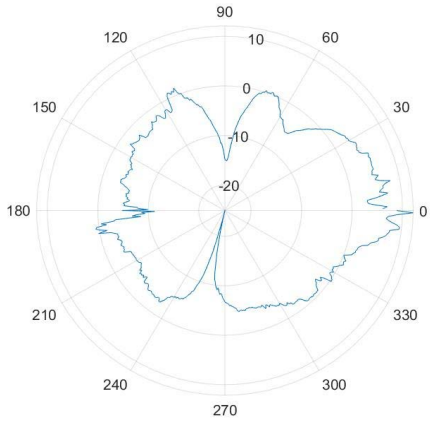

FIGURE 4. Residual plot for off-boresight detection at $\theta = 60°$.

Polar plots, like the one shown in Figure 4 for the case of a 60° off-boresight orientation in RX-2, depict the angular distributions of all such residuals. Five more residual plots were generated in a polar reference frame, eventually revealing the angular distributions and characteristics of the scattering centers in terms of variability in the magnitudes of such residuals.

IV. METHODOLOGY

A. DATA SETS

Two different ways were identified for analysis of the experimental data:

1) By taking the complete 4128 residual data per data set while considering all eight snapshots for each spatial position at once (516 spatial positions × eight snapshots per spatial position = 4128 residual data points per data set).
2) Residual data with 516 points by taking one snapshot for all the spatial positions (516 spatial positions × one snapshot = 516 residual data points). There are eight such snapshots per data set. For simplicity, the analysis considers every 4th snapshot in each dataset.

In all subsequent analyses, approach (1) was chosen to preclude any loss of information in the experimental data.

B. CLUSTERING ANALYTICS

An exercise classifying scatter centers is defined fundamentally based on their interaction with the C-band signals. The scattering centers' primary classification is either an absorbing or a reflecting type. While the former would diminish

the signal amplitude, the latter would leave a constructive attribute in the signal amplitude. In order to analyze the residuals, the K-means clustering technique has been employed as the first in a series of pre-processing tasks that would classify the residual data into 'n' clusters based on diversity in the dataset.

K-means belongs to a class called partitional clustering and has been chosen as the reference unsupervised learning method for data classification considering the diversity and size of the data being studied. The eventual outcomes of analyses would have a particular bearing on the study's specific context [50]. Besides, this data-driven algorithm is also one of the straightforward yet most robust when implemented for the dataset considered in the current work. The method aims at discovering inherent patterns in the data sets based on a measure of similarity [51].

The algorithm developer must define the number of clusters 'n' and their respective centroids. The initialization of the centroids is crucial since the algorithm is susceptible to the choice of starting points [52], *i.e.*, positioning the initial centroids causes different clustering results. The scheme adopted to address the initialization problem had been to determine the initial centroid counts, using the arithmetic mean formula, and locate them such that the distances between them is the maximum [53]:

$$q_i = \frac{1}{M}\sum_{\mathrm{j}=1}^{M} p_j \tag{5}$$

where j is the number of objects in cluster, the object values in a cluster are denoted by points p that have the coordinates (p_1, p_2,…., p_{j}), and the cluster centroids q have the coordinates (q_1, q_2,…., q_{j}). The task would be to minimize the sum of squared Euclidean distances between the points in the cluster referenced to the centroid of that cluster. The squared Euclidean distances for an n-dimensional Euclidean space are defined as [54]:

$$d^2\,(p, q) = \sum_{\mathrm{i}=1}^{n} (p_i - q_i)^2 \tag{6}$$

In order to extract all possible qualitative signatures of the potential scatterers from the dataset through clustering, it was necessary to determine the optimum number of clusters in the dataset. Applying convergence criteria such as a minimum in the data points' reassignments, centroid definitions, and decrease in sum-of-squared errors (SSE), the evident existence of more than three types of scatterers in the propagation channel had led to the choice in value of n as 3. Power and spatial domain data were considered subsequently to make the analysis even more effective. Since the centroids were initialized randomly for each value of k, to determine the optimum number of clusters, iterative analysis using K-means clustering was done with different values of k, followed by a silhouette analysis performed on the output of the K-means clustering algorithm.

1) SILHOUETTE ANALYSIS

Silhouette analysis is used to understand the degree of similarity among the various object in a cluster, *i.e.* cohesiveness, compared with those in other clusters indicative of divergence [55]. While most performance evaluation methods need a training set, the Silhouette analysis does not require a training set to evaluate the clustering results making it more appropriate as a clustering task in the present study. The values of such metrics in the silhouette analysis range from -1 to $+1$, with -1 indicating the worst possible clusters and $+1$ indicating the best possible clusters. The silhouette width for every data point in the cluster is calculated using the expression:

$$S = \frac{1}{n}\sum_{\mathrm{i}=1}^{n} \frac{(b_i - a_i)}{\max\,(a_i, b_i)} \tag{7}$$

Given a cluster, a_i is the average of distances within the clusters, and b_i is the minimum nearest cluster distance, *i.e.*, the distance to the point in the adjacent cluster. The greater an element's S-value (more positive), the higher its likelihood of being clustered in the correct group. Elements with negative S are more likely to be clustered in wrong groups, and the ensemble average within a dissimilarity is more significant than between dissimilarities. Only those clusters with average silhouette scores greater than a certain threshold are considered to have exhibited strong structures that are discernible in the input data [55].

2) ITERATIVE K-MEANS CLUSTERING

Silhouette analysis can also be used to determine the optimum number of clusters by iterating the K-means clustering analytics process for different numbers of clusters, k, say from 1 to 30. The variation of k with the average silhouette value can then be studied, and the value of k that results in the highest average silhouette value can be considered the most optimum clustering outcome.

One observation about this approach is that the centroids are initialized randomly for each value of k. The most optimum cluster is not considered for the computation of the silhouette value, and the resulting average silhouette value is not the best possible value for that number of clusters: even if clustering is repeated for the exact value of k, such as for $n = 5$, the resulting clusters are different each time. The method employed has been to iterate clustering multiple times for each value of k, say 20, and find the most optimum cluster by selecting the cluster with the lowest Euclidean distance [56]. Each time the clustering process iterates, the centroids are randomized within the range of the dataset, ensuring that all possible combinations get explored. While this iterative approach is not guaranteed to deliver the clustering that globally minimizes within-cluster variation, the algorithm nonetheless ensures that the most optimum cluster for that particular value of k is always selected, from which the average silhouette values can be computed.

3) K-MEDOIDS CLUSTERING

K-means clustering algorithm is sensitive to outliers, *i.e.* the mean centroid is pulled towards the side with the higher value of elements. Although the silhouette value obtained from the K-means clustering [56] indicates good clustering, the centroids in the clusters are not within the dataset. The K-medoids clustering approach has been used in place of the K-means clustering to correct this situation. In K-medoids clustering, objects in the datasets define the centroids instead of their mean values in a cluster [50]. The algorithm is thus immune to outliers since the centroid would always remain at the center of the dataset [57]. Instead of assuming that the scatterers in the propagation channel would belong to any one among the three categories of scatterers, *i.e.*, absorbing, neutral or reflecting, it is appropriate to conclude this property from the available dataset. The K-medoids approach, while still unable to achieve a global minimum, generally returns a higher value of the SSE at the expense of increased computation. It shall enable a robust decision on the optimum number of clusters and, hence, identify the numbers and types of scatterers in the propagation channel.

4) ELBOW METHOD

The Elbow method is one of the heuristic methods used to determine the optimum number of clusters. The within-cluster-sum-of-squares (WCSS), 'J', is a measure of the sum of the square of distances of each data point to their respective centroids. WCSS is defined by [58]:

$$J = \sum_{\mathrm{i}=1}^{k} \sum_{x \in C_i} |x - m_i|^2 \tag{8}$$

where 'x' is the element of the cluster $\mathrm{C_i}$; m_i is the centroid of cluster $\mathrm{C_i}$ (the mean vector of all the data points in $\mathrm{C_i}$); and k is the number of clusters $|\mathrm{C_i}|$. When k is less than the optimum number of clusters, the decrease in WCSS is more considerable. As k increases from $i = 1$ onwards, J progressively decreases, essentially indicative of an increase in the degree of clustering in the dataset. When k exceeds a particular value that is more than the one for the optimum number of clusters, any further decrease in WCSS is diminutive, and there is no further improvement to the degree of clustering, indicative of a convergence. The WCSS characteristics graph shows a distinctive feature that resembles an ''elbow''. The value of k for this reasonable decrease in the reduction of WCSS is considered the optimal number for a given cluster. The value of k at the 'elbow' is taken as the optimum number of clusters [59], and the iterative analysis is considered complete.

5) CORRELATION COEFFICIENT

In order to understand the variability of clustering across the angular off-boresight datasets (x_i for $\theta = 15°$, $30°$, $45°$ and $60°$) and to correlate them with the clusters in the bore-sight dataset (y_i for $\theta = 0°$), the correlation coefficient has to be determined. The correlation coefficient, Γ, is a measure of the linear relationship between these pair of variables (x_i and y_i) and is defined by:

$$\Gamma_{xy} = \frac{\sum (x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{9}$$

The mean of the variable pair is denoted by and, and Γ values range from -1 to $+1$. Equation 9 can also be defined as the covariance ratio in a pair of parameters to the product of their respective standard deviations in a given sample size. A value of 0 indicates no linear relationship; a value of $+1$ indicates a strong positive linear relationship, while a value of -1 indicates a strong negative linear relationship [60].

The correlation coefficient among the k clusters in each of the angular datasets (taken separately) vs the k' clusters of the boresight dataset would result in a ($k \times k$') matrix. Each cell in the matrix corresponds to a correlation coefficient value that is indicative of the degree of linearity that exists between a pair of clusters: *e.g.*, the value in the fifth row, second column cell, corresponds to the correlation coefficient between the fifth cluster in the angular dataset and the second cluster in the boresight dataset.

A histogram plot would have to be generated to visualize the distribution in the correlation coefficients with a careful choice of the bin size to enhance the resolution of the histogram. The correlation coefficient values that fall within the bins were compiled, and the histogram was plotted for each dataset.

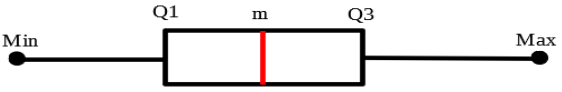

FIGURE 5. Box plot and its five-number summary.

6) BOX PLOTS

Box plots are suitable for visualizing the range and level of data and highlighting outliers, often used in explanatory data analysis. The ''box'' shows the inter-quartile range (IQR), with the ''whiskers'' representing the outliers. A generic box plot is a standard way of representing data based on a 5-number summary [61]. This 5-number summary derived out of the box plot, such as the one shown in Figure 5, consists of a minimum (Min); the first quartile (Q1: a quarter of the data is less than this value); the median (m: half of the data is greater than this value); the third quartile (Q3: a quarter of the data is greater than this value); and the maximum value (Max) of the given data set. With outlier data points up to a distance of 1.5 referenced to the respective quartile, box plots give a fair indication when comparing the distributions of several data groups since they summarize the center and spread of the data effectively. The size and degree of symmetry (skewness) of a box plot are directly reflective of the diversity in the data values: the closer the values, the smaller the box size, and *vice versa*.

The data would be quantitatively analyzed, in terms of its variability in received power and angular distribution, based

on two approaches employed in generating the box plots. A further consolidation of the results, based on derived trends in the 5-number summary, was performed as a culmination of the pre-processing efforts. The process also helped convincingly demonstrate the presence of regions in angular space that would be the most suitable (or otherwise) for locating the BS (and MD) units while defining a validated topology for operational mid-band 5-6G deployments.

TABLE 1. Clustering pipeline and context.

Algorithm type	Estimation parameter	Application in data set
K-means clustering	Squared Euclidean distance	Cluster count optimization
Silhouette analysis	Silhouette index	Agglomerations
Iterative K-means clustering	Nearest centroid	Least Euclidean distance in cluster
K-medoids clustering	Cluster center; partitioning around medoids	Intra-cluster distance minimization
Elbow method	Within cluster-sum-of-squares (WCSS)	Cluster count at elbow
Correlation analysis	Correlation coefficient	Inter-cluster linearity
Box plots	5-number summary	Box attributes & deviations

Table 1 summarizes the select group of well-established PALE contenders employed in a pipeline processing scheme. The table lists the seven topical clustering schemes that had been identified as the most suitable pre-processing models to operate on the data. In the lead up to generate a comprehensive profile of scattering centers, an overview of each of the respective constraining parameters and their application paradigms in the data set is identified; the last column has their citations to their prior art. Analyses of the extensive C-band spatial-temporal datum, with the view of efficiently classifying their residuals, had used this algorithm group.

V. RESULTS AND DISCUSSION

A. RESIDUAL PLOT

A qualitative differentiation among the scatterers in the propagating channel was performed on the residual data set using K-means clustering analysis. The number of clusters, 'k', was set to 3 to classify the scatterers as belonging to one of the three types, *viz*. absorbing, neutral or reflecting varieties.

The clustering analysis was initially confined to the power domain for the time-series data while setting aside information belonging to the spatial domain 'θ', understanding that scattering signatures have a topical representation in terms of the detected power variability. Most small magnitude variations (below 5σ) referenced to the ideal data set are explained due to the sensitive nature of the detection process. The presence of random noise in the neighborhood of the experimental space thereby gets detected, *e.g.* interference from MDs. All such data points were discarded from the further pre-processing stages. Beyond this point, by choosing the number of clusters to be 3, the average silhouette value for the clusters was determined to be 0.81544, which is a clear indication of robust clustering in the data set. Since only the power domain was given as the input to the clustering algorithm, the algorithm splits the data set into three partitions in the respective domain. Such distinct groupings as outcomes of residuals analysis are interpreted as qualitatively representing the characteristics of scatterers in the propagation channel.

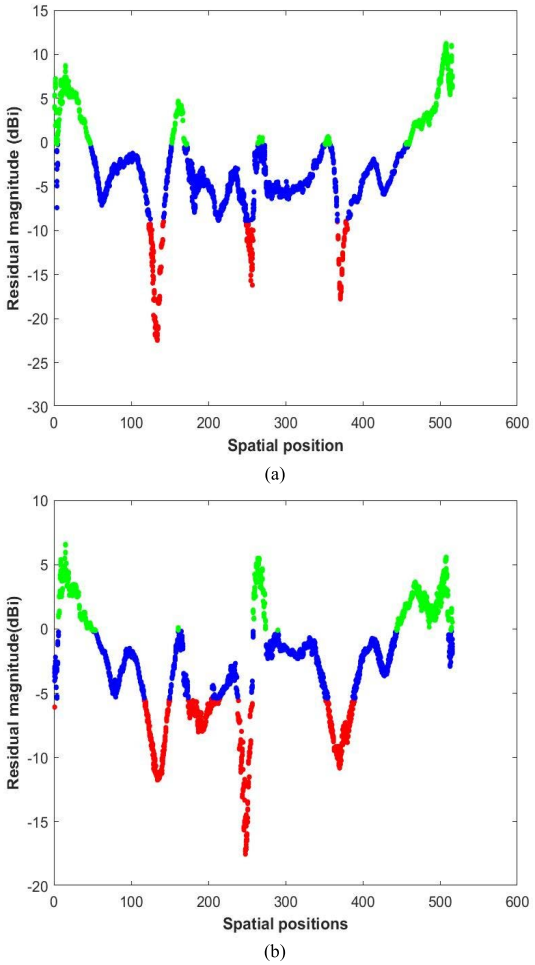

FIGURE 6. Residuals grouped in data from: (a) the LoS antenna (RX-1); and (b) the RX-2 antenna at θ =15°.

The following observations have been made, referenced in Figure 6:

1) The clusters represented by the color code blue are distributed across the data set. Data points in this cluster are distributed evenly and do not show extreme variations. This trend indicates that these residual types do not constitute a definitive signature left by the scatterers of concern in the propagation channel. Their low-profile and ubiquitous presence is discounted from any further active study.
2) The clusters represented by the color code red are data points in which the received signal power values are

significantly low compared to that of the reference data set. This set of points in the cluster arises due to absorbing scatterers which decrease the incident signal magnitude upon their interaction. Thus the power intercepted by the receiver antenna is also of a reduced scale.

3) The clusters represented by the color code green indicate the scattering at spatial positions from where the received power is relatively high compared to that in the reference data set. This set of points in the cluster arises due to specular-type scatterers that act as a mirror and bounce the incident signal off, resulting in the higher power levels intercepted by the receiver antenna from those directions.

As a further stage in this iterative analysis, both the power domain and the 'θ' domain values were offered as simultaneous inputs to the K-Means classification algorithm. The clusters were initially randomized, and clustering results differed each time the algorithm was invoked. The algorithm had partitioned the data only in the power domain during every subsequent iteration. Contrasting such outcomes with the clusters observed in Figure 6, where the variations noticed in the latter case of output clusters were primarily attributed to the random initialization of centroids, as discussed in Section IV-B. Thus, this method is found to be inconclusive in precisely defining clusters. However, partitioning the data in the power domain is a notable outcome of the qualitative analyses related to the property of scatterers in the propagation channel.

B. K-MEDOIDS CLUSTERING

K-medoids clustering is an iterative analysis approach used to determine the most optimum clustering configuration from among the residuals. The silhouette score for each value of 'k' was recorded based on an iterative analysis. Silhouette plots were then generated for each data set, corresponding to the various other off-boresight angles, θ, varying from 0° to 60°, obtained at 15° intervals. As shown in Figure 7(a), with $k = 3$ and for $\theta = 15°$, 60° and 0°, the first peak silhouette score was found. However, taking $k = 3$ is in broad agreement with the earlier interpretation of only three types of scatterers in the propagation channel, which is likely inaccurate.

The first peaks in the other data sets, *viz.* of $\theta = 30°$ and 45°, are observed for $k = 6$, as seen from Figure 7(b). This value is appropriate for the optimum number of clusters, as it also agrees with the presence of 6 possibly distinctive types of scatterers in the propagation channel. However, this value of k cannot remain invariant for all the data sets, given the significant dip in the respective silhouette value, which indicates a limitation in the clustering approach. Therefore, with a threshold having been defined as 0.71 for the average silhouette values, which, when exceeded, would count towards success in determining the attributes of solid structures in the input data [54], the optimum number of clusters was hard to determine. The number characteristics showed wild fluctuations while their optimum choice needed to maintain

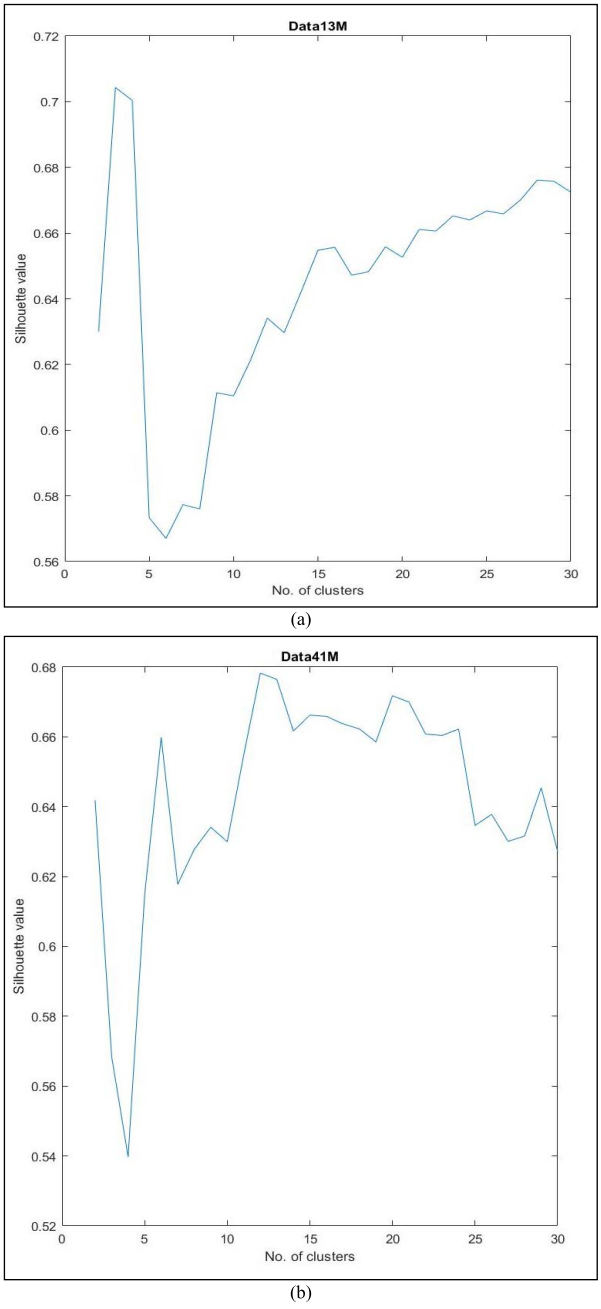

FIGURE 7. Variation in silhouette values with cluster count for: (a) $\theta = 15°$; and (b) $\theta = 45°$.

a constant value. The experiment was conducted under a near invariant laboratory setup, *i.e.*, amid a steady number of scatterers in the propagation channel. Having to cluster each data set with different silhouette values every time is non-ideal because the scatterers had remained spatially agnostic to the location of the transmitter antenna during the entire operational phase. From all such observations, a near-total lack of any discernible trend was noticeable with the increasing number of clusters. Hence, the silhouette analysis could not determine a definite optimum number of clusters. With vital insight derived from this method, an advanced methodology, *viz.*, the elbow analysis, was then employed for determining

the optimum number of clusters and precise classification of the scattering centers.

C. ELBOW METHOD

The elbow method involved the computation of the WCSS parameter for each value of k. A plot of the number of clusters vs WCSS is generated for each off-boresight data set. A precise selection of the point of the elbow is crucial here since it would be an indication of the location beyond which there shall be no further significant improvement to WCSS. The gradient of the plot starts to flatten out beyond such a point, and performing the clustering analysis on the data set beyond such a point that, *per se*, is indicative of the optimal number of clusters shall lead to an increase in the computational complexity with no significant rewards. It has thus, eventually, become possible to derive the optimum number of clusters from the elbow method derived plots. The elbow analysis scheme has possibly determined the optimum number of clusters, with the elbow set between 6 and 9 for all data sets, *i.e.* the set of points of the elbows is {6, 7, 7, 8, 9}, as seen in Figure 8 for two different cases of angular pointing.

The elbow method uses quantitative formulations to explain the variation in WCSS with the number of clusters, and inference about the point of the 'elbow' is subjective. The optimum number of clusters was found to be 7, determined from similar elbow plots for the various other off-boresight angles varying from 0° to 60°, obtained at 15° intervals. Hence, a median value of 7 in all further analyses would not affect the clustering as there is no significant change in the WCSS. Moreover, since the interpretation of the elbow is qualitative, a value of 7 for the optimal number of clusters is acceptable. The number of clusters is enhanced by a factor of 2, *i.e.* to 15, which can accommodate a potential scale-up in the number of power levels. The attended classification accuracy can ably handle complex propagation scenarios. A concomitant increase in the detection accuracy for several diverse scatterers in the propagation channel is also enabled. Each leaves a unique signature defined as part of its interaction with the RF signal that allows an enhanced estimation ability of its scattering characteristics.

An increase in the optimal number of clusters to 15 results in an observed decrease in the value of WCSS by 50%, as inferred from the various elbow plots. The individual clusters would thereby have a smaller footprint in the clustering plot compared to clusters partitioned using $k = 7$. It is necessary to increase detection resolution due to the high degree of variability in the scatterers located within the laboratory. Increasing the resolution of the clustering scheme is beneficial to the analysis due to the enhanced ability offered during the detection of a broader spectrum of spatial scatterers that impact the signal propagation, even in the case of a single off-boresight data set.

D. CORRELATION COEFFICIENTS

The clustering was done using $k = 15$, and the correlation coefficients between such clusters were determined as discussed in Section IV-5. The distribution in correlation coefficient values was studied from the related histograms. Correlation coefficient values within such windows are aggregated in the histogram plots, with the bin size set as 0.05. The statistical parameters derived from the histogram distributions, in all such cases of off-boresight angles varying from 0° to 60° at 15° intervals, are listed in Table 2.

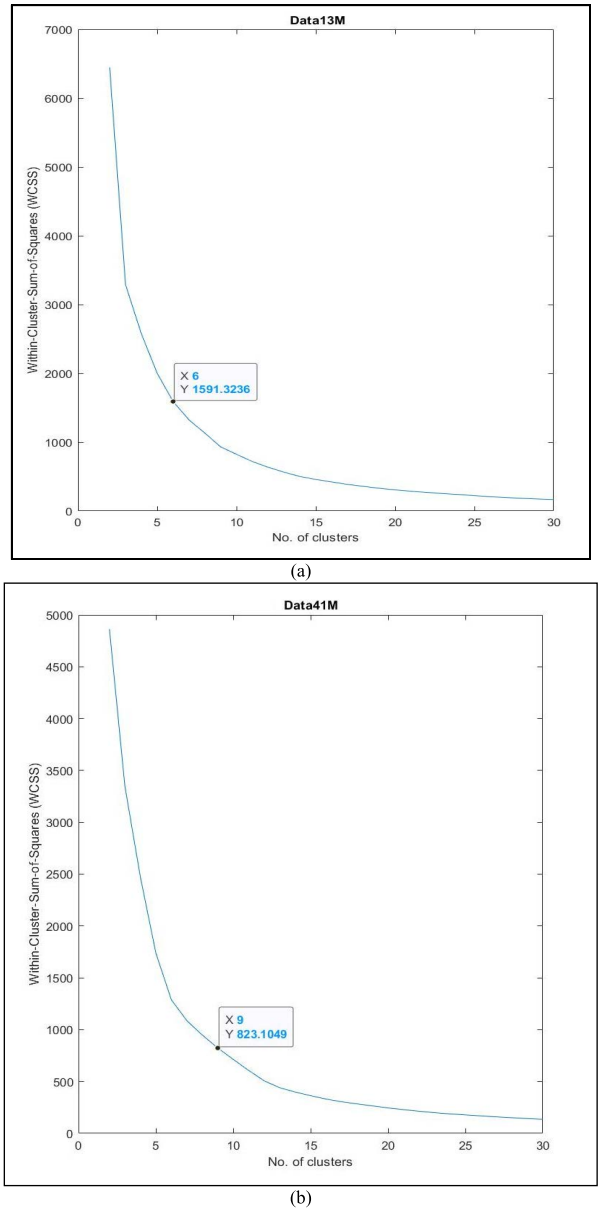

FIGURE 8. WCSS vs. cluster count for θ: (a) 15°; and (b) 45°.

TABLE 2. Histogram statistical parameters.

Angular data set at:	Mean	Median	Standard Deviation
θ = 15°	0.0026	0.0049	0.1180
θ = 30°	0.0091	-0.0089	0.1052
θ = 45°	0.0073	0.0017	0.1025
θ = 60°	0.00097	-0.00053	0.1083

The observation is that the number of instances where the clusters show no correlation is significantly higher. The inference from such an outcome is based on the fact that the correlation coefficient values between the clusters shall be increased when clusters with similar points are compared, which, in this case, is very unlikely given the random nature of K-medoids clustering. In all of the other spatially distributed data sets, the correlation coefficients shall be very nearly 0. The value corroborates the well-known understanding that any noise component is primarily uncorrelated at the RX. The correlation coefficient distribution study is limited mainly to detecting linear relationships between a pair of data sets. The absence of correlation coefficient values beyond a $\pm$ 0.4 indicates that no clusters from the angular data set precisely match those of the boresight data set, and no outliers exist in the data sets.

E. BAR CHART

With the primary elements of PALE having indicated a definitive profile in the spatial-temporal data, the inconclusive outcome of the correlation study needed an alternative approach to constrain the data and result in a standard classification. Data points belonging to similar clusters were grouped across all 5 data sets. The spatial distribution of the cluster and their rates of occurrence were visualized through the bar charts. The peaks in the bar charts of Figure 9 indicate the role of known scatterers at specific spatial locations.

The inferences derived from such results indicate the dimensions and locations of prominent scatterers in the laboratory environment, as defined by the layout in Figure 1. It is also observed from these plots that there are spatial distributions associated with signals of uncatalogued types, inferred from the presence of the smaller peaks. These could be reliable indications of unclassified ''targets''. Such types occur due to human interference in the operating environment, the effect of various other electronic devices or the movement made to a few other small objects such as chairs and tables during the experiment within the laboratory confines.

The spatial distributions of the test environment's absorbing- and reflecting type scatterers that have resulted in effects such as signal fading and enhancement, respectively, could be identified from the bar charts in Figures 9 (a and b). The spatial positions were then transformed into the corresponding angular coordinates, considering the relevance of the spherical reference frame to the signal propagation, acquisition and analysis.

The exercise in coordinate space transformation has helped with discerning the location coordinates in terms of a correlation with the equivalent physical dimensions and material characteristics of prominent scatterers; they include metallic cupboards, workbenches, window grates and doors that had already been defined in the floor plan of the laboratory (as depicted in Figure 1).

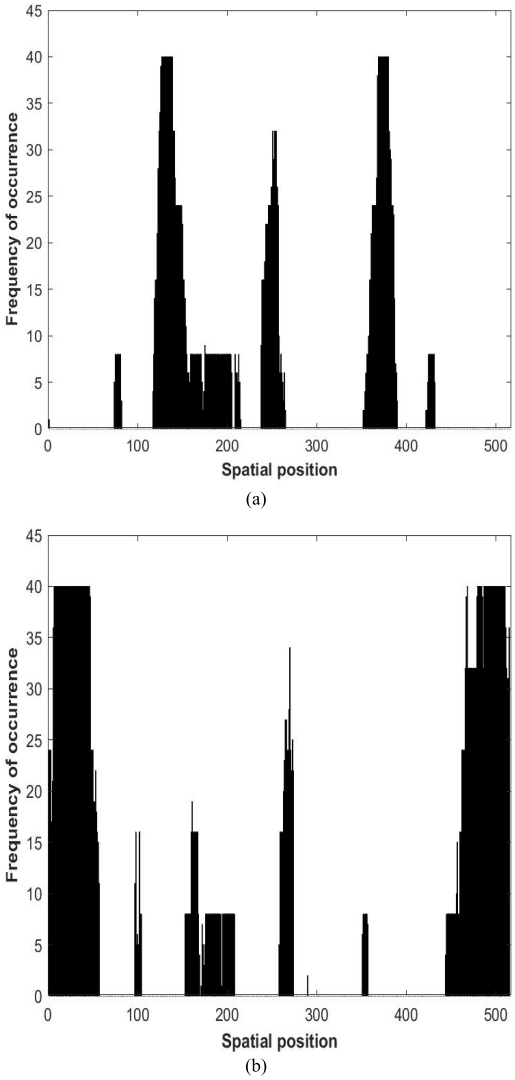

FIGURE 9. Spatial occurrences of scatterers with properties of: (a) absorption; and (b) reflection.

F. BOX PLOT

Inferences from kernel analytics and its associated methods are severely influenced by the cluster size and centroid definition, while histogram analysis is affected by bin sizes. Hence, for a conclusive inference, a refinement that uses box plots in quantitative analytics has been incorporated into this work. The data was analyzed quantitatively by generating two different box plots. Every row in the experimental data series was divided into 19 separate blocks, each consisting of 1024 data points; the power spectrum analysis helped determine the power values.

Two different data processing approaches were followed:

1) To understand the variance of power values over each step, a box plot was generated with the data corresponding to the exact angles. With each angle comprising eight snapshots, the combined yield is $19 \times 8 = 152$ power values for each angle. The

resulting box plots contained eight boxes, and each box indicated variation in these 19 values.

2) To understand the variance across all possible angle steps, a box plot was generated containing 516 boxes, where each box represents 19 values for that particular snapshot. A similar process was repeated for every snapshot, which yielded eight box plots.

Salient indicators regarding the morphology (geometry and material composition), location (range), and temporal variability (attenuation and interference) of the scatter centers (as shown in Figure 1) can be inferred with remarkable accuracy from the 5-number summary obtained using box plots. In all 5 cases in θ, similar plots were generated. The RX data identified the minimum and maximum values with the respective extremities that lie beyond the boxes. The median spatial positions were determined for each data interval and used to generate the box plots.

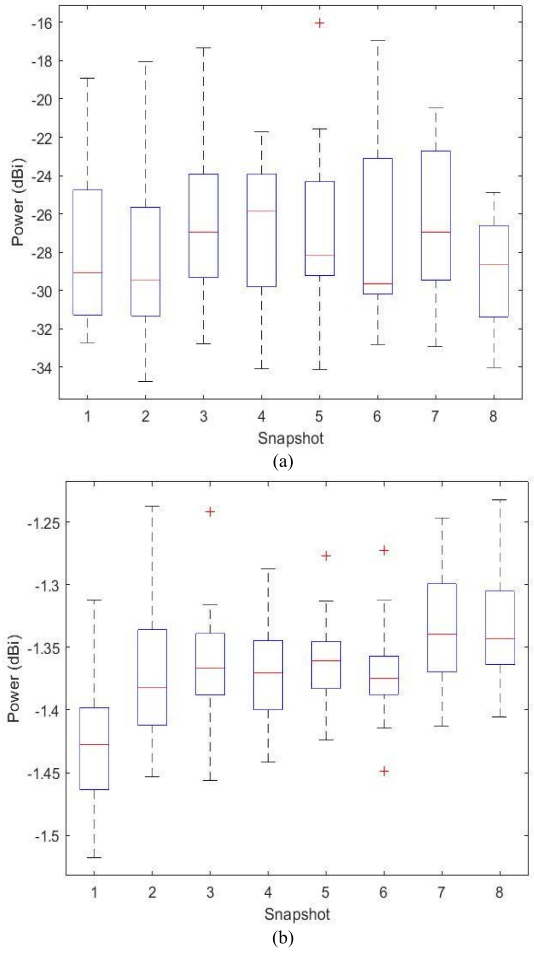

FIGURE 10. RX power levels defined across 8 snapshots: (a) the 251th; and (b) the 210th spatial positions.

A representative plot for the median value in the case of spatial position 251 that is associated with the TX: RX-1 geometry (Figure 2), shown in Figure 10(a), depicts a higher variation in the radiated power (Type-1). In marked contrast are the ones in Figure 10(b), which relate to the 210th spatial position (angle step) in the TX: RX-1 data (Type-2) and exhibit lower-order variability. Presence of outliers has been indicated with a '+' symbols in each case. The box plots corresponding to the various other RX-2 positions also closely resemble the Type-2 case across all five experimental data sets, indicating that there has been a negligible effect due to the known scatterers on the propagating signal in a majority of cases. The range of the boxes is also relatively low, as can be inferred from the ordinate values in this plot.

In the select instances that conform to Type-1, the scatterers' preferential location and location-specific attributes are evidenced by the reliable signature of a tell-tale scattering phenomenon noticed at specific angular ranges in the spatial distributions. Potential candidates include electrically conducting and dielectric types: such as the window grills and cupboards for the former and doors for the latter. In the practical field studies, these shall translate into a diverse assembly of composite and metallic civil structures, vehicles, and organic constituents within the cell served by FR1. The snapshots were taken at the exact spatial locations and within a minimal time interval as a further proof. Yet, the data sets exhibit significant variations in the received power.

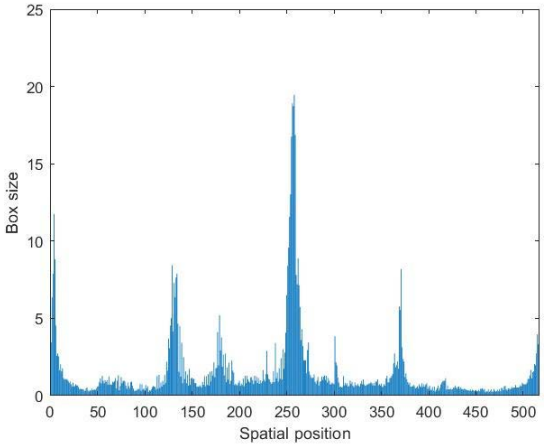

FIGURE 11. Variation of box size as a function of the spatial positions.

A further step in the analysis was to consolidate and accentuate the widespread trends in several such box plots. The specific attributes of the boxes, derived from the several plots similar to the one shown in Figure 10, could then be related to the non-LoS arrival paths of RF signals that correspond to the physical characteristics of the prominent scatterers in the propagation channel. The results of this consolidation effort become immediately apparent, as has been inferred from Figure 11, in that any relative variation is insignificant at positions with no scatterers (Type-2). At the same time, significant variabilities are also equally evident at the exact locations of the scatter centers (Type-1), whose locations have been identified in terms of their spatial positions. In this final segment of the study, specifics in the box plots were directly ascribed to the diversity of the RX data values; *i.e.*, the more tightly clustered these values are, the smaller the

box size, and *vice versa*. Similar plots were also generated for all the five experimental data sets corresponding to the various off-boresight angles θ. Inferences drawn about the scatterer profiles from all such similar plots were starkly consistent among the different data sets considered in this study. All of them indeed exhibit a trend identifiable from the representative plot of Figure 11.

In keeping with the principle goal of the study reported here, every prominent feature in the several plots generated (similar to Figure 11) directly impacts the type of interaction that the C-band signal had suffered since leaving the TX. The scatter centers announce their presence either individually (direct LoS return) or in combination (indirect LoS, as well as multiple scatters along a given LoS in situations involving more than one collinear center). The signatures take the form of reduced clustering scales in the corresponding data; hence, a larger box dimension is associated with every such data point indicated in the representative plot. The weak clustering concerns a destructive interference phenomenon to the received RF signals. In effect, a tomography of the experiment space, along specific azimuth and elevation planes, has been conducted from within such a space – against the conventional practice that involves a peripheral investigation procedure. In the practical case, equivalent mid-band 5-6G signals could now identify the ideal locations for their BSs, along paths corresponding to the Type-2 class. A total of 22 such Type-2 cases, each having a magnitude that is below the -12 dB significance threshold for the normalized box size, were found in this study. The relevance of the corresponding TX-RX locations lies in a unique combination of material types, spatial distributions of scattering centers, signal parameters and incident angles. Our approach is a lightweight unsupervised model that utilizes multipath undergoing multiple reflections. The approach can be potentially extended to other effects like diffraction. Our approach did not require prior data association labels between virtual anchors and extracted time of flight (ToF)s, since it is more challenging to extract virtual anchor locations when only one of the modalities among ToF, range measurements, or angle of arrival (AoA) of multipath components is available. SLAM-based positioning in wireless communications differs from RF-fingerprinting in explicitly modeling the multipath environment, while RF-fingerprinting creates a correlation model between channel measurements and user locations. In all such cases, multiple modes naturally arise in the distributions of locations after multiple measurements [62], [63], [64]. The underlying network can extract spatial information given any input, such as channel state information (CSI), time difference of arrival (TDoA), or ToF, if available. In contrast to such fingerprinting methods, our approach is unsupervised. Although the proposed method has considered a select set of anchor scenarios, it can certainly be extended to a multi-anchor case by concatenating several such instances, one per anchor. The location co-ordinates defined in this experiment are GPS-disciplined. In practical applications of the proposed technology, a wide variety of on-board location sensing solutions in the BS-MS shall enable the effective co-ordinates referencing and transformation among such hardware in arriving at the most optimum geometry and link.

Complex modeling or RT calculations, using such parameters, have been the conventional methods to arrive at distortion-free and efficient communication pathways in the MIMO channel that require further implementations of corrective, control and reconfiguration steps, contributing to sizable latency and inefficiency. The PALE-on-EDGE scheme proposed here has sidestepped such tasks with highly encouraging outcomes. The link attributes among BSs, and dynamic MD-based EDGE units deployed on-field could now be defined primarily as a data-driven, on-the-fly signal acquisition and processing (and, by extension, a control/reconfiguration) activity. The significant advantage here would be the inference-delay-free classification of the channel attributes even while the communication services are fully operational, performed in effect as an algorithmic pre-processing task.

Inferences derived from the distributions in the bar charts and the box plots have been instrumental in defining the properties of scatterers in more precise terms and with enhanced definiteness about their positions and characteristics. Such a quantitative analysis has primarily derived salient benefits from the outcomes of the many intensely data-driven approaches pursued during this study. The K-means clustering scheme was invoked to characterize the scatterers as part of an elaborate exercise to identify and deploy a suitable clustering strategy for the spatial-temporal data set. The latter have been progressively facilitating (and refining) an eventual lead onto the conclusive stages in scatterer profiling.

To effectively project the advantages in the current work, *vis-a-vis* the previous related attempts at discerning the characteristics of the scattering centers in the propagation channel, a detailed comparison has been made. The few highly topical evaluation parameters that had been considered in this exercise are:

1) *Channel type:* specifies the propagation medium, context, and application domain
2) *RF band:* operating frequency range
3) *Distortion type:* phenomenon affecting signal propagation
4) *Factors of influence:* primary contributors to distortion
5) *Methodology:* procedure employed
6) *Variability factors:* significant variables that yield salient results
7) *Evaluation metrics:* performance verification tools
8) *Detection accuracy:* exactitude of analyzed outcomes and validation
9) *Details of target:* contributors to distortion in the propagation channel
10) *Complexity in estimation:* dimensional, procedural, and computational rigor involved.

Several poignant attributes from such a comparison study have been made evident, and these are detailed in Tables 3-5. In a significant number of instances, a marked advantage has

been noticeable with the employ of the distributed PALE-on-EDGE implementations.

TABLE 3. Performance comparison.

Evaluation parameter	Ref. [4]	Ref. [5]	Ref. [8]
Channel type	Marine, sea surface	Marine, sea surface	Terrestrial: indoor and outdoor; 4G MIMO
RF band	C-band	L-band	S & C bands
Distortion type	Scattering	Diffuse and specular, scattering; interference; multipath fading	3D scattering
Factors of influence	Bandwidth, Number of sub-carriers, sub-carrier spacing, Tx power, Signal period, antenna design, platform dynamics	Elevation angle, sea surface wave height, Tx antenna gain, coherent and incoherent surface reflection coefficient, polarization	Probability density function of Rx signal's angle of arrival, time of arrival
Methodology	Channel sounding – model-base; measurements	Model-based; experiments	Comparison among analytical scattering channel models; simulations
Variability factors	Rx power; Tx antenna location in 3D; range	Sea wave height and wavelength; Rx antenna elevation angle	Relative locations and dynamics of base stations and mobile stations; MIMO antenna array topology
Evaluation metrics	Normalized Rx power; propagation delay	Scatter intensity; fading depth; reflection coefficient	Channel capacity; Spatial fading correlation
Detection Accuracy	Above average	Good	Average
Details of target	Known	Known	Generic
Complexity in estimation	High	Sizable large	High

TABLE 4. Performance comparison.

Evaluation parameter	Ref. [9]	Ref. [10]	Ref. [11]
Channel type	Terrestrial - hills and mountains	Terrestrial	Terrestrial urban microcell
RF band	L-band	UHF band	L & X bands
Distortion type	Diffuse scattering; multipath fading	Scattering; fading	2D scattering
Factors of influence	Terrain features	Non-LoS Tx-Rx	Probability density function of Rx signal's angle of arrival, angle of departure, time of arrival at base station
Methodology	Comparison of models to simulated data	Model-based; experiments	Comparison among analytical scattering channel models; and with empirical datasets
Variability factors	Rx signal time delay & angle spread; antenna dimensions	Angle of arrival; antenna dimensions	Angular distribution of scatterers, base station, and mobile station; antenna dimensions
Evaluation metrics	Path loss; link analysis	Antenna beamwidth; Correlation coefficient;	Angular and temporal curve fitting, histogram, and error vector magnitude
Detection Accuracy	Average	Average	Average
Details of target	Generic	Known	Generic
Complexity in estimation	High	High	Sizable large

The machine learning emissions calculator [65] has been used to estimate the equivalent black carbon footprint [66] of prominent cloud-based computing resources and on reduced scales by the many other bespoke ML initiatives – including the one considered in this work. The online calculator is a handy tool to compare and highlight the derived impact that massive data processing can have on the environment. In the case of PALE-on-EDGE implementation, the cumulative outcome had been a meager 0.07 kg black carbon emission at an efficiency of 45% of carbon released per unit of electric power utilized. On an identical scale of computation duration, a factor of 10–30 degrade to the environment has been imposed by cloud-based ML platforms. The trend in this

TABLE 5. Performance comparison.

Evaluation parameter	Ref. [12]	Ref. [68]	This work
Channel type	Terrestrial urban and rural roadside	Buildings; urban outdoors	Terrestrial urban
RF band	C-band VANET	L-band	C-band 5G
Distortion type	Diffuse scattering	Diffuse scattering	3D scattering; multpath fading and enhencement
Factors of influence	Doppler frequency shift; incoherent surface scatter cross section	Surface reflection coefficient; angular distribution and direction of RF power; field parameters;	Agnostic to material, design and deployment aspects in mid-band 5G propagation channel
Methodology	Comparison among models, and to prior art measurements	Antenna radiation pattern measurement; comparison among ray tracing models; simulation	Antenna radiation pattern measurement; data driven pre-processing on mobile EDGE computing platforms
Variability factors	Absolute and relative locations and dynamics of Tx and Rx antennas; angle of scatter from Tx side	Ray tracing models; angular distribution and direction of RF power; target and polarization diversity	Relative angular locations of Rx antenna in 3D space
Evaluation metrics	PSD; angular power density	RMS deviation in Rx power	Residual power; histogram; box plot; data processing efficiency
Detection Accuracy	Average	Average	Very high
Details of target	Generic	Known	Known – generic
Complexity in estimation	High	High	Facile; on-the-fly

gap is projected to widen, as the duration of activity increases, with diminishing returns on the quality of the classification outcome associated with the latter in the specific context involving the C-band dataset. The degree of environmental friendliness in the research practices adopted is underlined in this work, distinctly enabling its functionality without resorting to any of the attended carbon offset remedies.

In the case of the PALE-on-EDGE implementation, the actual "edge" is the completeness in the features it offers, in the form of a latency-free signal acquisition and inference-delay tolerant data analysis, that the cloud-based systems distinctly lack. With such a test arrangement, robust estimates on antenna placement and features in the topology can now be derived with enhanced thoroughness and relatively less effort, with no requirements for any further meaningful classification expected from the ML algorithms.

The proposed method would drastically enhance the efficiency of system-wide deployments of 5-6G technologies across an otherwise complex urban landscape. Complexity notwithstanding, the results obtained here are agnostic to the multitudes of scattering centers of varying scattering coefficients as against the precise knowledge of the types and locations in their limited numbers considered in the diffuse scattering models and RT approaches [15], [16]. The latter are conspicuous in their deployment of sophisticated test and measurement devices and simulation software that would prove disadvantageous and unbound on time and cost scales in the context of dynamic 5-6G field conditions.

There would be multiple such conducive arrangements in the actual field scenarios, and data from several among such receivers could be harnessed. All such data can then be subjected to an *in situ* analysis and interpretation, using the PALE model identified with this study and deployed on multiple MD-based EDGE platforms. The real-world, distributed large-scale profiling exercise would significantly weigh in favor of a collective effort made by the otherwise computationally light-weight EDGE platforms. A stark contrast now emerges to the HPC-defined ML platforms that have immense computing requirements (upwards of a 100-fold increase in the hardware cost [67]) and relative, unavoidable delays in retrieving the processed results; these had not been considered in this work.

The design and development of 5-6G systems is a massive undertaking. C-band 5G (and the future 6G) networks would require a different level of attention in terms of planning, deployment, and maintenance. To build and iterate till the system becomes operational is time-consuming and expensive. Time-to-market and network quality will depend on the of tests and measurements during the complete life cycle of the network. Complete virtual prototyping and simulation can substantially reduce costs and accelerate this design and deployment process.

The 5G technologies can create a revolution in operational flexibility, and the sixth-generation (6G) will be known for using AI to capitalize on this flexibility [69], [70], [71]. In 6G, intelligent services will span from cloud data centers to end devices and IoT devices, *e.g.*, self-driving cars, drones, and auto-robots. It is of prime importance to design ultra-low latency, ultra-low power and low-cost inference

processes. To overcome stringent computation, bandwidth, storage, power and privacy constraints on individual devices, increasing research interests are moving towards leveraging the dispersed computing resources across the cloud, network edge and end devices of 6G networks through the lens of MD-based EDGE computing [72]. For example, for a deep neural network (DNN), the initial features can be extracted on the end devices, which are then sent to the EDGE and cloud computing devices for further processing. However, with the heterogeneity in the computing capabilities and communication bandwidths among the computing devices, it becomes incredibly challenging to allocate the operations of the neural networks to the computing devices so that the latency and energy are optimized.

Earlier attempts, such as those by Nicoletti and Busiello [68], aimed to differentiate the various coupled systems in a propagation channel based on their degree of mutual information in complex and dynamic situations. A reasonably accurate prediction of channel characteristics based on diffuse scattering effects in restricted urban built spatial environments had been advanced by Degli-Esposti [68]. All these methods would suffer from a fundamental handicap when they are being scaled up to real-world scenarios, where complex 5G use cases that support service ubiquity, scalability, and continuity are the primary operational attributes. Such limitations, which either take the form of defining a restrictive theoretical framework or that of an algorithmic response devised for channel modeling, are entirely rooted out even at the pre-processing stage by the proposed data-driven approach. Highly dynamic, diverse and evolving channel characteristics will have to be negotiated with agile computing resources. The PALE algorithms suite, selected and assembled as a pipeline processing scheme, conforms to the requirements set to extract and interpret evident trends and patterns that, in turn, constitute keenly discernible features embedded in the scattered signals in real-time. Besides this, a multi-angle monitor methodology called microwave tomography identifies distributed topological elements in the practical world. In such an evolved scenario, there would be multiple instances of an on-device deployment of the PALE algorithms suite in the form of embedded software residing in MD-based EDGE computing platforms, such as the wide variety of portable digital communication devices. Such a distributed sensing and analysis scheme would effectively address the on-the-fly requirements of various operators in the sub-6 GHz spectrum. The vastly reduced inference-delay during modeling and classification of the channel parameters shall yield enhanced QoS outcomes.

VI. CONCLUSION AND FUTURE SCOPE

The objective of this research work has been to profile scatterers in the propagation channel, whose signature features were latent in the experimental data collected in a routine laboratory environment. The systematic application of various PALE-type data-driven approaches that conform to a broader gamut of pre-processing schemes has convincingly demonstrated its capability to profile the scatter centers. The task had been accomplished without resorting to an intensive ML algorithm-based classification regime that is computationally demanding. The proposed method, with its two parts, *viz.*, acquisition and analysis, is also in stark contrast to the conventional analytical model-driven schemes that have been extensively used to characterize various scatterers in the C-band signal propagating channel. The distinctive advantage here is the demonstrated ability of the PALE algorithms suite to closely track the evident patterns that the scatterers create during their interaction with the RF signals. A concurrent means has been identified to qualitatively characterize and quantitatively determine the influence of the scattering centers on the propagating signal. The derived interpretations have been the location, material attributes, morphology, and kinematics of the genuine scatterers. Such parameters have been identified even at the first-order processing of the tracked time series data.

The results obtained in this work shall find significant relevance in situations that involve the intentional incorporation of technologies that shall help evade the detection of targets of interest in aerospace situational awareness operations by being employed as potent electronic counter-countermeasures. The proposed method has the significant advantage of sensing and tracking such targets from the signatures of change in the reference signal in the wake of their interaction with the hostile constituents in the propagation channel. The results obtained are primarily based on the outcome of pre-processing a spatial-temporal data set.

AI/ML is evolving rapidly. The MD industry must take advantage of this technology and apply it to the mobile network architecture from the core to the RAN, the radio interface, and the end-user device itself. The federated learning models proposed in this work leverage 5-6G connectivity, on-device learning and inference techniques to take ML closer to the real-time processing needed for air interface optimization and superior end-user experience. Device-based learning enhances 5-6G QoS. The prospect of an adaptive, ML-native air interface, for example, could generate a radically simpler radio that generates unprecedented gains in efficiency and performance.

As a possible extension of the work reported here, a test system can be devised to perform in a manner that would constrain its acquisition parameters dynamically. Such a system would function with the precise understanding of scatter signatures obtained from the trends and patterns in the acquired data (primarily antenna radiation patterns) fed back in real-time. With such a closed-loop scheme, the malicious components in a sensitive RF signal detection procedure can now be subjected to adaptive excision, a task currently being pursued at an exorbitant cost and time in use of facilities such as the anechoic chamber. The profiling of antennas with unknown radiation characteristics will become a wholly computationally defined task that will prove cost-effective.

The process gets refined during the testing phases, in terms of noise detection, with its progressive exposure to additional data sets as part of a mainstream supervised ML classification exercise.

Highly dynamic, *ad-hoc* BS densification operations in cellular telephony are being initiated extensively to support complex radio techniques in 5-6G that need to offer enhanced data rates, capacity and coverage in the macro and micro spatial scales. Future data communication networks of 6G shall offer broadband cellular services over areas divided into tiny geographical areas called cells, a combination of features of existing 4G and evolving 5G network standards. The 6G networks are envisaged to offer superior data throughput services (50 times more than 5G). These shall incorporate pervasive ML & AI algorithms across their functional and operational domains as an integral element in designing optimum topologies and communication protocols. In such restricted scale operations, the PALE algorithms suite discussed in this article shall immediately accept deployable assets onboard the various types of embedded, connected EDGE computing platforms, such as MDs. Rigorous incorporation of ML techniques, beyond the first and promising stages in the analysis could be pursued. These schemes shall follow the pre-processing strategies discussed in this work. In concert with high-quality situational awareness models, the PALE suite shall yield rapid and time-bound estimates on theater-specific channel scatter attributes at the locations of antenna terminals in practical mid-band 5G-NR roll-out operations. These shall, in effect, serve as robust replacement solutions to the various conventional methods that require extensive modeling and outdoor test range studies at elevated costs and over lengthier time scales.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their vital observations, that have resulted in a vastly improved presentation of the scientific and technical content in this article. They also wholly express their gratitude to ''Grammarly'' (of Ukrainian origin) for thoroughly sprucing up the manuscript in terms of the spelling, grammar, punctuation, clarity, engagement, and delivery issues in the earlier drafts; the readability of this article has tremendously improved as a result.

REFERENCES

[1] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, ''Smart factory of industry 4.0: Key technologies, application case, and challenges,'' *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[2] K. Schwab. *The Fourth Industrial Revolution: What it Means, How to Respond*. World Economic Forum. Accessed: Jun. 17 2022. [Online]. Available: www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond

[3] D. Imkamp, J. Berthold, M. Heizmann, K. Kniel, E. Manske, M. Peterek, R. Schmitt, J. Seidler, and K. D. Sommer, ''Challenges and trends in manufacturing measurement technology—The 'Industrie 4.0 concept,'' *J. Sensors Sensor Syst.*, vol. 5, pp. 325–335, Oct. 2016.

[4] W. Wang, G. Hoerack, T. Jost, R. Raulefs, M. Walter, and U. C. Fiebig, ''Propagation channel at 5.2 GHz in Baltic Sea with focus on scattering phenomena,'' in *Proc. 9th EuCAP*, Apr. 2015, pp. 1–5.

[5] Y. Karasawa, M. Yamada, and J. E. Allnutt, ''A new prediction method for tropospheric scintillation on Earth-space paths,'' *IEEE Trans. Antennas Propag.*, vol. AP-36, no. 11, pp. 1608–1614, Nov. 1988.

[6] S. Srivatsa and G. A. S. Sundaram, ''PAM4-based RADAR countermeasures in hostile environments,'' in *Proc. 3rd ISTA Adv. Intell. Syst. Comput.*, vol. 683, no. 1. Singapore: Springer, 2018, pp. 390–400.

[7] P. A. van Walree, ''Propagation and scattering effects in underwater acoustic communication channels,'' *IEEE J. Ocean. Eng.*, vol. 38, no. 4, pp. 614–631, Oct. 2013.

[8] J. Zhou, H. Jiang, and H. Kikuchi, ''Generalised three-dimensional scattering channel model and its effects on compact multiple-input and multiple-output antenna receiving systems,'' *IET Commun.*, vol. 9, no. 18, pp. 2177–2187, 2015.

[9] J. S. Lu, X. Han, and H. L. Bertoni, ''The influence of terrain scattering on radio links in hilly/mountainous regions,'' *IEEE Trans. Antennas Propag.*, vol. 61, no. 3, pp. 1385–1395, Mar. 2013.

[10] W. C.-Y. Lee, ''Effects on correlation between two mobile radio base-station antennas,'' *IEEE Trans. Veh. Technol.*, vol. VT-22, no. 4, pp. 1214–1224, Nov. 1973.

[11] A. Maric, E. Kaljic, and M. Hadzialic, ''A unified hollow-disk scattering model: Angle-of-departure and time-of-arrival statistics,'' *IEEE Trans. Vehic. Technol.*, vol. 68, no. 5, pp. 4342–4352, Mar. 2019.

[12] L. Cheng, D. D. Stancil, and F. Bai, ''A roadside scattering model for the vehicle-to-vehicle communication channel,'' *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 449–459, Sep. 2013.

[13] S. Sundaresan and G. A. S. Sundaram, ''Simulation study on modeling the effects of wind turbine on communication signals (C and X bands) using XGtd,'' in *Proc. ICCSP*, Apr. 2015, pp. 77–173.

[14] Z. Chen and V. Rodriguez, ''Proposed changes and updates on IEEE std 1128—Recommended practice on absorber evaluation,'' in *Proc. Antenna Meas. Techn. Assoc. Symp. (AMTA)*, Oct. 2019, pp. 1–5, doi: 10.23919/AMTAP.2019.8906362.

[15] A. Navarro, D. Guevara, and D. A. Parada, ''Diffuse scattering implementation and verification in ray launching based tool at 5.4 GHz,'' *Proc. 32nd URSI GASS*, Aug. 2017, pp. 1–4.

[16] M.-H. Ren, X. Liao, J. Zhou, Y. Wang, Y. Shao, S. Liao, and J. Zhang, ''Diffuse scattering directive model parameterization method for construction materials at mmWave frequencies,'' *Internat. J. Antenna. Propag.*, vol. 2020, Dec. 2020, Art. no. 1583854.

[17] B. Derat, ''5G antenna characterization in the far field: How close can far-field be?'' *Proc. EMC/APEMC*, May 2018, pp. 959–962.

[18] C. Gentner, T. Jost, W. Wang, S. Zhang, A. Dammann, and U.-C. Fiebig, ''Multipath assisted positioning with simultaneous localization and mapping,'' *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6104–6117, Sep. 2016.

[19] E. Leitinger, F. Meyer, F. Hlawatsch, K. Witrisal, F. Tufvesson, and M. Z. Win, ''A belief propagation algorithm for multipath-based SLAM,'' *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5613–5629, Sep. 2019.

[20] H. Naseri and V. Koivunen, ''Cooperative simultaneous localization and mapping by exploiting multipath propagation,'' *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 200–211, Jan. 2017.

[21] E. Leitinger, P. Meissner, C. Rüdisser, G. Dumphar, and K. Witrisal, ''Evaluation of position-related information in multipath components for indoor positioning,'' *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2313–2328, Nov. 2015.

[22] VIAVI Solutions Inc. *Application Note: The Role of C-band in 5G*. Accessed: May 12 2022. [Online]. Available: https://www.viavisolutions.com/en-us/literature/role-c-band-5g-application-notes-en.pdf

[23] (Sep. 2017). *3GPP TR 38.901: Study on Channel Model for Frequencies From 0.5 to 100 GHz*. Accessed: Apr. 10, 2021. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.901/38901-h00.zip

[24] *3rd Generation Partnership Project Technical Specification Group Radio Access Network NR*, Standard 3GPP TS 38.101-3 V15.10.0 (2020-06), 3GPP Organizational Partners, 2020, p. 183.

[25] M. Kottkamp, *5G New Radio Fundamentals, Procedures, Testing Aspects, White Paper*. Berlin, Germany: Rohde & Schwarz GmbH & Co. KG, 2019.

[26] M. Shafi, ''5G: A tutorial overview of standards, trials, challenges, deployment, and practice,'' *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[27] E. Dahlman, ''5G wireless access: Requirements and realization,'' *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 42–47, Dec. 2014.

[28] M. Yao, M. Sohul, V. Marojevic, and J. H. Reed, "Artificial intelligence-defined 5G radio access networks," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 14–20, Mar. 2019.
[29] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart., 2019.
[30] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jul. 2020.
[31] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "Deep Learning's diminishing returns: The cost of improvement is becoming unsustainable," *IEEE Spectr.*, vol. 58, no. 10, pp. 50–55, Oct. 2021.
[32] V. Francois-Lavet, G. Rabusseau, J. Pineau, D. Ernst, and R. Fonteneau, "On overfitting and asymptotic bias in batch reinforcement learning with partial observability," *J. AI Res.*, vol. 65, pp. 1–30, May 2019.
[33] C. A. Balanis, *Advanced Engineering Electromagnetics*, 2nd ed. New York, NY, USA: Wiley, 2012, p. 1040.
[34] F. Saccardi, F. Rossi, F. Mioc, L. J. Foged, and P. O. Iversen, "Application of the translated-SWE algorithm for the characterization of antennas installed on cars using a minimum number of samples," in *Proc. Antenna Meas. Techn. Assoc. Symp. (AMTA)*, Oct. 2017, pp. 1–6, doi: 10.23919/AMTAP.2017.8123711.
[35] Ettus Research. *VERT2450 Antenna*. Products. Accessed: Apr. 5, 2022. [Online]. Available: https://www.ettus.com/allproducts/vert2450
[36] National Instruments Corp. *USRP-2954*. USRP Software Defined Radio Device. Accessed: Apr. 5 2022. [Online]. Available: https://www.ni.com/en-in/support/model.usrp-2954.html
[37] National Instruments Corp. *What is LabVIEW? Graphical Programming for Test & Measurement NI*. Product Features. Accessed: Apr. 5, 2022. [Online]. Available: https://www.ni.com/en-in/shop/labview.html
[38] National Instruments Corp. *PXIe-8880 Features*. PXI Controller. Accessed: May 10, 2022. [Online]. Available: https://www.ni.com/docs/en-U.S./bundle/pxie-8880-features/resource/374684c.pdf
[39] National Instruments Corp. *LabVIEW Communications System Design Suite NI*. Overview. Accessed: Apr. 5, 2022. [Online]. Available: https://www.ni.com/en-us/innovations/white-papers/14/overview-of-the-labview-communications-application-frameworks.html
[40] V. Monebhurrun, "Revision of IEEE antennas and propagation society standards 149, 211, and 1502," *IEEE Antennas Propag. Mag.*, vol. 58, no. 3, pp. 104–113, Jun. 2016.
[41] Arduino. *Shield_Motor-REV3c*. Reference Designs. Accessed: May 10, 2022. [Online]. Available: https://www.arduino.cc/en/uploads/Main/arduino_MotorShield_Rev3-schematic.pdf
[42] Raspberry Pi. *Raspberry Pi 2 Model B*. Specifications. Accessed: Apr. 5, 2022. [Online]. Available: https://www.raspberrypi.org/products/raspberry-pi-2-model-b
[43] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, I. Z. Kovacs, I. Rodriguez, O. Koymen, A. Partyka, and J. Jarvelainen, "Propagation path loss models for 5G urban micro- and macro-cellular scenarios," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6, doi: 10.1109/VTCSpring.2016.7504435.
[44] S. A. Hanna, "Convolutional interleaving for digital radio communications," in *Proc. 2nd Internat. Conf. Universal Pers. Commun.*, vol. 1, Oct. 1993, pp. 443–447.
[45] R. Harjani and T. A. Lee, "FRC: A method for extending the resolution of Nyquist rate converters using oversampling," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 4, pp. 482–494, Apr. 1998.
[46] S. Rapuano and F. J. Harris, "An introduction to FFT and time domain windows," *IEEE Instrum. Meas. Mag.*, vol. 10, no. 6, pp. 32–44, Dec. 2007.
[47] R. Jolly and G. A. S. Sundaram, "Phased antenna array design for L-band aerospace detection of radiolytic components," in *Proc. 11th Internat. Conf. Signal Process. Commun.*, Jun. 2016, pp. 5–81.
[48] S. A. Schelkunoff, and H. T. Friis, *Antennas: Theory and Practice*. New York, NY, USA: Wiley, 1952.
[49] C. A. Balanis, *Antenna Theory Analysis and Design*, 3rd ed. New York, NY, USA: Wiley, 2005, ch. 2.
[50] T. Velmurugan and T. Santhanam, "Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points," *J. Comput. Sci.*, vol. 6, no. 3, pp. 363–368, 2014.
[51] K.-L. Du, "Clustering: A neural network approach," *Neural Netw.*, vol. 23, no. 1, pp. 89–107, 2010.
[52] N. H. Spencer, *Essentials of Multivariate Data Analysis*. Boca Baton, FL, USA: Chapman & Hall, 2014, p. 186.
[53] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 200–210, 2013.
[54] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987.
[55] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: Wiley, 2009, p. 342.
[56] H. W. Choi, N. M. F. Qureshi, and D. R. Shin, "Comparative analysis of electricity consumption at home through a silhouette-score prospective," in *Proc. 21st Internat. Conf. Adv. Commun. Technol.*, Feb. 2019, pp. 589–591.
[57] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, and Y. He, "Classifier fusion with contextual reliability evaluation," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1605–1618, May 2017.
[58] C. L. Clayman, S. M. Srinivasan, and R. S. Sangwan, "K-means clustering and principal components analysis of microarray data of L1000 landmark genes," *Proc. Comput. Sci.*, vol. 168, pp. 97–104, Jan. 2020.
[59] N. Arbin, N. S. Suhaimi, N. Z. Mokhtar, and Z. Othman, "Comparative analysis between K-means and K-medoids for statistical clustering," in *Proc. 3rd Internat. Conf. Artif. Intell. Modeling Simulation*, Dec. 2015, pp. 117–121.
[60] B. Ratner, "The correlation coefficient: Its values range between +1/-1, or do they?" *J. Targeting, Meas. Anal. Market.*, vol. 17, no. 2, pp. 139–142, 2009.
[61] Statistics Canada, Ottawa, ON, Canada. *Measures of Dispersion: Visualizing the Box and Whisker Plot*. Statistics: Power From Data. Accessed: May 10, 2022. [Online]. Available: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214889-eng.htm
[62] F. Herranz, A. Llamazares, E. Molinos, and M. Ocaña, "A comparison of slam algorithms with range only sensors," in *Proc. ICRA*, May 2014, pp. 4606–4611.
[63] B. Boots and G. Gordon, "A spectral learning approach to range-only SLAM," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 19–26.
[64] J. Djugash and S. Singh, "A robust method of localization and mapping using only range," in *Experimental Robotics*. Singapore: Springer, 2009, pp. 341–351.
[65] A. Lacoste, *Quantifying the Carbon Emissions of Machine Learning*. Montreal, QC, Canada: Montreal AI Ethics Institute, 2019.
[66] H. Timonen, P. Karjalainen, P. Aalto, S. Saarikoski, F. Mylläri, N. Karvosenoja, P. Jalava, E. Asmi, P. Aakko-Saksa, N. Saukkonen, and T. Laine, "Adaptation of black carbon footprint concept would accelerate mitigation of global warming," *Environ. Sci. Technol.*, vol. 53, no. 21, pp. 12153–12155, 2019.
[67] *AI: OS for Machine Learning*. Accessed: Mar. 20, 2022. [Online]. Available: https://cnvrg.io
[68] V. Degli-Esposti, F. Fuschini, E. M. Vitucci, and G. Falciasecca, "Measurement and modeling of scattering from buildings," *IEEE Trans. Antennas Propag.*, vol. 55, no. 1, pp. 143–153, Jan. 2007.
[69] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
[70] M. Yao, M. Sohul, V. Marojevic, and J. H. Reed, "Artificial intelligence-defined 5G radio access networks," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 14–20, Mar. 2019.
[71] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G Wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul./Aug. 2019.
[72] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.
[73] G. Nicoletti and D. M. Busiello, "Mutual information disentangles interactions from changing environments," *Phys. Rev. Lett.*, vol. 127, Nov. 2021, Art. no. 228301.

G. A. SHANMUGHA SUNDARAM (Member, IEEE) was born in Coimbatore, Tamil Nadu, India. He received the Ph.D. degree in radio astronomy and solar radio astrophysics from the Indian Institute of Science, Bangalore, Karnataka, India, in 2004.

He has been a Research Faculty Member with the Department of Electronics and Communication Engineering, Amrita University, Coimbatore, since 2009. His earlier stints had been as a Research Scientist, in the SKA Project on next-generation instrumentation for radio astronomy, with JIVE-ASTRON, The Netherlands, from 2007 to 2009; and a Postdoctoral Scientist in JIVE-ASTRON and the Indian Institute of Astrophysics from 2006 to 2007 and from 2004 to 2006, respectively. From 2016 to 2020, he was PI for the project ''Design and Evaluation of a DRFM Mitigation System'' in the academic research grant funded by NI, USA. He has a U.S. patent on ''OAM Modulation in a Focal Plane Antenna Array.'' He has published around 70 articles till date. Over the past two decades, his research interest includes RF front end system design.

Dr. Shanmugha Sundaram is a member of professional organizations, such as the SPIE, COSPAR, URSI, NRAO, ASP, and AIAA. He was a recipient of the 2008 URSI Award for Young Scientists in the General Assembly held in Canada.

R. GANDHIRAJ was born in Coimbatore, Tamil Nadu, India. He received the Ph.D. degree in software-defined radio system design and applications from Amrita University, Coimbatore, in 2017.

He has been a Faculty Member with the Department of Electronics and Communication Engineering, Amrita University, since 2008. From 2015 to 2020, he had been a Co-PI of two funded projects, one of which was in ''Design and Evaluation of a DRFM Mitigation System'' the academic research grant funded by NI, USA. He has published around 60 articles till date. His research interests include digital communication, signal processing, and wireless communication.

B. N. BINOY was born in Kerala, India. He received the Ph.D. degree in applied soft computing for financial engineering from Amrita University, Coimbatore, India, in 2015.

He has been a Faculty Member with the Department of Electronics and Communication Engineering, Amrita University, since 2007. Since 2015, he had been a PI of a funded project, and a Co-PI of two other funded projects, one of which was in ''Design and Evaluation of a DRFM Mitigation System'' the academic research grant funded by NI, USA. He has published around 60 articles till date. His research interests include applying ML/AI to condition monitoring, advanced driver assistance systems, and financial markets.

S. I. HARUN was born in Coimbatore, Tamil Nadu, India. He received the B.Tech. degree from the Department of Electronics and Communication Engineering, ASE-Coimbatore, Amrita University, in 2020, and the master's degree from Nanyang Technological University, Singapore. He did the final year project activity in the related area to the one reported in this article.

S. N. SURYA was born in Coimbatore, Tamil Nadu, India. He received the B.Tech. degree from the Department of Electronics and Communication Engineering, ASE-Coimbatore, Amrita University, in 2020, and the master's degree from TU Delft, The Netherlands. He did the final year project activity in the related area to the one reported in this article.