RESEARCH ARTICLE

Physics-Informed Machine Learning Modelling of RF-EMF Exposure in Massive MIMO Systems

SAMUEL BILSON [1], **TIAN HONG LOH** [2], (Senior Member, IEEE),
FABIEN HÉLIOT [3], (Member, IEEE), AND **ANDREW THOMPSON** [1]
[1]Department of Data Science, National Physical Laboratory, TW11 0LW Teddington, U.K.
[2]Department of Electromagnetic and Electrochemical Technologies, National Physical Laboratory, TW11 0LW Teddington, U.K.
[3]Institute for Communication Systems (ICS), 5GIC and 6GIC, University of Surrey, GU2 7XH Guildford, U.K.

Corresponding author: Samuel Bilson (sam.bilson@npl.co.uk)

ABSTRACT Beamforming and massive multiple-input-multiple-output (mMIMO) technologies are key features of base stations (BSs) in the fifth-generation (5G) of mobile networks. This technology is used to focus more radio frequency (RF) energy towards actively connected users to improve their connection/performance, resulting in high variations in the radio frequency electromagnetic fields (RF-EMFs). This paper proposes a new methodology for modelling the RF-EMF exposure for 5G new radio (NR) mMIMO BS by means of a physics-informed machine learning (ML) approach using empirical measurement data. More precisely, the main focus of our work is to develop a suitable traceable RF-EMF exposure prediction tool in the context of 5G mMIMO BSs that can serve multiple mobile users (i.e. multiple-user MIMO (MU-MIMO)) within realistic real-world environments and scenarios. Our RF-EMF prediction tool relies on empirical measurement data acquired via a user-controllable mMIMO beamforming testbed and traceable RF-EMF measurement capability, where both indoor and outdoor RF-EMF measurement campaigns have been carried out. During the measurement campaigns various factors such as number of users, position of users and data duty cycles were considered. Using an ensemble of gradient boosted decision trees, we show that a physics-informed approach can improve predictive performance of RF-EMF compared with a purely data-driven approach, with the ability to extrapolate values of RF-EMF exposure to larger distances. Results show a coefficient of determination value of 0.86 on a 10-fold cross-validated experimental dataset. We also compare the sensitivity of RF-EMF exposure to various factors in the model, and show that model predictions become isotropic for large numbers of beam configurations, simplifying the exposure measurement methodology of 5G systems.

I. INTRODUCTION
Fifth-generation (5G) networks promise to deliver high data rate, low latency, and seamless connectivity between trillions of devices serving billions of people. However, when it comes to 5G mobile network deployment, the radio frequency electromagnetic field (RF-EMF) exposure limits (in terms

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

of field strength) have become a critical concern. This is mainly because 5G new radio (NR) base stations (BSs) rely on the beamforming massive multiple-input-multiple-output (mMIMO) technology [1], [2], [3], [4]. This technology is capable of adaptively forming narrow directional beams, allowing transmission of more information towards different desired users on the same frequency/time resource, i.e. focusing more radio frequency (RF) energy actively on connected users, resulting in high spatio-temporal variations

of the RF-EMF strength. This is quite different from previous generations of mobile networks and BSs, where RF power was assumed to radiate approximately uniformly in all directions.

Consequently, the conventional measurement of RF-EMF exposure from third-generation (3G) and fourth-generation (4G) BSs at the exclusion zone (a compliance boundary around the BS with no access to the general public), is currently not fit for purpose since it leads to non-realistic large exclusion zone areas when the beamforming technology is used to focus more RF energy in the direction of a specific mobile user [5]. In turn, this non-realistic information can make the deployment of 5G mMIMO BSs problematic for operators, especially on sites with pre-existing 3G and 4G BSs, or in countries, regions, and even cities, where RF exposure compliance regulation goes beyond the guidelines set out by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) [6], [7]. For example, in Poland, Italy and Switzerland, different regulations have been put in place where the current RF-EMF exposure limits are 7 V/m, 6 V/m and 4 V/m, respectively, which is much stricter than the ICNIRP guidelines at 61 V/m [8]. This more stringent exposure limit has had an impact on 5G network rollout and deployment. Furthermore, different contributions have already demonstrated that the old compliance methodology is not suitable for 5G BSs [5], [9]. Regulators, operators and 5G equipment suppliers all require reliable and rigorous assessment of RF-EMF exposure levels to support consistent and effective 5G regulation and network design. Scientific arguments and effective RF-EMF measurements on a mMIMO system are needed to support this vision. Furthermore, various factors such as number of users, position of users and data duty cycles need to be considered as they increase the measurement uncertainty in RF-EMF exposure evaluation.

In 5G systems, the RF-EMF exposure is considered to be more statistical/stochastic than with its evaluation done by using data from the synchronisation signal block (SSB) IEC 62232:2022 [10]. More specifically, the exposure level is extrapolated from the received signal reference power (RSRP), which is itself obtained through the SSB. The literature on 5G RF-EMF exposure generally acknowledges that statistical approaches should form the basis of RF-EMF exposure assessment in systems employing mMIMO and beamforming to ensure that high power user service beams are only transmitted on a need-to basis [5], [9], [11]. Theoretical statistical models of 5G BS exposure have been proposed in [5] and [9], which explore how the statistical dependency of RF-EMF on factors such as beam angle and distance from BS can impact existing EMF regulations. Following this theoretical work, the IEC has laid the foundation for developing a practical model-based method to evaluate and extrapolate the EMF exposure of in-situ 5G BSs based solely on measuring the strength of SSB of 5G BS signalling information [4], [12]. The principles of 5G exposure measurements based

on the measurement of SSB signals have been presented in different contributions [13], [14]. Furthermore, in practice, one envisages that BSs are not always operated with 100 % data payloads (i.e. they do not transmit at full power in each millisecond) and when multiple users are served by a BS, the power is split among different directions even with spatial multiplexing. Different contributions tend to demonstrate that the traditional approach for designing the compliance boundary might be over-conservative [9], [15]. In particular, [9] shows that using a statistical approach for assessing the actual transmission power in more realistic scenarios by taking into account mMIMO BS operations (e.g. deployment scenarios and channel models), the compliance boundary turns out to be actually smaller when compared to the one computed with the conservative traditional method. More recently, [16] worked on refining this practical method by either considering a more comprehensive set of parameters to extrapolate the RF-EMF exposure from the SSB measurements or defining a robust procedure mixing both classic field strength and SSB power measurements.

The authors of [2] and [4] carried out empirical studies of the statistical and stochastic nature of RF-EMF exposure from a fully reconfigurable mMIMO testbed system within a real-world indoor and outdoor environment with different complex operating scenarios. The aim was to focus on traceability establishment and development of suitable RF-EMF measurement methods in the context of 5G mMIMO BSs serving different numbers of mobile users with different data traffic profiles within realistic real-world environments and scenarios. This empirical measurement work was useful in establishing the relevant empirical statistical model. However, the definition and validity of an effective model is still an open problem for such complex operations. More work needs to be done to fully understand how RF-EMF exposure generated by a 5G mMIMO BS fluctuates as a function of the environment, such as spatial variation, user traffic profile and number as well as positions of the users.

Machine learning (ML) is a data-driven approach to modelling which has become extremely popular in wireless communication systems in recent years due to the abundant availability of data. ML has an advantage over traditional modelling approaches in its ability to learn complex systems with relatively low computational cost. This has led to numerous applications of ML in understanding different wireless communication phenomena, including various aspects of the physical layer in 5G wireless technology [17], [18]. For example, ML related work has been explored in the study of channel propagation [19], [20], [21], and beamforming [22]. The application of ML in the specific field of RF-EMF exposure is rapidly gaining interest. For example in [23], [24], [25], and [26], various neural networks are used to predict RF-EMF exposure by considering different input parameters under various conditions. Although neural networks tend to give more accurate predictions, they are less interpretable,

and lack the ability to predict outside the range of training data.

Due to the complex and stochastic nature of 5G mMIMO beamforming, and the abundance (in number of measurements and number of varying operational conditions) of empirical real-world data presented in [2] and [4], we propose a supervised learning approach using a physics-informed gradient-boosted ensemble of decision trees for modelling and predicting 5G RF-EMF strength exposure. This has the advantage over neural network approaches in that one can analyse the predictor importance (a measure of sensitivity) of the various factors affecting RF-EMF exposure (see section III-B). Also, by incorporating physical laws (namely the Friis formula [27]) into the model training (see section III-C), we show how the model can accurately interpolate and extrapolate the RF-EMF predictions to distances from the BS not present in the real-world data. Both indoor and outdoor environments, under various operational conditions are considered, with focus on developing a relevant and generic ML model to predict 5G RF-EMF exposure in many useful deployment scenarios.

In section II we describe the real-world experiments from which the RF-EMF was measured. In section III we present the methods used to train the ML model of RF-EMF exposure. We show that a physics-informed model is able to give more accurate predictions of RF-EMF exposure in spatial regions not measured in the experimental data. In section IV we present the results of the physics-informed model predictions, and analyse the dependency of RF-EMF on various factors. We show that the physics-informed model has high predictive performance on unseen test data, with a cross-validated R^2 value of 0.86, and exhibits isotropic RF-EMF exposure behaviour for a high number of users. Finally, in section V, we provide our conclusions of the work presented, and directions for future research.

II. EXPERIMENTAL DESIGN
This section explains how the data set of 5G RF-EMF exposure has been gathered through various measurement campaigns (reported in [1], [2], [3], [4]) by means of a 5G compliant mMIMO testbed. All the measurements (both indoor and outdoor) were performed around the 5G/6G Innovation Centre (5GIC) building at the University of Surrey [28], with the details of the testbed and experimental settings provided below.

A. mMIMO BEAMFORMING TESTBED
The mMIMO testbed used in [1], [2], [3], and [4] to gather 5G RF-EMF exposure data includes a user-controllable mMIMO beamforming system with up to 128 channels connecting to a transmit antenna array and a RF-EMF measurement system. The mMIMO beamforming system itself consists of:
1) A BEE7 synchronization and trigger generator;
2) A MegaBEE transceiver module (each module contents four input/output RF ports and could support up to 4 channels of IQ);

3) A White Rabbit time distribution system;
4) A transmitting antenna array with 128 (16 × 8) patch antenna elements.

The mMIMO testbed can perform phase-coherent, after over-the-air (OTA) mMIMO phase coherency calibration, and time synchronized MIMO baseband processing with user-programmable, reconfigurable, and real-time signal processing field-programmable gate arrays (FPGAs)-based software defined radio (SDR) capabilities. For downlink communications, up to 128 channels could be used simultaneously at the transmitting end by using all the 32 transceiver modules whereas, at the receiving end, up to 32 channels can be used.

The synchronization of the mMIMO testbed is controlled by the BEE7 synchronization and a trigger generator. The signal generation and analysis are all implemented using the MegaBEE transceivers. The clocking network that achieves sub-nano second time synchronization between channels is derived from the White Rabbit time distribution system, which synchronizes a reference clock to each of the MegaBEE transceiver modules over an optical fibre link using SFP+ (Small Form-factor Pluggable Plus) network adaptors. Note that optical fibers are employed for both data transport and the clocking network. Fig. 1(a) depicts the mMIMO beamforming system with the following MIMO downlink configuration features:

- RF operating frequency centric at 2.63 GHz
- CP-OFDM (Cyclic Prefix Orthogonal Frequency Division Multiplexing) waveform matching a 5G NR configuration
- Data frame generation with pilot signals
- 40 MHz instantaneous data bandwidth per channel
- Subcarrier spacing of 15 kHz in time division duplex (TDD) mode
- Symbol modulation formats: QPSK, 16-QAM, 64-QAM
- Flexible Massive MIMO as Tx (transmitting end) configuration for 16 up to 96 Antennas
- Flexible Receiver system as Rx (receiving end) configuration for up to 20 Antennas
- MATLAB baseband processing, algorithmic evaluation and zero-forcing beamforming
- Receive antennas separated by several wavelength distance
- Capable of MIMO precoding at Tx and channel estimation with receive pilots at Rx
- Transmission rate at 61.44 MSps for a fixed length of 65536 samples (i.e. 216)

The RF-EMF measurement system (also referred to as RF-EMF receiver) consists of up to five sets of 4-dipole-element receiving antenna arrays each connected to a MegaBEE receiver (see Fig. 1(b)). The measured traceable RF-EMF exposure is evaluated by converting the amplitude of the received signal at each of the receiver antennas into exposure information via the following equation
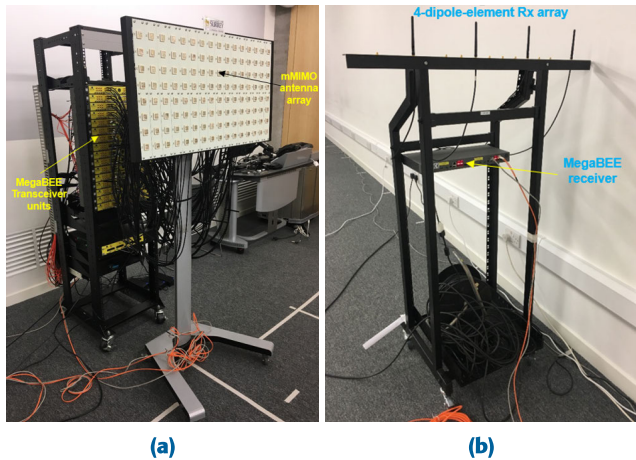
$$\varepsilon = V C_L A_F C_{AG}, \qquad (1)$$

FIGURE 1. Photos of the: (a) mMIMO beamforming system; (b) RF-EMF measurement systems.

(a) Indoor

(b) Outdoor

FIGURE 2. Diagrams showing the configuration of BS and receiving antenna for both indoor and outdoor scenarios. Rx # is a label indicating the receiver number.

TABLE 1. Table of antenna positions on the five receivers for both indoor and outdoor scenarios shown in Fig. 2. The x and y coordinates are taken to be the parallel and perpendicular distances from the centre of the BS, respectively.

(a) Indoor			(b) Outdoor		
Antenna	x (m)	y (m)	Antenna	x (m)	y (m)
Ant1Rx1	2.902	2.00	Ant1Rx1	0.402	3.00
Ant2Rx1	2.671	2.00	Ant2Rx1	0.171	3.00
Ant3Rx1	2.440	2.00	Ant3Rx1	-0.060	3.00
Ant4Rx1	2.209	2.00	Ant4Rx1	-0.291	3.00
Ant1Rx2	-1.098	3.00	Ant1Rx2	-0.948	4.05
Ant2Rx2	-1.329	3.00	Ant2Rx2	-1.179	4.05
Ant3Rx2	-1.560	3.00	Ant3Rx2	-1.410	4.05
Ant4Rx2	-1.791	3.00	Ant4Rx2	-1.641	4.05
Ant1Rx3	1.902	4.00	Ant1Rx3	1.452	4.95
Ant2Rx3	1.671	4.00	Ant2Rx3	1.221	4.95
Ant3Rx3	1.440	4.00	Ant3Rx3	0.990	4.95
Ant4Rx3	1.209	4.00	Ant4Rx3	0.759	4.95
Ant1Rx4	-0.098	5.00	Ant1Rx4	-0.348	6.00
Ant2Rx4	-0.329	5.00	Ant2Rx4	-0.579	6.00
Ant3Rx4	-0.560	5.00	Ant3Rx4	-0.810	6.00
Ant4Rx4	-0.791	5.00	Ant4Rx4	-1.041	6.00
Ant1Rx5	0.902	6.00	Ant1Rx5	0.852	7.05
Ant2Rx5	0.671	6.00	Ant2Rx5	0.621	7.05
Ant3Rx5	0.440	6.00	Ant3Rx5	0.390	7.05
Ant4Rx5	0.209	6.00	Ant4Rx5	0.159	7.05

where V is the received signal voltage measured at each antenna of the receiver, C_L accounts for the cable loss, A_F is the antenna factor to account for the field in the air, such that $A_F = E/V$, with E being the plane wave electric field (E-field). C_{AG} is the corrective factor that accounts for the automatic gain control that is implemented in each MegaBEE transceiver module. The values of C_L, A_F, and C_{AG} have been obtained through calibration and validated by ensuring that the EMF values obtained by each antenna of each receiver well-match EMF theoretical values [4].

B. EXPERIMENTAL SETTINGS

In the various experiments undertaken in [1], [2], [3], and [4], the mMIMO BS operated with 96 active transmitting antennas, while each RF-EMF measurement system operated as a receiver with four dipole receiving antennas. Each antenna was used as an RF-EMF probe to measure the RF exposure at the location they were placed whilst the mMIMO BS was not necessarily transmitting/steering a beam towards them in order to acquire more measurements and, hence, get a better statistical view of the RF-EMF exposure. The RF-EMF exposure was measured by using five RF-EMF receivers fixed in place at different locations within a well-defined grid, as it is depicted in Fig. 2 for both the indoor and outdoor scenarios. Thus, 20 RF-EMF strength measurements were taken for each set of operational settings. The exact locations of all antennas are shown in Table 1.

While the mMIMO Tx system was in operation, its associated RF-EMF was measured when considering the combination of the following varying factors:

- Number of virtual active users: Up to 4 (i.e. generate up to 4 beams)
- Direction of virtual active users (VUs): Up to 19 beam directions
- Data traffic pattern of frame (NRB): Up to 4 different data rates for each user (e.g. 25%, 50%, 75%, 100%)
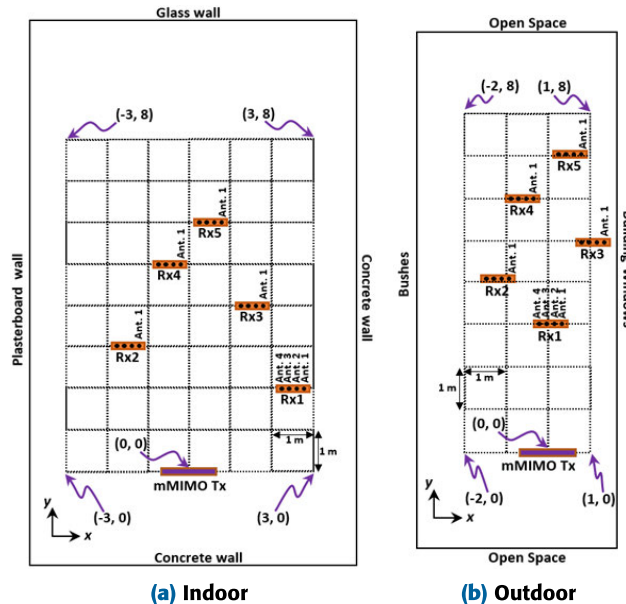
- The modulation coding scheme (MCS): Up to 29 different indices (0-28)
- The number of transmission antennas: Up to 3 different amounts (32, 64 and 96).
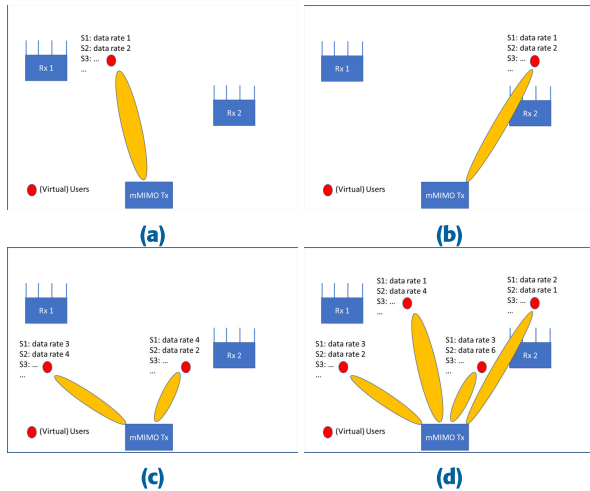
FIGURE 3. Illustrative diagrams showing virtual beamforming scenarios for different number of users, user data rate, and user location: (a) one active user with beam pointing towards surrey Rx1; (b) one active user with beam pointing towards surrey Rx2; (c) two active users; (d) four active users.

During the measurement campaign, various single-user (SU) and multiple-user (MU) MIMO downlink communication beamforming scenarios were considered, with different combinations of active beams and data traffic loading to mimic the performance of a realistic 5G BS. Fig. 3 shows some illustrations on how many beams were formed to transmit towards a virtual active user by considering the aforementioned varying factors. In Fig. 3, virtual users (red dot) do not have a real receiver at that point, the beams are pre-computed and loaded/transmitted in a static way to the field of interest for the duration of the measurement. Also, the static beams are computed taking account of different traffic profiles and numbers of virtual users. Note that the RF-EMF systems have been calibrated to ensure that the measured results could be used for cross validation.

In this measurement campaign, fully compliant 5G NR waveforms were used to transmit the data. The 5G NR beamforming baseband signals were generated by using the Keysight PathWave System Design platform (also known as SystemVue) [29]. This software can be used to generate up to 19 potential beam directions equally spaced every 5° between ± 45° in azimuth angle with an elevation of 0°. Each generated beam contains a physical downlink shared channel with a configurable data payload that is transmitted within a 1-ms period. The data rate of each beam payload can be controlled by adjusting the modulation coding scheme, i.e. QPSK (Quadrature Phase-Shift Keying), 16-QAM (Quadrature Amplitude Modulation), 64-QAM, as well as the number of allocated resource blocks (NRBs) used in the transmission. In the experiment, up to 216 NRBs were used per beam and up to 4 out of the 19 beams are active simultaneously to mimic a mMIMO BS beamforming data transmission towards up to 4 active users at the same time. Note that up to 864 total NRBs were used

for 4 simultaneously active beams and the power level of the mMIMO Tx system was not normalized. The carrier frequency of the downlink transmission was 2.63 GHz with a bandwidth of 38.88 MHz, and a subcarrier spacing of 15 kHz.

One envisages that the variations of the beam profiles and data rates are useful for assessing the spatial variation of RF-exposure surrounding the mMIMO testbed. In order to get insights into the effect of the stochastic nature of RF-exposure generated by non-ideal mMIMO beamforming operation due to potential hardware impairments and/or other factors (i.e. the beams may not be well defined and may not steer in the expected direction) in the studied environment, tens of thousands of electric field (E-field) measurements were acquired, after the multi-channel OTA calibration [3], [4] had been performed, with uniformly independently sampled numbers of active beams, beam directions, NRBs, MCSs and number of transmitters. In the measurement campaigns, a total of 65,980 and 87,880 measurements of E-field data were recorded for indoor and outdoor scenarios, respectively. The relevant RF-EMF is then evaluated using (1).

III. MACHINE LEARNING MODELLING
Since our goal in this paper is to predict the RF-EMF, which is a continuous variable, we need to solve a regression problem where, given a set of inputs, X, we wish to determine a function, f, which well approximates the output RF-EMF, y, as follows

$$y \approx f(X). \qquad (2)$$

A. DATA PREPARATION
Prior to determining f via our proposed physics-informed ML algorithm, we first needed to process the measurement data to identify the key input features/variables which influence the RF-EMF.

As was observed in [4], the peak E-field is measured by the antenna when the beam is directed at the receiver. Thus, the E-field strength depends on the angle of the beam θ_{beam} relative to the position of the receiving antenna θ_{ant}, which we denote as the relative angle

$$\Delta\theta = \theta_{beam} - \theta_{ant}, \qquad (3)$$

where $\theta_{ant} = \arctan(x/y)$ can be calculated using the receiver locations shown in Table 1.

According to the Friis transmission formula [27], the peak E-field also depends on the distance, d, between the BS and each receiver, in the form $E \propto 1/d$ (since the power available at the receiving antenna is proportional to $1/d^2$ and the E-field strength E (RF-EMF) is proportional to the square root of the channel power at a distance d from the E-field source). The distance from the BS for each measurement can be determined according to $d = \sqrt{x^2 + y^2}$, using the receiver locations shown in Table 1. Fig. 4 illustrates how $\Delta\theta$ and d were calculated for a specific receiving antenna.

To confirm that the acquired RF-EMF exposure data have the correct/expected dependence on d, the peak E-field
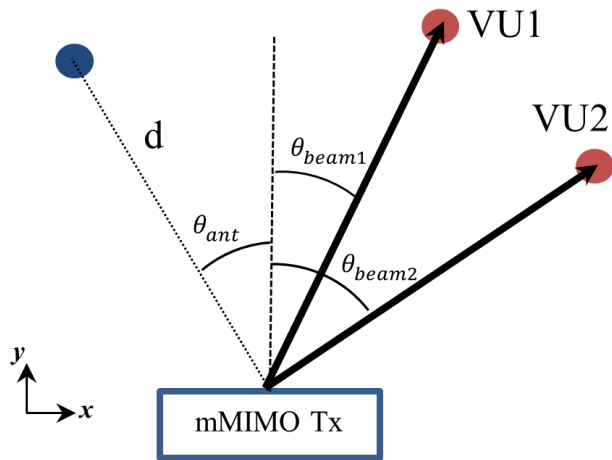
FIGURE 4. Diagram showing the relative angles and distance of one antenna RF-EMF measurements for two active beams.
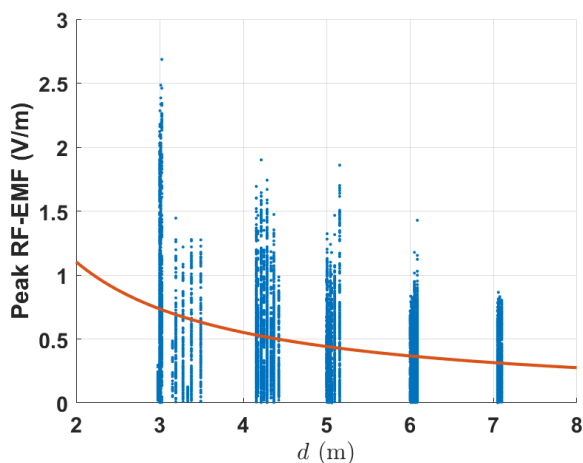

FIGURE 5. Plot showing the dependency of peak electric field of receiver with distance from BS d. A least-squares $1/d$ fit is also plotted.
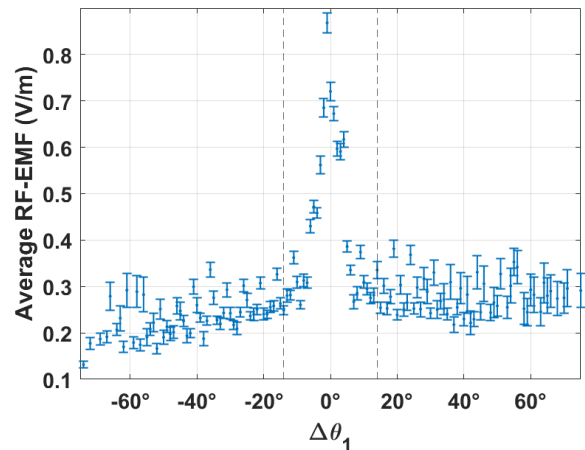

FIGURE 6. Plot of average RF-EMF for $\Delta\theta_1$ with bin widths of 1°. The 95 % confidence interval in the mean RF-EMF is also shown, as well as threshold values $|\Delta\theta| = 14°$.

TABLE 2. Description of input variables. N_{beams} and N_{trans} are the number of beams and transmitters respectively. The subscripts 1-4 for the variables indicate the beam number as represented in Fig. 4.

Input variable	Data Type	Value range
N_{beams}	Discrete	[1, 2, 3, 4]
N_{trans}	Discrete	[32, 64, 96]
Config	Categorical	[indoor, outdoor]
d	Continuous	(2.98 m, 7.11 m)
$|\Delta\theta_1|, \ldots, |\Delta\theta_4|$	Continuous	(0.0°, 14.0°)
$NRB_1, \ldots, NRB_4$	Discrete	[54, 108, 162, 216]
$MCS_1, \ldots, MCS_4$	Categorical	[0, 1, \ldots, 28]

strength (based on the measurements for which $\Delta\theta$ is between $-10°$ and $10°$) is plotted against d in Fig. 5. By performing a least-squares quadratic fit: $E = a+b(1/d)+ c(1/d)^2 + \epsilon_E$, we verified that the data follows the correct $1/d$ behaviour. The p-values for the null hypothesis that the coefficients a and c are zero are much greater than 0.05 ($p_a = 0.25$ and $p_c = 0.83$ for the coefficient t-statistic, respectively), but the b coefficient rejects the null hypothesis with $p_b < 10^{-4}$. The maximum range of $|\Delta\theta|$ for which there is a significant $1/d$ dependency ($p < 0.05$), but no significant constant or $1/d^2$ term ($p > 0.05$) for our measurement dataset is $(0°, 14°)$. This is confirmed by numerical results in Fig. 6, where we see that for angles beyond this range, the signal of the transmitted beam is lost to the environmental noise. Thus to properly model the signal behaviour, for which the correct $1/d$ dependency holds, and avoid environmental boundary effects due to the placement of the antennas, the observed measurements were restricted to the range $|\Delta\theta| \leq 14°$ for all beams, corresponding to 18,774 (out of 162,939)

measurements (observations) in total. We also observe from Fig. 6 symmetry in the E-Field exposure measurements about $\Delta\theta_1 = 0°$, as expected from physical intuition. To ensure symmetry of the model predictions, we take the absolute value $|\Delta\theta_i|$ as input to training the model (see Table 2), so that model predictions for $-\Delta\theta_i$ will be equivalent to $+\Delta\theta_i$. We found that similar predictive performance of RF-EMF is obtained when symmetry is not enforced, and so this symmetry requirement does not significantly reduce model performance.

Along with the key features previously mentioned (i.e. angle, distance), there are in total 16 input variables to our ML model, with their type and range reported in Table 2. All other environmental variables that could affect the RF-EMF exposure, such as walls, bushes, obstacles, sunlight, are taken as noise in the ML model.

B. METHODOLOGY
The choice of ML model to obtain f in (2) depends on the nature of the input variables. Since we have continuous, discrete, and categorical variables, a binary regression tree [30] can be a suitable choice as it can handle varying types of inputs, as well as providing a high level of interpretation of the model.

A binary regression tree works on the principle of nested binary partitions of the observations. All partitions are based on single predictors and at each node in the tree a partition is selected which optimises some metric designed to promote similarity of predictions within each of the two child nodes. Intuitively, a good split is one which maximally reduces the variance of the child nodes, and for regression problems this is often captured by the residual sum-of-squares metric.

However, binary regression trees are weak learners and suffer from over-fitting, and so we instead use a model consisting of an ensemble of decision trees, an approach which is often shown to improve predictive performance [31]. Ensembles of decision trees can be used to directly interpret (compared with deep neural networks) the influence of each predictor on RF-EMF predictions by using feature importance. This is crucial for understanding the individual influence that each predictor has on RF-EMF exposure. Ensembles of decision trees also have a faster computational training speed than deep neural networks, with the latter typically requiring GPUs for model training.

In this work, we use a least-squares gradient boosting algorithm to create the ensemble, which has been shown to improve prediction accuracy compared with random forests [32]. For every new tree in the ensemble, the least-squares boosting algorithm minimises the expression

$$\frac{1}{N} \sum_{i=1}^{N} (y_i - \eta f_{ens}(X_i))^2 , \qquad (4)$$

where X_i, y_i are the input and output training data of observation i, with N observations in total. f_{ens} is the predicted output from the current ensemble, and η is the learning rate. Thus each new tree uses the weights of the previous trees to improve the model fit. The branches and nodes of each tree were determined using the standard CART algorithm [30].

The model has two key hyperparameters which control the bias-variance trade-off in the ML model:

1) The learning rate, η, controls the boosting speed, and ranges from 0 (no boosting, high bias), to 1 (maximal boosting, high variance).
2) The minimum leaf size, l_{min}, which is a parameter of f_{ens}, determines the minimum number of measurements that a leaf of a tree in the ensemble f_{ens} can contain. This ranges from 1 (high variance) to half the number of observations (high bias).

To determine the optimal hyperparameters of the model, the data was first split into a training set (containing 80 % of the measurements), and a test set (containing the remaining 20 %). The split was performed through random sampling, whilst ensuring a similar distribution of indoor and outdoor configurations in both sets. A Bayesian optimisation algorithm was performed over the hyperparameters on the training set to minimise the 5-fold cross-validated mean squared error (MSE) in the model. The optimisation procedure consists of a Gaussian process regression (GPR) model
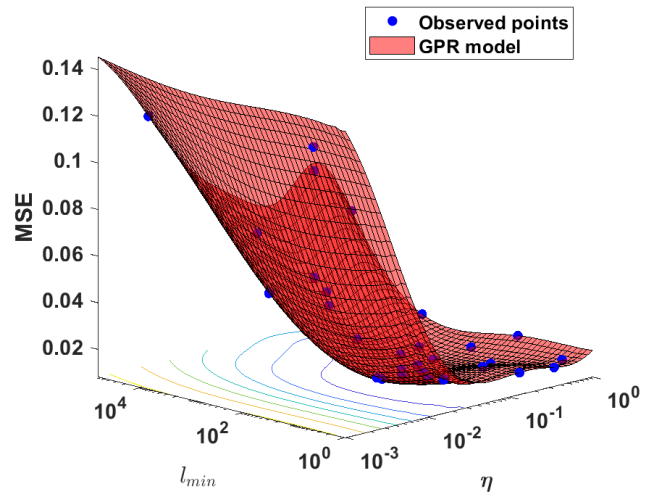

FIGURE 7. Bayesian optimisation of the ML model hyperparameters, using a GPR model fitted over 30 observed points in hyperparameter space, resulting in optimal hyperparameters (η, l_{min}) = (0.10, 67).

of the MSE using an automatic relevance determination (ARD) Matérn 5/2 kernel with Gaussian noise [33]. New observed points are determined by maximising an acquisition function based on the posterior distribution of the GPR model. The acquisition function avoids local minima by modifying the kernel function when the standard deviation of the posterior objective function becomes too small at the observed point (implemented in `Matlab` R2023a using *bayesopt*, Statistics & Machine Learning Toolbox). The optimisation performed on the training data is illustrated in Fig. 7. We observe that there is a unique and smooth global minimum of the GPR model, which implies that the boosted ensemble model accuracy is less sensitive to choice of hyperparameters, and is thus a good choice of ML model for this data.

The boosted ensemble of trees was trained (implemented in `Matlab` R2023a using *fitrensemble*, Statistics & Machine Learning Toolbox) using 100 trees with the optimal hyperparameters calculated in Fig. 7. To interpret the behaviour of the model predictions, the predictor importance estimates [34], which determines how much influence each input variable has on the predicted RF-EMF, were calculated (implemented in `Matlab` R2023a using *predictorImportance*, Statistics & Machine Learning Toolbox) for each input variable of the trained model. The algorithm computes the node importance ni$_i$ for each branch node i using the following equation

$$ni_i = p_i \epsilon_i - p_{l(i)} \epsilon_{l(i)} - p_{r(i)}, \epsilon_{r(i)} \qquad (5)$$

where p_i is the probability that a training observation reaches node i, ϵ_i is the MSE of node i, and $l(i)$ or $r(i)$ are the left or right child node of the branch node i. The predictor importance, pi$_j$ for predictor j is then determined by averaging the node importance across all nodes in the ensemble which
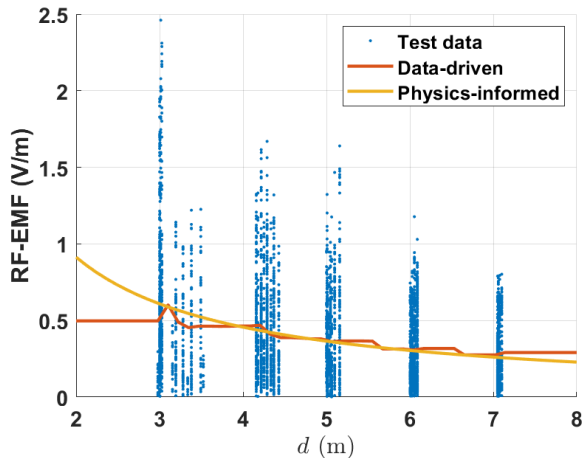
FIGURE 8. Partial dependency plot of ML model predictions of RF-EMF with distance d on the test dataset, by averaging over all the other variables in Table 2.

splits on that predictor, that is

$$\text{pi}_j = \frac{\sum_{i:\text{node } i \text{ splits on predictor } j} \text{ni}_i}{\sum_i \text{ni}_i}. \tag{6}$$

C. PHYSICS-INFORMED MACHINE LEARNING

We know from section III-A and especially from Fig. 5 that RF-EMF is dependent on the distance from the BS to any receivers. However, our method outlined in section III-B to train a ML model does not correctly predict the expected $1/d$ dependency (see Fig. 8) when used directly on the held-out test data. Indeed, the boosted ensemble model predictions have no dependency on distance for distance intervals for which there is no training data, since the ML model only learns from the data available. This inhibits the ability of such models to interpolate or extrapolate the value of the exposure when there is no available training data. One method to solve this issue is to incorporate the distance dependency into the ML training procedure. Incorporating physical laws into machine learning algorithms is an active area of current research, with the main focus on incorporating partial differential equations (PDEs) describing physical systems into neural networks [35], [36], [37]. Given that we have a much simpler physical law ($E \propto 1/d$) in our case, we can take a simpler, but effective approach to adapt the training procedure.

In general, given a set of input variables X_{phys} for which the output y depends on a given physical law, and a remaining set of input variables $X_{stat} = X \setminus X_{phys}$ for which there is a complex dependency on the output y, one can model such a system as a function composition

$$y = g_{X_{phys}}(f(X_{stat})), \tag{7}$$

where $f(X_{stat})$ represents the unknown coefficients of the physical law $y = g_{X_{phys}}$. If there exists an inverse $g_{X_{phys}}^{-1}$ over the range of y, which holds for all values of X_{phys}, then we

can rewrite (7) as

$$g_{X_{phys}}^{-1}(y) = f(X_{stat}), \tag{8}$$

and train the ML model $\hat{f}$. We can then predict new values of y through $\hat{y} = g_{X_{phys}}(\hat{f}(X_{stat}))$.

In our particular example, the physics-informed model is given by
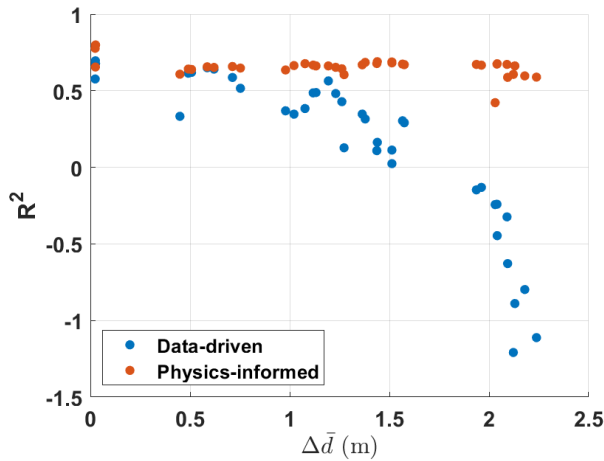
$$E = g_d(f(X_{stat})) = f(X_{stat})/d, \tag{9}$$

which is invertible for $d > 0$ to give $Ed = f(X_{stat})$, where X_{stat} consists of all input variables described in Table 2 except d (which is treated as an independent input). By following the same methodology as in section III-B, the model dependency on distance is shown in Fig. 8. It is clear that by integrating the physical dependency of E on d, our modified model now predicts the correct $1/d$ dependence, and improves the results of the purely data-driven model for distance intervals where there is no training data. In essence, our physics-informed ML model does a better job at interpolating/extrapolating the measured exposure data than the purely data-driven model.

In order to formally evaluate and compare the overall predictive performance of both purely data-driven and physics-informed models, we computed the coefficient of determination, defined as,
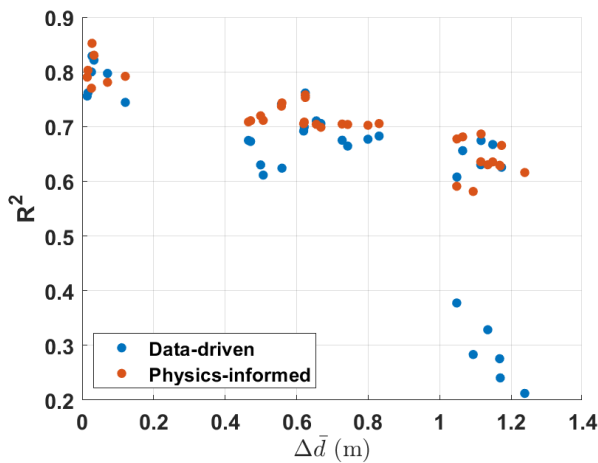
$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{10}$$

on the training set, 10-fold cross-validated set, and the held-out test set. In (10), y_i, $\hat{y}_i$, and $\bar{y}$ are the measured, predicted and mean measured RF-EMF values of observation i, respectively. An R^2 value of 1 corresponds to a perfect prediction, whereas a value of 0 would indicate that the model is no better than a constant model. Table 3, which includes the R^2 results for both models and each type of data set, points out that even though the physics-informed model is more constraining than the purely data-driven model, it exhibits similar predictive performance.

Fig. 9 compares the predictive performance of data-driven and physics-informed models when performing extrapolation and interpolation. In both cases, we divide the data into a training data set and a test set, based on the distance d_{thresh}. For extrapolation, each observation i with distance, d_i, such that $d_i < d_{thresh}$, is placed in the training set, while the others are part of the test set. For interpolation, meanwhile, a training observation i belongs to the training set if $|d_i - 5| > d_{thresh}$, with the remaining observations belonging to the test set.

TABLE 3. A comparison of R^2 values computed using (10). The model hyperparameters were for data-driven: $(\eta, l_{min}) = (0.10, 67)$; and physics-informed: $(\eta, l_{min}) = (0.28, 170)$.

Model type	Training	Cross-validated	Test
Data-driven	0.93	0.85	0.87
Physics-informed	0.91	0.84	0.86

(a) Extrapolation

(b) Interpolation

FIGURE 9. A comparison of extrapolation and interpolation performance between purely data-driven and physics-informed ML models. The R^2 values are calculated using (10) on test data. $\Delta\bar{d}$ is the average distance of all the test data points to their closest training data point.

The distance d_{thresh} is then varied to determine how the the accuracy of the interpolation/extrapolation changes with distance. We see from Fig. 9(a) that the physics-informed model is more robust at extrapolation than its purely data-driven counterpart, especially for larger $\Delta\bar{d}$ distances, since the latter exhibits a steep drop-off in R^2 performance in this case. For interpolation in Fig. 9(b), a similar trend can be observed, where the predictive performance of the physics-informed model is either equal to, or greater than that of the purely data-driven model for larger $\Delta\bar{d}$.

IV. RESULTS

It is clear from Section III-C that the physics-informed ML model is preferable to a purely data-driven approach, so in this section, we report the performance of our physics-informed model in terms of RF-EMF exposure predictions when trained on all of the data prepared in section III-A
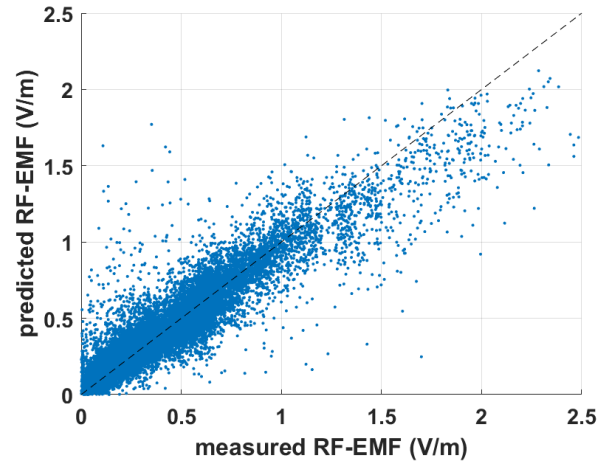

FIGURE 10. Cross-validated predictions of the physics-informed ML model against the original measured RF-EMF values, along with a perfect fit line.

(i.e. no splitting of the data as in Table 3), with its hyperparameters optimised accordingly. Overall, the physics-informed model has a resubstitution R^2 of 0.92 and a 10-fold cross-validated R^2 of 0.86 for RF-EMF predictions, when using the optimal hyperparameters $(\eta, l_{min}) = (0.35, 175)$. The model predictions shown in Fig. 10 show that there is good predictive performance for RF-EMF values up to 1.5 V/m (which contains the bulk of the data), but the model tends to underestimate the RF-EMF for larger values, where there is less training data available.

One of the key advantages of our mMIMO beamforming testbed compared to commercial 5G BSs measurements is the fact that it can measure RF-EMF exposure for various varying factors, as already explained in Section III-B. Fig. 11 determines the importance of each factor/input variable in predicting RF-EMF exposure. The results show that, as expected, the number of beams, beam angle, and number of transmit antennas are the most important factors influencing RF-EMF exposure. This is a useful validation that the model predictions follow physical intuition. However, the predictor ranking gives additional insights that are less expected. The choice of modulation scheme in each beam is of similar importance to the number of resource blocks in determining RF-EMF exposure. It also shows that indoor and outdoor environments, and proximity of different receiving antennas, have far less impact on RF-EMF exposure than the other variables under consideration.

To analyse the sensitivity of the RF-EMF exposure to variation in a subset of the input variables, we compute the partial dependence as defined in (11). Let X^S be a subset of the whole predictor variable set X, and X^C be its complement set, i.e. $X^C = X \setminus X^S$, then, the partial dependence $\hat{f}^S$ using observed data is given by

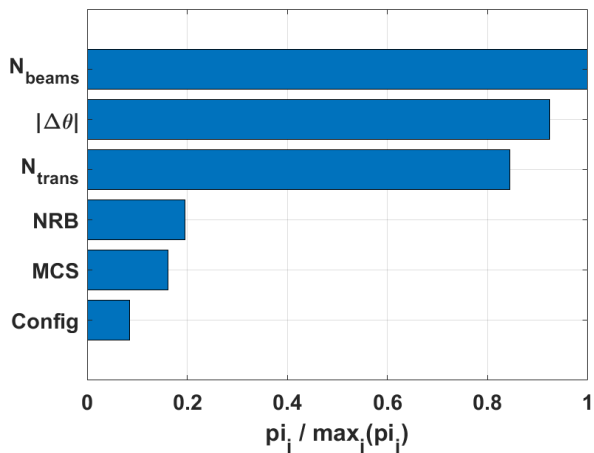$$\hat{f}^S\left(X^S\right) \approx \sum_{i=1}^{N} \hat{f}\left(X^S, X_i^C\right), \tag{11}$$

FIGURE 11. Normalised predictor importance scores calculated using (8), with variable descriptions given in Table 2. The predictor importance scores for $\Delta\theta$, MCS and NRB are averaged over the scores for each beam.

when assuming that each observation $i = 1, \ldots, N$ is equally likely, and where $\hat{f}$ is the trained physics-informed ML model. Note that the data used to compute (11) is generated by uniformly sampling each predictor variable over the ranges given in Table 2.

Fig. 12 and 13 depict the partial dependence of RF-EMF over two variables at a time. We observe an expected clear peak in RF-EMF exposure when $|\Delta\theta_1| = 0°$ (where all other beam angles are averaged), as well as an increase in RF-EMF exposure with NRB and N_{beams}. However, Fig. 13(b) and 13(c), also show opposing effects of NRB and N_{beams} with beam angle. Namely that for small $|\Delta\theta_1|$ values, i.e. when the receiver is in the general direction of a beam, RF-EMF exposure is more dependent on NRB than the number of beams, whereas for larger $|\Delta\theta_1|$ values, it is the other way round.

The benefit of the physics-informed ML model described in section III-C is the ability to predict RF-EMF exposure outside the range of training data. This enables a spatial map to be generated of RF-EMF exposure under different beam configurations. In Fig. 14 we show heatmaps of the average exposure predictions for different number of uniformly sampled input variables (beam configurations). The heatmaps can be interpreted as a time-average of RF-EMF exposure for varying number of users, resource blocks, modulation scheme etc. We see that for small numbers of configurations, the directions of the beams have a large impact on average exposure variation with direction, whereas for large numbers of configurations, the RF-EMF exposure predictions become isotropic (uniform in direction). This isotropy is built in to the ML model through considering $|\Delta\theta|$ (and not θ_{ant}) as model input. However, the high predictive performance of the model on the test data shown in Table 3 suggests that the original experimental data was close to isotropic, and so the heatmaps in Fig. 14 also validate that the ML model is giving the expected predictions.
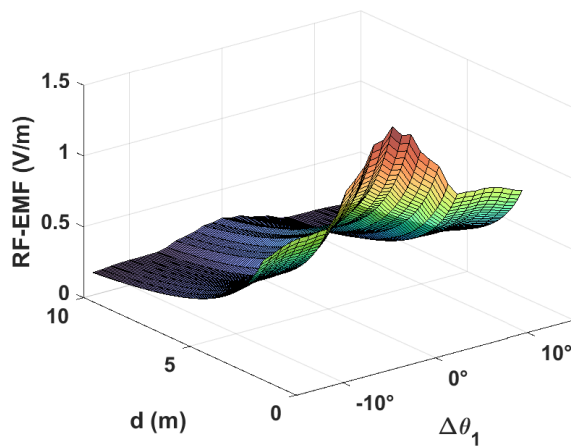
(a)
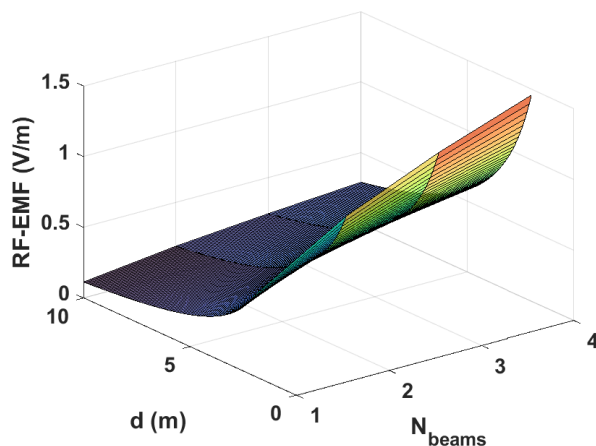
(b)
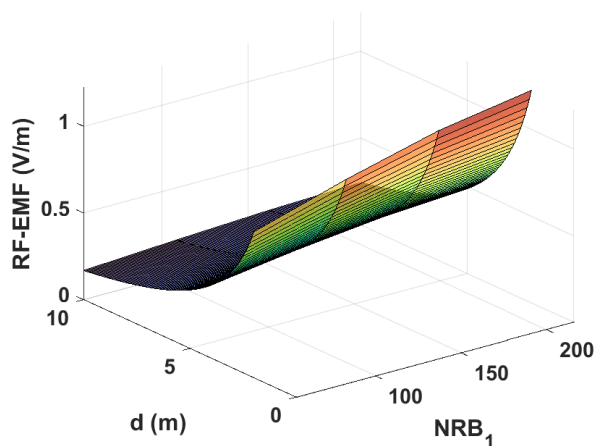
(c)

FIGURE 12. Partial dependence plots of model predictions of a variable against distance from BS d, using (11).

Even though the the RF-EMF exposure is considered to be more statistical/stochastic in 5G, Fig. 14(d) indicates that the exposure becomes more deterministic, as in 4G, when we look at its spatial-temporal average. This could
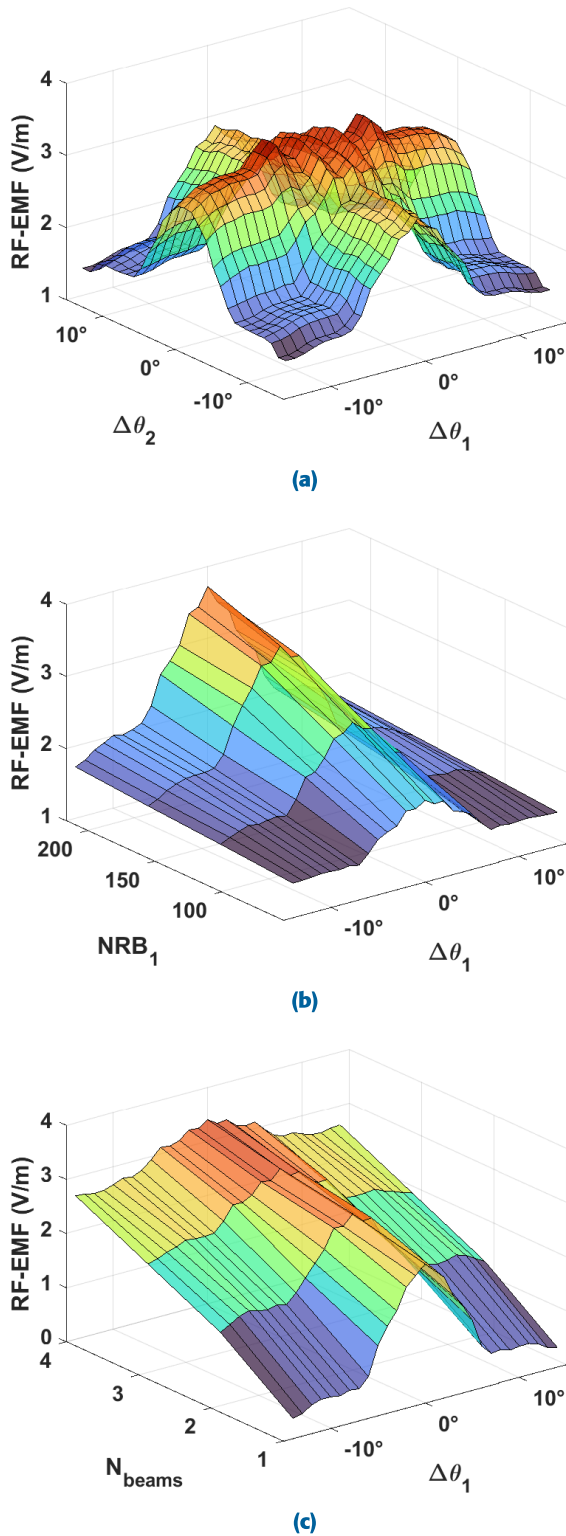
(a)

(b)

(c)

FIGURE 13. Partial dependence plots of model predictions for two variables, using (11). $d = 1$ m in all cases.

have an impact on the definition of the BS exclusion zone, which was traditionally based on the effective isotropic radiated power (EIRP). For instance, based on the results of
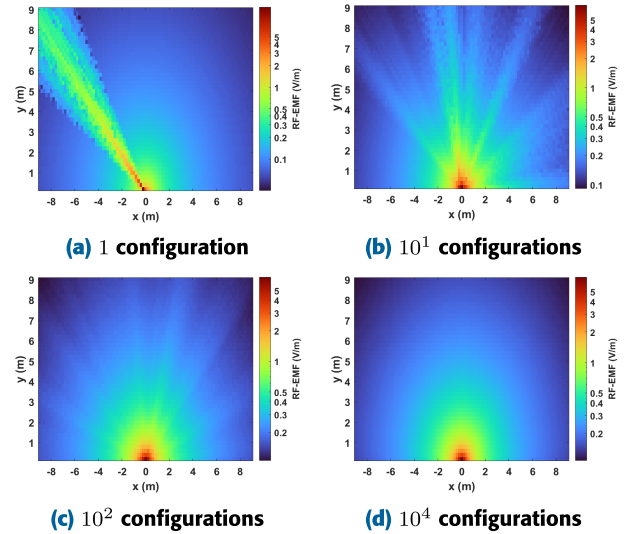

(a) 1 configuration

(b) 10^1 configurations

(c) 10^2 configurations

(d) 10^4 configurations

FIGURE 14. Heatmaps of average RF-EMF exposure using ML model predictions on a random sample of input variables. The mMIMO BS is located at (0,0) m. Input values for $|\Delta\theta|$ and d are calculated for each grid point using (3). Grid points are spaced every 0.2 m in x and y directions.

Fig. 14(a) and 14(d), the radiated power within the single beam in Fig. 14(a) is more than two times larger than the radiated power averaged over multiple beams in Fig. 14(d). Given that the BS exclusion distance is based on the square-root of the EIRP, it can result in a difference of more than 40 % if this distance is set according to Fig. 14(a) and 14(d). In that sense, by using our ML method in conjunction with measurement data, we can predict the spatial-time average exposure around a BS and, it turn, use this data to define a less stringent BS exclusion zone.

V. CONCLUSION

In this paper we propose a novel interpretable and general-isable physics-informed ensemble of decision trees model, trained on real-world experimental data, for evaluating the RF-EMF exposure of 5G BSs. The physics-informed model exhibits high predictive performance, with a cross-validated R^2 of 0.86, which is comparable in accuracy to other ML models presented in the literature, even though the model is more constrained by the physical law. The model has the additional ability to extrapolate beyond the training data to predict the level of exposure at distances further from the BS. Such a model architecture could also be used for predicting the RF-EMF exposure of 6G systems, by taking into account other propagation aspects related to higher frequency transmission.

The model predictions show that the number of transmitters, beam angle, and number of virtual users have the greatest effect on RF-EMF exposure. The predictions also show that number of resource blocks and modulation scheme have less influence on RF-EMF exposure, with the environmental setting having the least influence. Model predictions also show that as the number of beam configurations becomes large, the RF-EMF exposure becomes isotropic (due to

statistical averaging of large numbers), essentially following the Friis formula in all directions. This can help to simplify the exposure measurement methodology of 5G systems.

ACKNOWLEDGMENT

REFERENCES

[1] T. H. Loh, F. Heliot, D. Cheadle, and T. Fielder, "An assessment of the radio frequency electromagnetic field exposure from a massive MIMO 5G testbed," in *Proc. 14th Eur. Conf. Antennas Propag. (EuCAP)*, Mar. 2020, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9135291

[2] T. H. Loh, D. Cheadle, F. Heliot, A. Sunday, and M. Dieudonne, "A study of experiment-based radio frequency electromagnetic field exposure evidence on stochastic nature of a massive MIMO system," in *Proc. 15th Eur. Conf. Antennas Propag. (EuCAP)*, Mar. 2021, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9411325

[3] F. Heliot, D. Cheadle, Y. Gui, and T. H. Loh, "Metrology for RF exposure from massive MIMO system," in *Metrology for 5G and Emerging Wireless Technologies*. Edison, NJ, USA: IET, Dec. 2021, pp. 675–701.

[4] F. Héliot, T. H. Loh, D. Cheadle, Y. Gui, and M. Dieudonne, "An empirical study of the stochastic nature of electromagnetic field exposure in massive MIMO systems," *IEEE Access*, vol. 10, pp. 63100–63112, 2022. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9794655

[5] B. Thors, A. Furuskär, D. Colombi, and C. Törnevik, "Time-averaged realistic maximum power levels for the assessment of radio frequency exposure for 5G radio base stations using massive MIMO," *IEEE Access*, vol. 5, pp. 19711–19719, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8039290

[6] International Commission on Non-Ionizing Radiation Protection (ICNIRP), "Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz)," *Health Phys.*, vol. 118, no. 5, pp. 483–524, May 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32167495/

[7] J. T. Bushberg, C. K. Chou, K. R. Foster, R. Kavet, D. P. Maxson, R. A. Tell, and M. C. Ziskin, "IEEE committee on man and radiation—COMAR technical information statement: Health and safety issues concerning exposure of the general public to electromagnetic energy from 5G wireless communications networks," *Health Phys.*, vol. 119, no. 2, pp. 236–246, Aug. 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32576739/

[8] D. Urbinello, W. Joseph, A. Huss, L. Verloock, J. Beekhuizen, R. Vermeulen, L. Martens, and M. Röösli, "Radio-frequency electromagnetic field (RF-EMF) exposure levels in different European outdoor urban environments in comparison with regulatory limits," *Environ. Int.*, vol. 68, pp. 49–54, Jul. 2014. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24704639/

[9] P. Baracca, A. Weber, T. Wild, and C. Grangeat, "A statistical approach for RF exposure compliance boundary assessment in massive MIMO systems," in *Proc. WSA 22nd Int. ITG Workshop Smart Antennas*, Mar. 2018, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8385449

[10] *Determination of RF Field Strength, Power Density and SAR in the Vicinity of Base Stations for the Purpose of Evaluating Human Exposure*, document IEC 62232:2022, Oct. 2022. [Online]. Available: https://webstore.iec.ch/publication/64934

[11] C. Törnevik. (Dec. 2017). *Impact of EMF Limits on 5G Network Roll-Out*. [Online]. Available: https://emraustralia.com.au/blogs/news/impact-of-emf-limits-on-5g-network-roll-out

[12] IEC. (Apr. 2019). *Case Studies Supporting IEC 62232—Determination of RF Field Strength, Power Density and SAR in the Vicinity of Radiocommunication Base Stations for the Purpose of Evaluating Human Exposure*. [Online]. Available: https://www.emf-portal.org/en/article/38267

[13] H. Keller, "On the assessment of human exposure to electromagnetic fields transmitted by 5G NR base stations," *Health Phys.*, vol. 117, no. 5, pp. 541–545, Nov. 2019.

[14] METAS, "Measurement method for 5G NR base stations up to 6 GHz," METAS, Menlo Park, CA, USA, Tech. Rep. 154.1-2020-5218-1016, Feb. 2020. [Online]. Available: https://www.metas.ch/metas/en/home/dok/publikationen/meldungen/2020-02-18.html

[15] D. Colombi, P. Joshi, B. Xu, F. Ghasemifard, V. Narasaraju, and C. Törnevik, "Analysis of the actual power and EMF exposure from base stations in a commercial 5G network," *Appl. Sci.*, vol. 10, no. 15, p. 5280, Jul. 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/15/5280

[16] S. Aerts, L. Verloock, M. Van Den Bossche, D. Colombi, L. Martens, C. Törnevik, and W. Joseph, "In-situ measurement methodology for the assessment of 5G NR massive MIMO base station exposure at sub-6 GHz frequencies," *IEEE Access*, vol. 7, pp. 184658–184667, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8937514

[17] J. Tanveer, A. Haider, R. Ali, and A. Kim, "Machine learning for physical layer in 5G and beyond wireless networks: A survey," *Electronics*, vol. 11, no. 1, p. 121, Dec. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/11/1/121

[18] S. Bilson and A. Thompson, "A review of applications of supervised learning to future networks," NPL, U.K., Report/Guide MS 40, Oct. 2022, doi: 10.47120/npl.MS40.

[19] H.-S. Jo, C. Park, E. Lee, H. K. Choi, and J. Park, "Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process," *Sensors*, vol. 20, no. 7, p. 1927, Mar. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/7/1927

[20] C. Huang, R. He, B. Ai, A. F. Molisch, B. K. Lau, K. Haneda, B. Liu, C.-X. Wang, M. Yang, C. Oestges, and Z. Zhong, "Artificial intelligence enabled radio propagation for communications—Part II: Scenario identification and channel modeling," *IEEE Trans. Antennas Propag.*, vol. 70, no. 6, pp. 3955–3969, Jun. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9713743

[21] C. Huang, R. He, B. Ai, A. F. Molisch, B. K. Lau, K. Haneda, B. Liu, C.-X. Wang, M. Yang, C. Oestges, and Z. Zhong, "Artificial intelligence enabled radio propagation for communications—Part I: Channel characterization and antenna-channel optimization," *IEEE Trans. Antennas Propag.*, vol. 70, no. 6, pp. 3939–3954, Jun. 2022.

[22] J. Gante, G. Falcão, and L. Sousa, "Beamformed fingerprint learning for accurate millimeter wave positioning," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8690987

[23] J. Bang and J. H. Kim, "Predicting power density of array antenna in mmWave applications with deep learning," *IEEE Access*, vol. 9, pp. 111030–111038, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9507505

[24] G. Tognola, D. Plets, E. Chiaramello, S. Gallucci, M. Bonato, S. Fiocchi, M. Parazzini, L. Martens, W. Joseph, and P. Ravazzani, "Use of machine learning for the estimation of down- and up-link field exposure in multi-source indoor WiFi scenarios," *Bioelectromagnetics*, vol. 42, no. 7, pp. 550–561, Oct. 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/bem.22361

[25] T. Mazloum, S. Wang, M. Hamdi, B. A. Mulugeta, and J. Wiart, "Artificial neural network-based uplink power prediction from multi-floor indoor measurement campaigns in 4G networks," *Frontiers Public Health*, vol. 9, Nov. 2021, Art. no. 777798. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8669482/

[26] S. Wang, T. Mazloum, and J. Wiart, "Prediction of RF-EMF exposure by outdoor drive test measurements," *Telecom*, vol. 3, no. 3, pp. 396–406, Jun. 2022. [Online]. Available: https://www.mdpi.com/2673-4001/3/3/21

[27] H. T. Friis, "A note on a simple transmission formula," *Proc. IRE*, vol. 34, no. 5, pp. 254–256, May 1946. [Online]. Available: https://ieeexplore.ieee.org/document/1697062

[28] University of Surrey. *5G/6G Innovation Centre*. Accessed: Mar. 8, 2023. [Online]. Available: https://www.surrey.ac.uk/institute-communication-systems/5g-6g-innovation-centre

[29] Keysight. *PathWave System Design (SystemVue)*. Accessed: Mar. 8, 2023. [Online]. Available: https://www.keysight.com/gb/en/products/software/pathwave-design-software/pathwave-system-design-software.html

[30] L. Breiman, *Classification and Regression Trees*. New York, NY, USA: Routledge, Oct. 2017.

[31] L. Breiman, "Random forests," *Machine Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[32] T. Hastie, R. Tibshirani, and J. Friedman, "Boosting and additive trees," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), T. Hastie, R. Tibshirani, and J. Friedman, Eds. New York, NY, USA: Springer, 2009, pp. 337–387, doi: 10.1007/978-0-387-84858-7_10.

[33] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, vol. 25. Red Hook, NY, USA: Curran Associates, 2012. [Online]. Available: https://papers.nips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

[34] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003. [Online]. Available: https://dl.acm.org/doi/10.5555/944919.944968

[35] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021999118307125

[36] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, no. 6, pp. 422–440, May 2021. [Online]. Available: https://www.nature.com/articles/s42254-021-00314-5

[37] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific machine learning through physics–informed neural networks: Where we are and what's next," *J. Sci. Comput.*, vol. 92, no. 3, p. 88, Jul. 2022, doi: 10.1007/s10915-022-01939-z.

SAMUEL BILSON received the Ph.D. degree in theoretical physics from the University of Durham, U.K., in 2011. Since then, he has been a Machine Learning Data Scientist in both industry and academia, including a Postdoctoral Research Fellowship in statistical epidemiology with the Department of Mathematics, University of Warwick, and as a Senior Training Engineer with The MathWorks. He joined the Department of Data Science, National Physical Laboratory (NPL), as a Higher Scientist, in 2022, specializing in applied machine learning. He has authored five refereed publications and has extensive experience in teaching and industry applications of machine learning. His research interests include trustworthy and physics-informed machine learning.
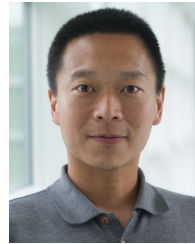
TIAN HONG LOH (Senior Member, IEEE) received the Ph.D. degree in engineering from the University of Warwick, U.K., in 2005. He has been with the National Physical Laboratory (NPL), U.K., since 2005. He is currently a Principal Scientist with NPL. He leads work with NPL on a wide range of applied and computational electromagnetic metrology research areas to support the telecommunications industry. He is also a Visiting Professor with the University of Surrey, U.K., and the President of the International Union of Radio Science (URSI) U.K. Board, Chair of the European Association on Antennas and Propagation (EurAAP) Measurement Working Group. He holds seven patents, one book, and nine book chapters. He has authored or coauthored over 200 refereed publications. His research interests include 5G/6G communications, smart antennas, small antennas, metamaterials, body-centric communications, wireless sensor networks, electromagnetic compatibility, and computational electromagnetics.

FABIEN HÉLIOT (Member, IEEE) is currently a Lecturer in wireless communications with the Institute for Communication Systems (ICS), University of Surrey. He has published in excess of 100 peer-reviewed articles, book chapters, and technical reports in diverse areas of communications theory and signal processing for wireless communications. He has lead the EPRSC funded project "Electromagnetic field exposure reduction/avoidance for the next generations of wireless communication systems" and has been actively involved in other funded U.K./EU projects related to the topic of this work, such as metrology for RF EXposure from 5G base station (5GRFEX) and Low Exposure Future Networks (LExNet). He received an Exemplary Reviewer Award from the IEEE COMMUNICATIONS LETTERS, in 2011, and the Best Paper Award from the EuCNC Conference, in 2011, 2021, and 2022.

ANDREW THOMPSON received the Ph.D. degree in mathematics from the University of Edinburgh, U.K., in 2012. He is currently a Senior Scientist in Data Science with the National Physical Laboratory (NPL). His research interests include inverse problems and machine learning. He leads NPL's work on trustworthy machine learning. Before joining NPL, he was a Visiting Assistant Professor with Duke University and a Departmental Lecturer with the Mathematical Institute, University of Oxford.

● ● ●