Enhanced Wireless Channel Estimation Through Parametric Optimization of Hybrid Ray Launching-Collaborative Filtering Technique

FRAN CASINO, (Member, IEEE), PEIO LOPEZ-ITURRI, (Member, IEEE), ERIK AGUIRRE, LEYRE AZPILICUETA, (Senior Member, IEEE), FRANCISCO FALCONE, (Senior Member, IEEE), AND AGUSTI SOLANAS, (Senior Member, IEEE)

1Department of Informatics, University of Piraeus, 18534 Piraeus, Greece
2Electric, Electronic and Communication Engineering Department, Public University of Navarre, 31006 Pamplona, Spain
3Institute of Smart Cities, Public University of Navarre, 31006 Pamplona, Spain
4School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey 64849, Mexico
5Department of Computer Engineering and Mathematics, Rovira i Virgili University, 43007 Tarragona, Spain

Corresponding author: Francisco Falcone (francisco.falcone@unavarra.es)

This work was supported in part by the European Commission LOCARD Project under Grant 832735 and under Project RTI2018-095499-B-C32 and Project RTI2018-095499-B-C31, and in part by the Ministerio de Ciencia, Innovación y Universidades, Gobierno de España (MCIU/AEI/FEDER, UE). The work of Agusti Solanas was supported in part by the Government of Catalonia (GC) under Grant 2017-DI-002 and Grant 2017-SGR-896, and in part by the Fundació PuntCAT with the Vinton Cerf Distinction.

Abstract In this paper, an enhancement of a hybrid simulation technique based on combining collaborative filtering with deterministic 3D ray launching algorithm is proposed. Our approach implements a new methodology of data depuration from low definition simulations to reduce noisy simulation cells. This is achieved by processing the maximum number of permitted reflections, applying memory based collaborative filtering, using a nearest neighbors’ approach. The depuration of the low definition ray launching simulation results consists on discarding the estimated values of the cells reached by a number of rays lower than a set value. Discarded cell values are considered noise due to the high error that they provide comparing them to high definition ray launching simulation results. Thus, applying the collaborative filtering technique both to empty and noisy cells, the overall accuracy of the proposed methodology is improved. Specifically, the size of the data collected from the scenarios was reduced by more than 40% after identifying and extracting noisy/erroneous values. In addition, despite the reduced amount of training samples, the new methodology provides an accuracy gain above 8% when applied to the real-world scenario under test, compared with the original approach. Therefore, the proposed methodology provides more precise results from a low definition dataset, increasing accuracy while exhibiting lower complexity in terms of computation and data storage. The enhanced hybrid method enables the analysis of larger complex scenarios with high transceiver density, providing coverage/capacity estimations in the design of heterogeneous IoT network applications.

Index Terms Collaborative filtering, 3-D ray launching, pattern recognition, wireless channel.

I. INTRODUCTION

Wireless communication systems are one of the main enablers of highly interactive scenarios in future Heterogeneous Network architectures. In this context, multiple wireless systems cooperate, using physical layer mechanisms and network supported services, to optimize systems’ operation in terms of overall interference minimization and energy consumption reduction. In this sense, foreseen applications within Internet of Things (IoT) and within 5G communications, such as Device to Device (D2D) or machine type communications (MTC), will substantially increase transceiver density [1], [2]. This increase in the number of wireless sources and inherent limitation owing to size and form factors (wearables, embedded transceivers) requires radio channel as well as system planning tasks to optimize coverage/capacity
relations and hence overall quality of service metrics, mainly constrained by interference. Deterministic wireless channel modelling techniques, such as ray launching or ray tracing, provide accurate results both in terms of received power estimation and time-dependent variables (e.g., power delay profiles or delay spread distributions). The main drawback of those techniques is the potentially large computational cost, which is dependent on scenario size, consideration of detailed scenario topology and the inclusion of additional effects, such as diffraction or diffuse scattering [3], [4].

Recently, approaches based on artificial intelligence have been explored in the field of electromagnetic analysis, in multiple fields such as EM scattering, inverse Scattering, direction of arrival estimation, radar and remote sensing [5], as well as in other network oriented aspects, such as dynamic resource allocation algorithms [6]. In the case of wireless channel modelling, some works on channel estimation have been presented in relation with: massive MIMO systems [7], node distribution in wireless sensor networks [8], machine learning assisted path loss prediction [9], empirical connectivity model for an extended monitoring network of environmental parameters optimized by machine learning on an extensive data set [10], low altitude propagation model based on a machine learning approach [11], or an enhanced empirical propagation model combined with machine learning techniques from extensive measurement sets in the UHF focused on coverage analysis of DTV systems [12], among others. Future trends are foreseen for upcoming beyond 5G systems, in which multi state, multi-dimensional networks can be analyzed and optimized by the aid of quantum machine learning techniques [13]. Different applications within wireless channel characterization, resource allocation optimization or system level enhancement [14]–[20], are presented in Table 1.

Deterministic based methods can provide accurate wireless channel estimations for complex scenarios with high density of constitutive elements, particularly for the case of indoor scenarios. However, as previously stated, accuracy comes with the tradeoff of high computational cost. This is given mainly by a precise definition of the elements within the scenario under consideration and the discretization level of the physical propagating wave front and the equivalent set of rays within the defined solid angle in the case of volumetric approaches. In order to reduce computational complexity, several approaches have been proposed, based on the combination of deterministic Ray Launching techniques with other approaches, such as neural networks and the electromagnetic diffusion equation [21], [22]. This has given rise to the use of hybrid simulation, an approach that has provided improved results in elements such as tracking in nonlinear systems, supported by fuzzy systems [23]. In [24], we proposed the combination of in-house 3D Ray Launching (3D RL) code with collaborative filtering (CF) recommender systems [25]. The main idea was to use the ability of CF methods to predict rates and infer the values of empty cells in matrices obtained in low definition (LD) simulations, reducing the computational complexity of high definition (HD) simulations. The proposed methodology was applied to RF power distribution in complete volumes of several scenarios, and it could be extended to other parameters as well.

In this article, we present an optimization methodology to decrease computational cost, based on the analysis on the permitted maximum number of rebounds (NR) of the launched rays, initially described in [26]. The study of multiple simulation databases (created with different NR values) enables the

| Ref  | Overview                                                                 | Description of Solution Implemented                                                                 |
|------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| [9]  | Path loss prediction based on machine learning                          | Improvement as compared with conventional empirical based RF path loss estimation techniques, by using different machine learning approaches. |
| [10] | Enhanced connectivity model in large scale low power wireless sensor networks | A radio propagation model based on 2.4GHz ISM band is presented for distributed wireless sensor network connectivity, scalable and adaptive to other network conditions with the aid of machine learning techniques |
| [12] | Optimized UHF propagation model based on machine learning techniques     | Based on an extensive set of measurements and the application of hybrid KNN-KBT techniques, a radio propagation model for DTV system modeling is presented. |
| [14] | Deep neural networks to support multi-channel cognitive radio operation  | Transmit power allocation for secondary users minimizing interference for primary users in cognitive radio schemes. |
| [15] | Optimal resource allocation in wireless systems supported by deep neural networks | General purpose approach in the use of deep neural networks, with examples provided in AWGN channel capacity and Interference channel conditions |
| [16] | Optimized MAC framework based on deep neural networks user detection for Het-LoT operation | Implementation of MAC functionalities to handle coexistence between IoT-Wi-Fi devices in ISM band, with enhanced channel usage information. |
| [17] | Robust modulation employing a deep neural network - recursive neural network-based classifier | A novel classifier scheme is proposed, which is robust in the case of considering a non-ideal noisy SISO channel, reducing training process time. |
| [18] | Pedestrian dead reckoning indoor positioning systems implemented with deep neural networks | Indoor location system implemented over Wi-Fi signals fused with Smartphone opportunistic sensor data, with the aid of deep neural network-factor graph model |
| [19] | Joint Optimization in IoT Fog Enabled systems of multiple characteristics based on the use of deep neural networks | By using multiple deep neural networks, multiple variables such as memory caching, edge computing and radio resource management can be jointly enhanced, with the main goal of reducing end to end delay |
| [20] | Improvement of radio resource management in vehicular 5G networks supported by machine learning techniques | By using different machine learning based strategies (i.e., deep neural network -recursive neural network), radio resource management is enhanced in a 5G SDN/NFV network implementation. |
analysis of the minimum number of rays per simulation cell, which in turn reduces estimation errors in the LD to HD result association phase. The proposed approach provides increased accuracy, while reducing computational costs regarding HD simulations. The new proposed method increases efficiency by discarding results of those cells in which the total number of rays detected is lower than a set value (NR), acting as an effective threshold. The corresponding threshold is obtained, by means of 3D ray launching simulation analysis. These values exhibit high error as compared to high definition ray launching simulations, being equivalent to noise. The application of collaborative filtering techniques significantly improves the overall accuracy.

The rest of article is organized as follows: Section II provides background on recommender systems and collaborative filtering. Section III describes validation with 3D RL simulation tools and wireless channel measurement results. Section IV describes the hybrid CF-3D RL methodology with optimized NR parameter analysis. Section V presents the experimental validations, ending with the concluding remarks.

II. BACKGROUND
A. RECOMMENDER SYSTEMS AND COLLABORATIVE FILTERING
Nowadays, data management is facing a paradigm shift due to the widespread adoption of cyber-physical systems, which will increase both the volume and the way data are exchanged [27]. As a consequence, real-time services face several changes due to increased regulations, client demands and big data challenges [28]. In this context, automatic recommendation systems [28] are gaining momentum due to their inherent characteristics, which provide manageable and personalized information to users [30], [31]. Collaborative filtering [24] encompasses disparate recommendation methods and is nowadays the most widely used technique due to its adaptability according to the input data. CF relies on the assumption that users that share similar behavior/experience in specific topics will have similar tastes or interests according to some quantifiable metric. Usually, the relationships between users and items are stored in the form of \( n \times m \) matrices (i.e. \( n \) users and \( m \) items), where each cell \((i, j)\) stores the evaluation of user \( i \) on item \( j \). Fig. 1 shows an example of such data representation.

The literature classifies CF methods into three main categories according to the data they manage [24], [32]: (i) Memory-based, which use all the available data about users, items and relationships, (ii) model-based, which create a model (e.g. by using machine learning, dimensionality reduction or statistical models) from the complete set, and (iii) hybrid-based, which incorporate other data sources (e.g. social networks, demographic data). Nevertheless, despite the benefits provided by CF methods, there are several challenges that such systems need to face, being the most acute the cold start, scalability, sparseness and privacy issues [32]–[36]. For more on CF, we point the interested reader to [32], [37] for a review of the state-of-the-art and the most relevant advances and trends.

In this paper, we adopt the most well-known memory-based CF variant with the nearest neighbors approach (KNN), where users compute their similarities according to a metric (e.g. Euclidean distance, cosine similarity) to find which are their closest neighbors (i.e. their corresponding most similar profiles). Therefore, given a pattern with inconsistent/erroneous values, we will select its \( k \) most similar patterns (according to a ground truth database) to infer/predict the RF power level [38], [25].

III. 3D RAY LAUNCHING TOOL VALIDATION
As stated previously, an in-house 3D Ray Launching algorithm has been used in this work to predict radio wave propagation in a complex indoor environment. The proposed algorithm is a geometry-based deterministic approach where different parameters can be considered as inputs, namely the number of reflections, operation frequency, transmitted power, bit rate, angular and spatial resolution and the radiation pattern of the considered antennas. A detailed 3D scenario is created considering all the obstacles within it, by means of the conductivity and relative permittivity of all the materials at the frequency of operation of the system. The main drawback of these methods is their high computational complexity due to 3D space analysis. To overcome this problem, several articles in the literature analyze convergence analysis of different approaches in terms of the number of reflections or the launching ray’s density. In [39], a quasi-analytical ray propagation model to obtain the RF field within an aircraft cabin is proposed. The convergence analysis of the algorithm is presented in terms of rays’ propagation time and number of bounces, showing that the high content of metallic parts inside the cabin shows a slow rate of convergence. Another study in [40] presents a hybrid method of GO/PO and physical theory of diffraction where the dependence of field convergence on the maximum number of reflections is

![FIGURE 1. Example of data matrix. Each row corresponds to a user profile, and each column corresponds to an item.](image-url)
investigated. The work in [41] proposes a ray density normalization within a novel stochastic ray launching approach to accurately predict signal levels in curved geometries. In [42], a mixed ray launching/tracing method propagation modelling for large areas is proposed, analyzing the convergence of the number of reflections and diffractions of such areas. The study in [43] analyzes the number of rays to achieve convergence in a ray tracing approach for RCS modeling of large complex objects.

In the same way, the convergence analysis in terms of launching rays and number of reflections has been performed for the in-house 3D ray launching algorithm, and it is presented in [44], showing the optimal parameters to be used in the algorithm to achieve good accuracy with affordable computational time. Considering these parameters, in this section, the 3D RL validation is presented for a scenario under test (SUT).

The SUT is ‘Laboratori 231’, an indoor scenario of 8.8m × 4.7m × 3.7m, located at the School of Engineering of the Rovira i Virgili University, in Tarragona, Catalonia, Spain (cf., Fig. 2). As it can be seen, the scenario is a small conference hall where mainly tables and chairs are present. Due to the reduced size of the scenario and the high density of obstacles, the scenario is a complex one in terms of multipath propagation components.

In order to perform the validation of the proposed 3D RL tool, the SUT has been created for its simulation (cf., Fig. 3 and Fig. 5). The dimensions of the scenario, the shapes and sizes of the elements within, and their material properties have been set as close as possible to the real scenario. Table 2 shows the dispersive material properties used in the simulations.

The used 3D RL parameters are summarized in Table 3. Note that the parameters have been chosen in order to suit the equipment employed in the measurement campaign: As a transmitter, an XBee mote (ZigBee) with a whip antenna. As a receiver, a monopole antenna (Titanis 2.4 GHz Swivel SMA Antenna from Antenova) coupled to an Agilent FieldFox
TABLE 3. Ray Launching simulation parameters.

| Parameter                        | Value         |
|----------------------------------|---------------|
| Operation Frequency              | 2.4 GHz       |
| Antenna Type                     | Monopole      |
| Antenna Gain                     | 1.2 dB        |
| Transmitted Power                | 0 dBm         |
| Cuboid size (Mesh resolution)    | 10 cm x 10 cm x 10 cm |
| Launched rays angular resolution | 1 degree      |
| Permitted maximum rebounds       | 6             |
| Diffraction phenomenon           | Activated     |

N9912A spectrum analyzer. The location of the transmitter (red dot) and the measurement points (green dots) are represented in Fig. 3.

Fig. 4 shows the comparison between the measured RF power level and the estimated values obtained by the 3D RL algorithm for 12 different measurement points. As expected, the results show good agreement, with a mean error of 0.23 dB with a standard deviation of 2.43 dB. Therefore, the simulation tool has been considered validated to be used in this SUT in order to perform the hybrid ray launching-collaborative filtering technique study proposed in this paper.

IV. METHODOLOGY

The proposed optimization is applied to LD simulations performed by the presented 3D RL algorithm for both creation of DBs and simulation of the SUT, which is where the whole improved methodology to obtain RF power distribution is applied.

The permitted maximum NR in a 3D RL simulation gives the maximum number of interactions between launched rays (from the transmitter) and the obstacles within the scenario. This parameter is fixed by the user before the simulation. The NR affects the accuracy of the obtained results, being more accurate the higher this number is. However, a higher NR implies more calculation time. Besides, the accuracy of the results tends to converge at a specific NR, which for the kind of indoor scenarios evaluated is six rebounds. In other words, for NR > 6 the accuracy improvement is negligible, but the required calculation time increases significantly. As an illustrative example of the effect of the permitted maximum NR, Fig. 5 shows the RF power distribution estimations obtained at height 2m of the SUT, for NR = 0 (LD1), NR = 2 (LD3), NR = 4 (LD5) and NR = 6 (LD7). In the previous methodology [23], we built DBs using LD simulations with NR = 3. However, in this new approach, we use a set of indoor scenarios, described in Section III.A, to create DBs from NR = 0 to NR = 6. Each scenario has been simulated in LD for the seven cases.

NR affects the accuracy of LD simulation results because it plays a role in the number of rays that reach each simulation cell: the smaller NR, the smaller the number of rays that reach each cell, and the higher the number of cells that are not reached by any launched ray. These empty cells (i.e., cells without an RF power level) in LD simulations are filled by our 3D RL–CF method. However, empty cells are not the only problem. We detected non-empty cells that lessen the accuracy of the 3D RL–CF results. These cells usually exhibit a very low RF power level, because they have been reached by a very low number of rays during the simulation. To address this problem, we propose a depuration method, explained in Section III.C, to suppress cells with inaccurate values.

A comparison between the previous and the new methodologies is shown in Fig. 6. The depuration process is highlighted in orange.

A. DATA COLLECTION – DATABASES CREATION

Ten scenarios have been defined to build the DBs needed to apply the proposed methodology. All scenarios are indoor and similar in terms of morphology and density, like in [3]. In this case, a different PC has been used for the simulations: Intel (R) Core(TM) i5-4690 CPU @ 3.50GHz, with 32 GB RAM. The features of the scenarios used to build the DBs are summarized in Table 4. The Size column shows the l: length, w: width and h: height of the scenario, and column Density shows the percentage of the volume of the scenario which is occupied by objects (i.e., not air). It is important to note that the features of the scenarios used to build the DBs limit the scenarios that can be analyzed. As can be seen in Table 4, the chosen SUT features are between the limits of the DB scenarios. Moreover, the presented methodology can be applied to every scenario as long as the DBs are built accordingly to the required features.

Each scenario has been simulated in LD using 3D RL with several values for NR (from 0 to 6). A total of 70 simulations have been used to create seven DBs (i.e., {DBLDi, ∀i ∈ [1,7]}), each containing the results of simulations with a different NR value. Fig. 7 shows the simulation time for all cases.
In this section, we detail the cumulative distributions obtained for all scenarios in terms of number of rays $R_i$ per cell, MAE and sparseness. Note that we represent the aggregate numbers after applying LD simulations using NR values from 0 to 6. These data are used to determine the value of $R_{\text{min}}$ in Section III.C.

Fig. 8 shows the aggregate distribution of number of cells that had a specific number of rays $R_i$. We can observe that, as expected, the higher the number of rays the lower the number of cells. Such outcome is related with the one depicted in Fig. 9. In this case, we show the cumulative percentage of sparseness of the scenarios according to all values. For example, in the case of $R_i = 0$, we have a 0% of sparseness and thus, all cells have at least this number of incident rays. In the case of $R_i = 1$, we observe that only a 69% of all cells have more than $R_i > 1$. In the case of $R_i = 30$, we observe a sparseness value above 97%, which means that less than 3% of cells registered a $R_i > 30$. The outcomes state that most of the cells of the system have a low number of $R_i$. Considering that $R_{\text{min}}$ is set to 4 as later discussed in Section III.C, this means that in average we discard more than 60% of values (cf. Fig. 9) due to their bad quality when
performing LD simulations. In this regard, Fig. 10 shows the maximum absolute error (MAE) of all values according to each $R_i$. It is obvious that the higher $R_i$ the less error in the values. However, as seen in Fig. 9, selecting only cells with high $R_i$ would lead to a high sparseness, hindering the prediction process.

### C. DATA DEPURATION PROCESS

We state that if the number of rays, $R_i$, passing through a cell, $C_i$, during the simulation is small, the value in that cell will be inaccurate. With the aim to identify those non-empty cells with small $R_i$, containing inaccurate values, we set a quality threshold for the minimum number of rays, $R_{\text{min}}$, which should pass through a cell to consider its value valid. $R_{\text{min}}$ is set once for the whole volume of the scenarios.

It is worth emphasising the difference between the number of rebounds, $NR$, and the number of rays, $R_i$, in a cell. Although they are closely related concepts, the former determines how many times rays interact with objects in the simulation, while the latter counts the number of rays that pass through each cell.

In order to determine $R_{\text{min}}$, we analyse the error (i.e., the difference between obtained values in LD and HD simulations) and the sparseness of LD simulations for several values of $R_{\text{min}}$. First, we compute the error of all scenarios using the mean absolute error (MAE) as follows:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |HD_i - LD_i|}{n}$$

(1)

where $n$ is the number of non-empty cells, $LD_i$ is the value of the LD simulation for cell $i$, and $HD_i$ is the value of cell $i$ in the HD simulation. We only compare non-empty LD cells with their HD counterparts (i.e., empty cells are discarded).

The error is classified depending on the number of incident rays per cell to obtain an aggregated error for all simulations. Next, in order to compute the sparseness, we classify cells depending on their number of incident rays and we count the percentage of cells that have, at least, a given number of incident rays.

The $R_{\text{min}}$ value is selected to balance the trade-off between sparseness and MAE. In a nutshell, we want to eliminate inaccurate cells, but we have to minimise sparseness to preserve enough patterns/information in the DBs. Note that a high $R_{\text{min}}$ value means better quality simulation results (values) but a smaller number of patterns in the DBs. A shortage of patterns affects CF predictions, since it prevents the recommendation algorithm from finding proper analogies. A simple procedure to determine the value of $R_{\text{min}}$ is to find the discrete value of the minimum intersection point of both functions (i.e., error & sparseness). The result, depicted in Fig. 11, is that the quality threshold value $R_{\text{min}}$ should be set to 4. Therefore, in the depuration process, cells with $R_i < R_{\text{min}} = 4$ are discarded. Fig. 12 shows a graphical example of the refinement process.
Following the procedure described in [24], recommender/CF

DEPURATED

DBR

that in the case of LD1, any value reaches the quality threshold and thus, $DBR_{min}$ remains empty.

FIGURE 13. Number of patterns stored by each set of databases. Note that in the case of LD1, any value reaches the quality threshold and thus, $DBR_{min}$ remains empty.

TABLE 5. Number of patterns created from original simulation results ($DB_{Manhattan}$) and their depurated counterparts ($DB_{Rmin}$).

| Pattern | $DB_{Manhattan}$ | $DB_{Rmin}$ | Reduction (%) |
|---------|------------------|--------------|---------------|
| 2x2 Pattern | 8,499,957 | 5,787,176 | 31.92 |
| 3x3 Pattern | 6,584,238 | 3,986,426 | 39.46 |
| 4x4 Pattern | 5,220,039 | 3,016,904 | 42.21 |

D. DATABASE SIZE COMPARISON - ORIGINAL VS. DEPURATED

Following the procedure described in [24], recommender/CF databases, containing squared 2D patterns of several sizes $q \times q$ have been created from the simulations (cf., Section III.A). We have created two sets of databases: One set ($DB_{Manhattan}$) obtained using the original simulation results, and another set ($DB_{Rmin}$) obtained using their depurated versions resulting from the application of the procedure described in Section III.B. That is, discarding all the patterns, which have one or more cells below the $R_{min}$ incident rays threshold. In each set, we distinguish 21 different databases depending on the size of the 2D pattern (i.e., $2 \times 2, 3 \times 3$ and $4 \times 4$) and the NR of the LD simulations (i.e., $LD1(NR=0), LD2(NR=1), LD3(NR=2), LD4(NR=3), LD5(NR=4), LD6(NR=5)$ and $LD7(NR=6)$). Hence,

$$DB_{Manhattan} = \{DB_{Rmin}^{LD_i}, \forall q \in [2, 4] \land \forall i \in [1, 7]\},$$

and

$$DB_{Rmin} = \{DB_{Rmin}^{LD_i}, \forall q \in [2, 4] \land \forall i \in [1, 7]\}.$$  

Fig. 13 shows a comparison between these two sets regarding the number of generated patterns. Table 5 shows the aggregated results. The proposed depuration procedure reduces the size of databases (i.e., number of stored patterns) by more than 30% in all cases.

V. EXPERIMENTAL RESULTS OF THE PROPOSED ALGORITHM

The proposed depuration algorithm could be applied in two different phases: (Phase I) During the creation of knowledge databases to reduce the number of patterns and filter inaccurate values from LD simulations, and (Phase II) Before the prediction on the SUT simulation to filter inaccurate values of the SUT LD simulation before applying the prediction algorithm. Depending on the phase(s) in which the depuration algorithm is applied, we can distinguish four cases:

- Case A: No depuration is applied (original approach [23])
- Case B: Depuration is applied only in Phase I
- Case C: Depuration is applied only in Phase II
- Case D: Depuration is applied in both phases

Our depuration algorithm reduces the number of patterns in the databases, contributing to a better efficiency of the overall solution (cf., Section III.D). However, beyond efficiency improvements, in this section, we assess the effect of the depuration algorithm on the prediction accuracy. To do so, we apply the 2D CF-RL hybrid method proposed in [24] with several prediction strategies (cf., Table 6) on the original simulation of the SUT and the depurated simulation of the SUT.

Prediction strategies differ on the size of the 2D pattern (i.e., $q \in [2, 4]$) and on the database source (i.e., original simulation results − $DB_{Manhattan}$, or their depurated counterparts − $DB_{Rmin}$).

Each strategy is applied to the existing simulations obtained with different NR (i.e., from LD2 to LD7). LD1 (i.e., $NR=0$) is discarded because $DB_{Rmin}^{LD1}$ is empty for any $q \in [2, 4]$, (cf., Fig. 13). In each strategy, we set $k=100$ as the maximum number of neighbours used to compute the prediction. Results accuracy is measured in terms of the MAE between our predictions and the HD simulation values. All results are shown in Table 7 highlighted in different colours depending on the case. The average MAE for each case is given in Table 8.

It can be observed that using our depuration algorithm helps to reduce the error. The original approach, Case A, [24] is the one with the worst performance, whilst Case D is the best performer. However, the differences are not substantial. Comparing cases A and B, and cases C and D we observe that the effect of using the original simulations ($DB_{Manhattan}$) or the depurated ones ($DB_{Rmin}$) is minimal, slightly favoring the use of the depurated ones (i.e., $-0.02\ dB$). Similarly,
TABLE 7. Mean absolute error LD + CF vs. HD simulations (in dB – the lower the better). σ indicates the standard deviation Case A: Original databases & Original SUT (Light Red), Case B: Depurated databases & Original SUT (Light Yellow), Case C: Original databases & Depurated SUT (Light Blue), Case D: Depurated databases & Depurated SUT (Light Green).

| Case | LD2 | LD3 | LD4 | LD5 | LD6 | LD7 | Strategy Average | σ |
|------|-----|-----|-----|-----|-----|-----|-----------------|---|
| A    | 11.57 | 8.99 | 10.83 | 12.38 | 10.98 | 12.89 | 10.67 | 13.68 | 10.38 | 14.80 | 10.56 | 12.95 |
| B    | 10.98 | 8.78 | 11.24 | 9.56 | 11.63 | 10.08 | 12.55 | 13.65 | 10.08 | 14.92 | 10.31 | 12.50 |
| C    | 10.51 | 8.65 | 10.52 | 9.08 | 11.29 | 9.57 | 12.44 | 9.79 | 13.66 | 9.94 | 14.82 | 10.09 | 12.21 |
| D    | 12.39 | 9.43 | 13.67 | 10.56 | 12.96 | 11.47 | 12.82 | 10.57 | 13.18 | 10.18 | 14.53 | 10.50 | 13.16 |
| 2D.1 | 11.44 | 9.04 | 11.61 | 9.69 | 12.31 | 10.70 | 12.32 | 12.80 | 10.63 | 14.13 | 10.14 | 12.44 |
| 2D.2 | 10.71 | 8.80 | 10.68 | 9.15 | 11.63 | 9.97 | 12.21 | 9.65 | 12.75 | 9.82 | 14.01 | 10.01 | 12.00 |

TABLE 8. Average MAE for each Case (in Db – The lower the better).

| Case | Average MAE |
|------|-------------|
| A    | 12.55       |
| B    | 12.53       |
| C    | 11.63       |
| D    | 11.61       |

TABLE 9. Sparseness values of sut before/after depuration (in %).

| SUT       | LD2 | LD3 | LD4 | LD5 | LD6 | LD7 |
|-----------|-----|-----|-----|-----|-----|-----|
| Original  | 24.09 | 11.73 | 5.48 | 3.12 | 1.89 | 1.25 |
| Depurated | 71.89 | 43.66 | 26.06 | 16.38 | 10.87 | 6.99 |

comparing cases A and C, and cases B and D we can see the effect of using the original simulations of the SUT or its depurated counterpart. In this case, using the depurated SUT simulations is clearly better (i.e., −0.98 dB). It is worth noting that applying our depuration algorithm to the SUT increases its sparseness significantly (cf., Table 9). However, despite this apparent loss of information, the results exhibit a lower error, which supports the usefulness of our proposal not only in terms of efficiency but also in terms of error reduction. Regarding strategies, there is no clear evidence supporting the use of one or another, although there is a slight trend towards reducing the error whilst increasing the pattern size. Surprisingly enough, using simulations with a small NR (e.g., LD2, LD3) proves to perform better than those obtained with larger NR (e.g., LD7). This occurs because the higher the NR, the more specific and characterised the scenario is, and finding similar patterns becomes more difficult. Therefore, the obtained results indicate that the proposed methodology effectively maintains power level estimation accuracy, with a reduction in database size. Therefore, computational complexity is reduced, enabling an increase in the computational volume of the scenario under test.

VI. CONCLUSION

We have proposed an optimization of our previous 2-dimensional RL–CF approach, based on the analysis of the NR parameter. We presented a methodology to obtain the $R_{\text{min}}$ value, which is computed finding a trade-off between sparseness and the error between LD and HD values. The $R_{\text{min}}$ value enables the depuration of simulations by removing invalid values, which reduces DBs size and increases efficiency and accuracy. While the previous approach applied CF to the empty cells of LD of 3D RL simulations, this new method discards the results of the cells reached by a number of rays lower than a set value (NR). This discarded cell values are considered noise due to the high error that they provide comparing them to high definition ray launching results. Thus, applying the collaborative filtering technique both to empty and noisy cells, the overall accuracy of the proposed methodology is significantly improved. The main contributions of this article are summarised as follows: (i) we studied the NR parameter to observe its effect on LD simulations (ii) we computed the $R_{\text{min}}$ depuration value according to sparseness and error values obtained for each simulated scenario considering different NRs (iii) we showed that our optimized, depuration-based proposal obtains faster and more accurate results when applied to both databases and the SUT. Future work will focus on the clustering of scenarios according to their features and characteristics. Therefore, scenarios with similar contexts will be grouped, enhancing the results, especially in simulations with high NR. The proposed methodology enables performing coverage/capacity analysis whilst reducing the computational cost of wireless systems in scenarios with high node density, morphological complexity and size. Note that the methodology can be applied to every kind of scenario building and feeding the databases accordingly. The proposed methodology deals with frequency dependent power level characterization. Great deal of interest lays within the analysis of time delay characteristics and their effect in overall system performance [45]. Application of hybrid 3D RL + CF techniques within time domain parameters will also be explored as a future work line.

REFERENCES

[1] S. Mattison, “An overview of 5G requirements and future wireless networks: Accommodating scaling technology,” IEEE Solid State Circuits Mag., vol. 10, no. 3, pp. 54–60, 2018.
B. Yang, X. Cao, Z. Han, and L. Qian, “A machine learning enabled MAC,” IEEE Access, vol. 3, pp. 1089–1100, 2015.

R. Brem and T. F. Eibert, “A shooting and bouncing ray (SBR) modeling framework including dielectrics and perfect conductors,” IEEE Trans. Antennas Propag., vol. 63, no. 8, pp. 3599–3609, Aug. 2015.

A. Massa, D. Marcantonio, X. Chen, M. Li, and M. Salucci, “DNNs as applied to electromagnetics, antennas, and Propagation—A review,” IEEE Antennas Wireless Propag. Lett., vol. 18, no. 11, pp. 2225–2229, Nov. 2019.

M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, “Artificial neural networks-based machine learning for wireless networks: A tutorial,” IEEE Commun. Surveys Tuts., vol. 21, no. 4, pp. 3039–3071, 4th 2019.

H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, “Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system,” IEEE Trans. Veh. Technol., vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

D. Praveen Kumar, T. Ang moth, and C. S. R. Annavarapu, “Machine learning algorithms for wireless sensor networks: A survey,” Inf. Fusion, vol. 49, pp. 1–25, Sep. 2019.

Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, “Path loss prediction based on machine learning: Principle, method, and data expansion,” Appl. Sci., vol. 9, no. 9, p. 1908, May 2019.

C. Orosa, Z. Zhang, T. Watteyne, S. D. Glaser, “A machine-learning-based connectivity model for complex terrain large-scale low-power wireless deployments,” IEEE Trans. Cogn. Comput. Netw., vol. 3, no. 4, pp. 576–584, Dec. 2017.

F. A. Almalki and M. C. Angelides, “A machine learning approach to evolving an optimal propagation model for last mile connectivity using low altitude platforms,” Comput. Commun., vol. 142–143, pp. 9–33, Jun. 2019.

I. R. Gomes, C. R. Gomes, H. S. Gomes, and G. P. D. S. Cavalcante, “Empirical radio propagation model for DTV applied to non-homogeneous paths and different climates using machine learning techniques,” PLoS ONE, vol. 13, no. 3, 2018, Art. no. e0194511.

S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, “Quantum machine learning for 5G communication networks: State-of-the-art and vision for the future,” IEEE Access, vol. 7, pp. 46317–46350, 2019.

W. Lee, “Resource allocation for multi-channel underlay cognitive radio network based on deep neural network,” IEEE Commun. Lett., vol. 22, no. 9, pp. 1942–1945, Sep. 2018.

M. Eisen, C. Zhang, L. F. O. Chamon, D. D. Lee, and A. Ribeiro, “Learning optimal resource allocations in wireless systems,” IEEE Trans. Signal Process., vol. 67, no. 10, pp. 2775–2790, May 2019.

B. Yang, X. Cao, Z. Han, and L. Qian, “A machine learning enabled MAC framework for heterogeneous Internet-of-Things networks,” IEEE Trans. Wireless Commun., vol. 18, no. 7, pp. 3697–3712, Jul. 2019.

S. Hu, Y. Fei, P. P. Liang, and Y.-C. Liang, “Deep neural network for robust modulation classification under uncertain noise conditions,” IEEE Trans. Veh. Technol., vol. 69, no. 1, pp. 564–577, Jan. 2020.

Y. Wang, Z. Li, J. Gao, and L. Zhao, “Deep neural network-based WiFi/pedestrian dead reckoning indoor positioning system using adaptive robust factor graph model,” IET Radar, Sonar Navigat., vol. 14, no. 1, pp. 36–47, Jan. 2020.

Y. Wei, F. R. Yu, M. Song, and Z. Han, “Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor–critic deep reinforcement learning,” IEEE Internet Things J., vol. 6, no. 2, pp. 2061–2073, Apr. 2019.

S. Khan, H. A. Khattak, A. Almogren, M. A. Shah, I. Ud Din, I. Alkalifa, and M. Guizani, “5G vehicular network resource management for improving radio access through machine learning,” IEEE Access, vol. 8, pp. 6792–6800, 2020.

L. Azpilicueta, M. Rawat, K. Rawat, F. M. Ghannouchi, and F. Falcone, “A ray launching-neural network approach for radio wave propagation analysis in complex indoor environments,” IEEE Trans. Antennas Propag., vol. 62, no. 5, pp. 2777–2786, May 2014.

L. Azpilicueta, F. Falcone, and R. Janaswamy, “A hybrid ray launching-diffusion equation approach for propagation prediction in complex indoor environments,” IEEE Antennas Wireless Propag. Lett., vol. 16, pp. 214–217, 2017.
FRAN CASINO (Member, IEEE) received the B.Sc. degree in computer science, the M.Sc. degree in computer security and intelligent systems, and the Ph.D. degree (Hons.) (cum laude) in computer science from Rovira i Virgili University, Tarragona, Spain, in 2010, 2013 and 2017, respectively. He was a Visiting Researcher in ISCTE-IUL (Lisbon-2016). He is currently a Post-doctoral Researcher with the Department of Informatics, University of Piraeus, Piraeus, Greece. He has participated in several European-, Spanish-, and Catalan-funded research projects. His research focuses on pattern recognition and data management applied to different fields such as privacy and security protection, recommender systems, smart health and blockchain. He received the Best Dissertation Award for his Ph.D. degree.

LEIRE AZPILICUETA (Senior Member, IEEE) received the degree in telecommunications engineering, in 2009, and the master’s degree in communications and the Ph.D. degree in telecommunication technologies from the Public University of Navarre (UPNa), Spain, in 2011 and 2015, respectively. In 2010, she worked in the Research and Development Department of RFID Osés as a Radio Engineer. She is currently working as an Associate Professor and a Researcher with Tecnológico de Monterrey, Monterrey, Mexico. She has over 150 contributions in relevant journals and conference publications. She was a recipient of the IEEE Antennas and Propagation Society Doctoral Research Award 2014, the Young Professors and Researchers Santander Universities 2014 Mobility Award, the ECSA 2014 Best Paper Award, the IISA 2015 Best Paper Award, the Best Ph.D., in 2016, Awarded by the Colegio Oficial de Ingenieros de Telecomunicación, the N2Women: Rising Stars in Computer Networking and Communications 2018 Award, and the ISSI 2019 Best Paper Award.

FRANCISCO FALCONE (Senior Member, IEEE) received the degree in telecommunication engineering and the Ph.D. degree in communications engineering from the Universidad Pública de Navarra (UPNA), Spain, in 1999 and 2005, respectively. From February 1999 to April 2000, he was a Microwave Commissioning Engineer at Siemens-Italtel, deploying microwave access systems. From May 2000 to December 2008, he was a Radio Access Engineer at Telefónica Móviles, performing radio network planning and optimization tasks in mobile network deployment. In January 2009, as a co-founder, he has been the Director of Tafco Metawireless, a spin-off company from UPNA, until May 2009. In parallel, he was an Assistant Lecturer with the Electrical and Electronic Engineering Department, UPNA, from February 2003 to May 2009. In June 2009, he becomes an Associate Professor with the EE Department, being the Department Head, from January 2012 to July 2018. From January 2018 to May 2018, he was a Visiting Professor with the Kuwait College of Science and Technology, Kuwait. He is also affiliated with the Institute for Smart Cities (ISC), UPNA, which hosts around 140 researchers. He is currently acting as the Head of the ICT Section. His research interests are related to computational electromagnetics applied to the analysis of complex electromagnetic scenarios, with a focus on the analysis, design, and implementation of heterogeneous wireless networks in plug-and-play environments. He has over 500 contributions in indexed international journals, book chapters, and conference contributions. He received the CST 2003 and CST 2005 Best Paper Award; the Ph.D. Award from the Colegio Oficial de Ingenieros de Telecomunicación (COIT), in 2006; the Doctoral Award UPNA, in 2010; the First Juan Gomez Peñalver Research Award from the Royal Academy of Engineering of Spain, in 2010; the XII Talgo Innovation Award 2012; the IEEE 2014 Best Paper Award, in 2014; the ECSA-3 Best Paper Award, in 2016; and the ECSA-4 Best Paper Award, in 2017.

AGUSTÍ SOLANAS (Senior Member, IEEE) received the M.Sc. degree (Hons.) in computer engineering from Rovira i Virgili University (URV), in 2004, the Diploma degree in advanced studies from the Technical University of Catalonia, in 2005, and the Ph.D. degree from the Department of Telematics Engineering, Technical University of Catalonia, in 2007. He is currently a Professor with the Department of Computer Engineering and Mathematics and the Head of the Smart Technologies Research Group, URV. His current research interests include smart technologies, health informatics, behaviour analysis, multivariate analysis, privacy protection, and computer security. He serves as a Scientific Coordinator at APWG.EU.

PEIO LOPEZ-ITURRI (Member, IEEE) received the degree in telecommunications engineering, the master’s degree in communications, and the Ph.D. degree in communication engineering from the Public University of Navarre (UPNA), Pamplona, in 2011, 2012, and 2017, respectively. He has worked in ten different public and privately funded research projects. He has over 120 contributions in indexed international journals, book chapters, and conference contributions. He is affiliated with the Institute for Smart Cities (ISC), UPNA. His research interests include radio propagation, wireless sensor networks, electromagnet dosimetry, modeling of radio interference sources, mobile radio systems, wireless power transfer, the Internet of Things (IoT) networks and devices, 5G communication systems, and EMI/EMC. He received the ECSA 2014 Best Paper Award and the IISA 2015 Best Paper Award. He also received the 2018 Best Spanish Ph.D. thesis in Smart Cities in CAEPIA 2018 (Third Prize), sponsored by the Spanish Network on Research for Smart Cities CI-RTI and Sensors.

ERIK AGUIRRE received the M.Sc. degree in telecommunications engineering from the Public University of Navarre (UPNA), in 2012, and the Ph.D. degree in 2014. From 2012 to 2014, he worked in a research project at the University of Vigo, related to dispersive propagation. Since 2015, he has been working in Tafco Metawireless. Since 2016, he has been an Assistant Lecturer with UPNA. His research interests include radio propagation in dispersive media, body centric communications, and wireless sensor networks.

[44] L. Azpilicueta, M. Rawat, K. Rawat, F. Ghannouchi, and F. Falcone, “Convergence analysis in deterministic 3D ray launching radio channel estimation in complex environments,” Appl. Comput. Electromagn. Soc. J., vol. 29, no. 4, pp. 256–271, 2014.

[45] B. Wang, W. Chen, B. Zhang, and Y. Zhao, “Regulation cooperative control for heterogeneous uncertain chaotic systems with time delay: A synchronization errors estimation framework,” Automatica, vol. 108, Oct. 2019, Art. no. 108486.

[46] B. Wang, W. Chen, B. Zhang, and Y. Zhao, “Regulation cooperative control for heterogeneous uncertain chaotic systems with time delay: A synchronization errors estimation framework,” Automatica, vol. 108, Oct. 2019, Art. no. 108486.