

Understanding sub-GHz Signal Behaviour in Deep-Indoor Scenarios

Malarski, Krzysztof Mateusz; Thrane, Jakob; Christiansen, Henrik Lehrmann; Ruepp, Sarah Renée

Published in: leee Internet of Things Journal

Link to article, DOI: 10.1109/JIOT.2020.3027829

Publication date: 2021

Document Version Peer reviewed version

Link back to DTU Orbit

Citation (APA): Malarski, K. M., Thrane, J., Christiansen, H. L., & Ruepp, S. R. (2021). Understanding sub-GHz Signal Behaviour in Deep-Indoor Scenarios. *Ieee Internet of Things Journal, 8*(8), 6746-6756. https://doi.org/10.1109/JIOT.2020.3027829

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.

- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Understanding sub-GHz Signal Behaviour in Deep-Indoor Scenarios

Krzysztof Mateusz Malarski, Student Member, IEEE, Jakob Thrane, Student Member, IEEE, Henrik Lehrmann Christiansen, Member, IEEE, and Sarah Ruepp, Member, IEEE

Abstract—Critical Internet of Things (IoT) services require seamless connectivity, which is not always simple to provide and particularly in deep-indoor scenarios, it can be even impossible in some cases. The existing outdoor-to-indoor path loss models lack the accuracy in the underground situations, thus IoT coverage planning in such areas cannot rely on robust tools and becomes a process of trial and error. In this work, we derive and analyse various environmental features that can be useful in understanding sub-GHz deep-indoor signal propagation. Based on a large-scale field trial in an underground tunnel system. we formulate several parameters related to TX-RX distance and tunnel geometry. Through feature relevance studies in linear (Ordinary Least Squares (OLS) regression) and non-linear (Gaussian Process Regression) realms we show that 2D indoor distance and the distances to the tunnel walls may be useful in sub-GHz signal strength prediction in deep-indoor situations. We construct a linear and a Gaussian Process model for indoor pathloss prediction that outperform the 3rd Generation Partnership Project (3GPP) model by 1.8 dB and 4.1 dB, respectively.

Index Terms—sub-GHz, NB-IoT, signal propagation, deepindoor, path-loss, gaussian process regression, tunnel, LIDAR.

I. INTRODUCTION

THE vital and dynamic development of numerous Internet f Things (IoT) protocols, architectures and technologies has transfigured 5G visions about connected industry and society into a real and already ongoing process. Already in 2019, it was forecasted that 93% of companies use IoT to optimise and automate their processes [1]. However, it was the advent of Low-Power Wide-Area Network (LP-WAN) standards that has really enabled the services involving big distances (more than 1 km) and challenging connectivity conditions, such as smart metering and object monitoring. A prominent LP-WAN member, Narrowband IoT (NB-IoT), provides 20 dB link budget improvement with respect to Long-Term Evolution (LTE) [2]. Considering the specification alone, it may appear clear that NB-IoT is the right choice for such use-cases as water metering, asset tracking or intelligent alarm systems. However, as these services may require the sensors to operate in so called, deep-indoor scenario¹ one must be

Copyright (c) 20xx IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

¹We classify a communication scenario deep-indoor when there are at least 2 layers of walls or ground between the communicating parties, e.g. underground tunnel, underground parking or metro station level -2.

aware of NB-IoT behaviour in such extreme situation before planning the actual service deployment. In deep-indoor areas, it is inevitable to encounter coverage holes, i.e. spots when the User Equipment (UE) cannot operate with at least 160 bps throughput in uplink and/or downlink [3]. Yet, even in the case of full coverage, the energy consumption of a NB-IoT system may grow undesirably in difficult radio conditions [4], as the NB-IoT system employs message repetition scheme to increase the coverage and the power consumption of the UE is proportional to the number of repetitions [5].

From the service provider and user perspectives, it is crucial that the packets are reliably transmitted and the devices are deployed optimally, so that infrastructure investments and the cost of replacing the batteries are minimised. Although there are many studies based on simulations and measurements proving good NB-IoT coverage in outdoor and indoor settings [6]-[8], less efforts have been done to study the deepindoor coverage, and according to the existing findings, it is significantly poorer than in outdoor and indoor scenarios [9]. Moreover, the simplicity of the official deep-indoor pathloss models defined in [10], [11] does not reflect the true nature NB-IoT signal propagation in underground areas [12]. As a result, service providers and users interested in deepindoor deployments are left with neither a strong experimental evidence about robust NB-IoT operation nor with precise statistical models that could be used to reliably simulate a deep-indoor NB-IoT application.

In this article we investigate NB-IoT signal behaviour in the underground tunnel and explore the potential of several environmental features for explaining sub-GHz signal attenuation in deep-indoor scenarios. In [13] we presented the measurement procedure and several parameters that could be included in deep-indoor path-loss modelling. In this work, we introduce more features and extend the data analysis by studying feature relevance by means of Gaussian Process Regression (GPR). The contributions of this work are the following:

- we introduce deep-indoor features: number of the closest corridors, the distance and the angle to the farthest tunnel corner, and the distances to the tunnel walls and ceiling,
- we analyse statistically the derived features jointly with the dataset introduced in [13]; we apply feature selection, linear regression and Gaussian Process Regression,
- we discuss the prediction performance of the 3GPP 38.901 model and the proposed linear and Gaussian Process models based on the most relevant features; we explain to what extent the findings of this work can be

K. M. Malarski, J. Thrane, H. L. Christiansen and S. Ruepp are with the Department of Photonics Engineering, Technical University of Denmark, Kongens Lyngby, 2800 e-mail: {krmal, jathr, hlch, srru}@fotonik.dtu.dk

This work was partially supported by Innovation Fund Denmark through the Eureka Turbo project IoT Watch4Life.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3027829, IEEE Internet of Things Journal

ТΔ	BIF	Ŀ	Notation	used in	this	article
IА	DLE	11	INOLALION	used m	uns	anticle

PL_{o2i}	outdoor-to-indoor path loss
PL_b	basic outdoor path loss
PL_{tw}	building penetration path los
PL_{in}	indoor path loss
G_{TX}	transmitter gain
G_{RX}	receiver gain
L_{TX}	transmitter noise figure
L_{RX}	receiver noise figure
P_{RX}	received power
P_{TX}	transmitted power
L_{misc}	miscellaneous losses
$\sigma_n 2$	local variability term
$\dot{\omega}$	vector of weights
$oldsymbol{x}$	input vector
\boldsymbol{y}	output vector
X	training data
X_*	test data



The remainder of this paper is organised as follows. The related work is introduced in Section II. We introduce the methodology of the work in Section III. Section IV describes the dataset acquired by the underground measurement campaign and the feature engineering process. We analyse the relevance of the features from the linear perspective in Section V and in non-linear domain in Section VI. Section VII contains a discussion and additional considerations for LP-WAN deep-indoor coverage modelling, and Section VIII concludes the work. The explanation on the general symbols used in this article can be found in Table I.

II. RELATED WORK

IoT connectivity in underground areas is an active research topic due to the potential critical applications, such as: cable monitoring [14], mine safety warning [15] and urban drainage systems [16]. In their survey, Hrovat, Kandus and Javornik reviewed the state-of-the-art research focusing on radio propagation modelling in underground tunnels [17]. The discussed techniques spanned from numerical methods of Maxwell equations, through Ray Tracing, to empirical statistical modelling. The authors explain the factors affecting signal propagation in the tunnels. Tunnel geometry, especially cross section shape and the presence of big obstacles may have significant impact on the attenuation, whilst the effect of electromagnetic features of the tunnel materials can be neglected in most cases [17]. Even though LP-WAN technologies are relatively new, they have attracted attention in terms of underground applications. This is not surprising, as in the most prominent standards: Long Range Wide-Area Network (LoRaWAN), Sigfox and NB-IoT the Maximum Coupling Loss (MCL) is increased by at least 10 dB, comparing to General Packet Radio Service (GPRS) (as shown in Table II) [8].

TABLE II: MCL in uplink for LP-WAN technologies.

Standard	GPRS	LoRaWAN	Sigfox	NB-IoT
MCL (dB)	144	154	158	164

A simulation-based study in [18] considers applying Lo-RaWAN for underground sensor networks for agriculture. It is



Fig. 1: Deep-indoor path-loss component derivation procedure.

shown that LoRaWAN sensors can achieve a multi-kilometre range, if buried less than 70 cm under the ground; moreover, the moisture level of the soil has substantial influence on the signal attenuation. In [19], the authors claim that both LoRaWAN and NB-IoT are suitable for underground deployments of up to 1 metre of depth, based on Ray Tracing simulations and single-point measurement in the manhole. The measurements of 2 network operators in Oslo showed that NB-IoT provides better deep-indoor coverage than LTE [20], however, a deep-indoor measurement campaign presented in [9] implies that NB-IoT can only provide underground connectivity up to 400 metres away from the evolved Node-B (eNB). Based on the ray tracing study, the authors in [21] state that unless novel optimisation techniques (such as multi-hop relaying) are implemented, the current infrastructure cannot provide reliable coverage in deep indoor scenarios. The work in [13] presents the first large-scale NB-IoT underground measurement campaign (895 points) and initially evaluates indoor features engineered from measurements' metadata in terms of path loss prediction.

The area of NB-IoT deep-indoor coverage modelling remains unexplored. To the best of the authors' knowledge, this article presents the first empirical study of Outdoor-To-Deep-Indoor (O2DI) for NB-IoT in a deep indoor propagation scenario.

III. METHODOLOGY

In the IoT deployment process, it is essential to use accurate and versatile modelling tools able to predict the coverage condition in various environments, and at a reasonable cost. Ideally, the model should be based on generic (i.e., scenarioagnostic) parameters that are cheap to derive. Therefore, considering deep-indoor scenarios, it is understandable to discard Ray Tracing modelling, as the tool requires geographical data, which may be unavailable or difficult to acquire. Moreover, the models are complex, thus costly in terms of computations and software price. Instead, the use of models formulated with the aid of statistical data seems appealing. As such models often describe the phenomena by means of generalised mathematical equations, they are simple to implement and consume little computational resources. At the same time, the desired quantities depend on generic parameters (such as distance and power) that can be calculated or measured in situ. There are, however, certain issues to be addressed for the statistical models to be accurate:

- Complexity level. By adjusting the number of terms and the corresponding coefficients, one needs to face a tradeoff between model simplicity and accuracy.
- 2) Parameter calibration with the experimental data. The more unbiased measurements are provided at the model creation, the more versatile the model becomes. Unfortunately, conducting deep-indoor measurement campaigns is a challenging and costly task.

Let us consider the example of Outdoor-To-Indoor (O2I) path-loss model defined in [10]:

$$PL_{o2i} = PL_b + PL_{tw} + PL_{in} + \mathcal{N}(0, \sigma_n^2)$$
(1)

Where, PL_b is the *basic* outdoor path-loss, PL_{tw} are constant and frequency dependent building penetration losses, $\mathcal{N}(0, \sigma_p^2)$ is a log-normal distribution with local variability σ_p and PL_{in} are losses dependent on the depth inside the building. The component related to indoor losses is formulated as follows:

$$PL_{in} = 0.5 \cdot d_{in,2d} \tag{2}$$

Where $d_{in,2d}$ is the distance indoor, e.g. the distance to the outermost wall closest to the transmitter. Noteworthy, the model is only limited to regular buildings and does not consider underground cases (level -1, -2, etc.). It means that the PL_{in} of the considered path-loss model has not been calibrated with empirical deep-indoor data, thus does not correspond to the complexity of underground signal propagation, as shown in [12]. Nevertheless, even though there exist other outdoor and indoor models that consider environmental particularities, such as COST 231 [22], we decided to utilise the 3GPP 38.901 model, presented in this section. We find it advantageous that 3GPP models are general and applicable to wide spectrum of outdoor and indoor scenarios; therefore, the purpose of our work is to 1) evaluate the 3GPP 38.901 model performance underground and 2) propose an enhancement based on the additional deep-indoor measurements. In this paper, we propose a novel PL_{in} formula following the procedure in Figure 1. We base on the measurements taken in an underground tunnel system and we engineer features with the aid of the available Light Detection and Ranging (LIDAR) data of the measured area. We perform Ordinary Least Squares (OLS) regression analysis of statistical significance and relevance analysis with Gaussian Process Regression (GPR).

A. Linear regression

Linear regression is a simple, well-know and easy to interpret method of statistical inference in the cases where the underlying dependence resemble linear behaviour [23]. In this type of regression, the output variable y is modelled as a linear dependency of the input features x, and the fit of the model can be adjusted by optimised the assigned weights ω . In a M feature situation, linear regression may be represented as follows:

$$\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\omega}) = \omega_0 + \omega_1 x_1 + \dots + \omega_M x_M$$
 (3)

One of the most straightforward methods of linear model optimisation is adjusting the weights so that the deviations between the observed and predicted output values are minimised. In Ordinary Least Squares method, the term subjected to minimisation (Sum of Squares Error) is given as follows:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$
(4)

where *n* is the number of output samples, y_i is the actual observation and \hat{y}_i is the predicted output value [24].

In our work, we investigated whether it is feasible to propose a more accurate PL_{in} term of the Equation 1 with the aid of the engineered features. We applied linear regression using various combinations of the features as parameters (vector **x** in Equation 3) and Reference Signal Received Power (RSRP) of the signal as the output variable. Finally, we constructed a linear regression model for PL_{in} prediction, which was compared with the baseline model (Equation 2) and the GPR model.We utilised a Python implementation of OLS regression [25] and examined the goodness of fit coefficient (R^2) and Residual Mean Square Error (RMSE) of the considered models.

B. Gaussian processes

Gaussian Process is a probabilistic machine learning method, applicable both to regression and classification problems. It allows for a tractable inference for (non-)linear output functions and an arbitrary number of dimensions, under the assumption that both the input and the output follow gaussian distribution. In the regression, the resultant process (i.e. functional distribution) can be sampled to yield possible x,y dependencies; additionally it is possible to evaluate the significance of such samples-functions by analysing the given probabilities.

The GPR bases on 2 fundamental properties of gaussian distributions, namely that a gaussian distribution retains its characteristics even when subjected to: 1) Marginalisation, when certain dimensions are discarded (marginalised out) and 2) Conditioning under the training data [26].

Contrary to linear regression, GPR does not use any weightlike parameters to produce a function, but instead uses the input dataset (say, M samples) to create M-dimensional distribution described by its mean \bar{f}_* and covariance matrix $cov(f_*)$. The covariance matrix can be derived with the aid of the selected kernel function, which provides the similarity measure between any two input points. A well-known kernel example, also applied in this work, is squared-exponential kernel, defined as follows:

$$cov(f(\boldsymbol{x}), f(\boldsymbol{x}_*)) = K(X, X_*) = \exp\left(-\frac{1}{2}|\boldsymbol{x} - \boldsymbol{x}_*|^2\right)$$
 (5)

The ultimate goal of GPR is to derive a predictive distribution given the training and test data:

^{2327-4662 (}c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 13,2020 at 08:37:05 UTC from IEEE Xplore. Restrictions apply.



 $PL_{in} = P_{RX} - P_{TX} - G_{TX} - G_{RX} - L_{TX} - L_{RX} - L_{misc} - PL_{o2i} - PL_b$

Fig. 2: PL_{in} calculation.

$$\boldsymbol{f_*}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X_*} \sim N(\bar{\boldsymbol{f_*}}, cov(\boldsymbol{f_*})) \tag{6}$$

where the mean is expressed as:

$$\bar{f}_{*} = K(X_{*}, X)[K(X, X) + \sigma_{n}^{2}I]^{-1}y$$
 (7)

and the covariance function:

$$cov(f_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)$$
(8)

The samples drawn from the distribution from Equation 6 are different output functions, and the mean of such samples represents the prediction of the output variable given the test inputs. Additionally, since the hyperparameters of the covariance function, called length-scales, determine the distance in the input space at which the function values become uncorrelated, it is possible to perform so called, Automatic Relevance Determination (ARD), simply based on the fact that the larger the value of the length-scale, the lower is the relevance of the parameter [27].

In this article, we applied the GPR as follows. In order to complement the linear regression findings on the relevance of the engineered features, we performed the GPR with ARD and obtained relevance scores in the non-linear realm. We chose the most relevant features and constructed the GPR model for PL_{in} prediction, compared with the model from Equation 2 and the linear regression model.In our analysis, we applied a GPR implementation in Python [28].

C. Indoor path-loss estimation

In scenarios, such as the O2DI scenario investigated in this paper, the link budget is complex and composed of many factors, related to transmitter and receiver's gains, noise figures, antenna radiation patterns, antenna polarisation, and others. In our study we did not know all of the aforementioned terms, thus we formulated a set of assumptions related to some of the components (TX gain and losses) and calculated the link budget accordingly. We calibrated the loss terms by adjusting an additional constant term corresponding to miscellaneous losses and also compensating for inaccuracies in the assumptions. The constant could be found by minimising the Mean Square Error (MSE) of the measurements towards a link budget where the constant composes the unknown attenuation and gain terms. Figure 2 shows how the indoor path loss was derived from the power measurements; TX/GX gains are denoted G_{TX} , G_{RX} , noise figures are represented by L_{TX} , L_{RX} and L_{misc} refers to the miscellaneous losses and is our link budget calibration term.

IV. EXPERIMENTAL DATA COLLECTION



Fig. 3: Layout of tunnel system with coloured signal power samples along the corridors where the measurements were conducted. The tunnel system is considered between level -1 and -2. The base station is placed at 30 m above top ground level.

In order to increase the accuracy of the statistical path-loss modelling, a field trial was conducted to collect RSRP samples



Fig. 4: The profile of the campus area.

from a NB-IoT device located in the underground tunnel system of DTU Lyngby Campus. The data were captured along 1.6 km in total; the area of the experiment is depicted in Figure 3 and the parameters of the campaign are present in Table III. The university buildings are of 2-4 floors of height and their walls are made from brick or glass/steel structure. The tunnels are located directly under the buildings and under the parking area, where high vegetation is also present. A simplified terrain cross-section is depicted in Figure 4. The signal samples were taken only along the main corridors, i.e. those connecting the group of buildings, or campus quadrants. Each measurement series started and finished in a certain distance from the entry of the tunnel system, as there was no direct access to a main corridor from above and one had to pass some smaller corridors to reach the main tunnels. Thus, the impact of the tunnel entries was not analysed.

The uniqueness of this trial lays in the fact that precise locations of the measurement points could be obtained, even though there was no Global Navigation Satellite System (GNSS) signal in the tunnels. The indoor localisation was conducted basing on high resolution LIDAR plan of the measured area. Thanks to the dataset containing x, y, z coordinates of the complete tunnel system, we could localise the measurement positions with a 1 cm precision with the aid of the following process. 1) We documented the exact location of the first and the last measurement point of each independent study (i.e. tunnel interval); 2) such start and end positions were then localised in the LIDAR point cloud and their Global Positioning System (GPS) coordinates were calculated; 3) since all the measurement points were equidistant (1 or 2 metres of distance), knowing the amount of points and the exact start/end positions it was possible to interpolate the locations of all intermediate measurement points. In total, 1048 measurement points were taken and each of the 1048 samples was averaged over 10 measurements to limit the impact of shadowing and large-scale fading impairments. In Figure 5, the captured RSRP samples are presented with respect to 3D distance and 3GPP 38.901 UMa model fit.

A. Feature engineering

As both the GPS position and transmission parameters of the neighbouring NB-IoT base station were known (see Table III and Figure 3), combining this knowledge with the precise GPS locations of the deep-indoor measurement points enabled us to derive a number of advanced features analysed in this article. Apart from calculating 3D distance between the NB-IoT device and the eNB (denoted as d_{3D}), it was also possible to compute azimuth and elevation angles (θ and ϕ , respectively). The latter ones alone were treated as quantities assisting in the inference rather than path-loss feature candidates. With the aid of 3D trigonometry, indoor distance $d_{in,2D}$ and penetration distance $d_{pen,3D}$ could be found. $d_{in,2D}$ corresponds to the distance between the UE and the tunnel edge, and $d_{pen,3D}$ represents the distance between the tunnel edge and the ground surface; both of them are calculated along the straight line path toward the eNB. The rest of the features can be categorised as tunnel-related, as they only refer to the tunnel geometry and not to the distance from the eNB. It has to be mentioned that for each tunnel, along which the measurements were taken, the locations of the entrances to the corridors crossing the main tunnel were identified. This allowed for the formulation of the average distance to the nearest corridor $d_{cor,ava}$, formulated as follows:

$$d_{cor,avg} = \frac{d_{cor,closest} + d_{cor,farthest}}{2} \tag{9}$$

where $d_{cor,closest}$ and $d_{cor,farthest}$ are the distances between the measurement point and the closest/farthest corridor entrance in the measured tunnel, respectively. Another feature included in the study is the number of close corridors² n_c . Furthermore, we defined 2 more features related to the tunnel corners, namely, the distance to the farthest corridor corner and the angle between the measurement point and the farthest corner with respect to the tunnel wall (d_{corner} and a_{corner} , respectively). Last, but not least, we derived 3 parameters involving the position of the receiver with respect to the tunnel walls along x,y,z coordinates ($d_{wall,x}$, $d_{wall,y}$ and $d_{wall,z}$).

The features investigated in this work can thus be classified into 2 groups:

- 1) distance-related, i.e. referring to the straight line distance between the NB-IoT sensor and the eNB: d_{3D} , $d_{in,2D}$ and $d_{pen,3D}$, and
- 2) tunnel-related, corresponding to the geometry properties of the tunnel: $d_{cor,avg}$, n_c , d_{corner} , a_{corner} , $d_{wall,x}$, $d_{wall,y}$ and $d_{wall,z}$.

Figure 6 illustrates all tunnel-related features and the distancerelated features are depicted in Figure 7.

V. FEATURE ANALYSIS: LINEAR PERSPECTIVE

The purpose of the statistical analysis with linear tools was to observe and evaluate linear dependencies between the engineered features and the measured RSRP values. This approach is a natural first step of the parameter examination, as the prior assumption on the indoor losses involves a linear relation to the 2D indoor distance (see Equation 2).

We performed relevance evaluation by means of both automated and analytical feature selection, and we studied the performance of a selection of linear models composed of the engineered features.

²Particularly, n_c denotes the number of crossing corridor entrances within the threshold distance from the measurement point.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3027829, IEEE Internet of Things Journal



Fig. 5: Reference Signal Received Power (RSRP) measurements versus 3D distance, compared with 3GPP predictions.



Fig. 6: Tunnel features in an example 3D tunnel fragment. In this particular example, n_c is equal to 1, as is the number of corridor entrances within the distance of d_{thr} from the measurement point.



Fig. 7: Visualisation of distance-related features.

TABLE III: Experiment parameters

# of measurement points	1048
TSMW/UE measurements per point	1 <i>e</i> 6/10
Operating frequency	820.5 MHz
Bandwidth	180 kHz
Noise figure (TX/RX)	5 dB/3 dB
TX power	46 dBm
Receiver antenna type	Dipole-like
Receiver antenna position	Vertical
TX/RX antenna gain	5 dBi/5.8 dBi

A. Feature selection

1) Automatic methods: In this work, we applied automated feature selection using 3 different methods: 1) Feature filtering based on Pearson Correlation, 2) Recursive Feature Elimination (RFE) and 3) Lasso Regularisation. The results of the test are presented in Table IV. Depending on the specific method, different features were selected. From the results of the filtering method one can learn that most of the engineered features have little correlation with RSRP and only d_{3D} and $d_{cor,avg}$ remain important. Taking into account merely the resultant R^2 coefficient and the statistical significance of the

TABLE IV: Automatic feature selection

Method	# of features selected	Features selected
Filtering	2	$d_{3D}, d_{cor,avg}$
RFE	13	$ \begin{array}{c} d_{3D}, d_{in,2D}, d_{pen,3D}, \\ d_{cor,avg}, d_{corner}, a_{corner}, \\ cos(\phi), sin(\phi), cos(\theta), sin(\theta), \\ d_{wall,x}, d_{wall,y}, d_{wall,z} \end{array} $
Lasso	5	$d_{pen,3D}, d_{cor,avg}, d_{corner}, \ d_{wall,x}, \ d_{3D}$

features (which is the case in the RFE method), only n_c is not explanatory for RSRP in a linear fashion. The iterative method selects 5 features: $d_{pen,3D}$, $d_{cor,avg}$, d_{corner} , $d_{wall,x}$ and d_{3D} . Regardless of the automatic method used, d_{3D} and $d_{cor,avg}$ are always selected.

2) Analytical selection: We repeated the process of feature selection, now in a manual way, considering both the model performance (R^2 and RMSE) and the number of features involved. Thus, performing backward feature selection we discovered the following: 1) Starting with all features, one may drop 9 features and leave only d_{3D} , $d_{in,2D}$, $d_{pen,3D}$, $d_{cor,avg}$, and $d_{wall,x}$ at the cost of 1.3 dB increase of the RMSE; 2) Starting with all tunnel-related features, removing all except $d_{cor,avg}$, n_c , d_{corner} and $d_{wall,y}$ comes at a negligible increase of RMSE (0.004 dB).

B. Goodness-of-fit analysis

1) Single-feature models: Figure 8a summarises the RMSE scores of the engineered parameters considered individually. In accordance with the feature selection results, d_{3D} and $d_{cor,avg}$ are the most explanatory for the RSRP measured. Both the azimuth angle θ and the elevation angle ϕ represent a noticeable share of the output variance and indicate that there exist other geo-statistical features not considered in this paper, that affect the signal attenuation. Interestingly, $d_{pen,3D}$ and $d_{in,2D}$ alone do not influence the signal power at the receiver; this is noteworthy having in mind that the current 3GPP indoor path-loss model bases solely on $d_{in,2D}$ (see Equation 2). Regarding the features corresponding to the tunnel geometry, the average distance to the nearest corridor impacts the RSRP much more than the distances to the tunnel walls and corners (especially, $d_{wall,z}$).

2) *Multiple feature models:* In this part, we limited the amount of investigated models to the following set:

- *all_feats*, containing all features mentioned in this article,
- *dists_only*, including distance-related features: d_{3D} , $d_{in,2D}$ and $d_{pen,3D}$,
- *tun_angles*, involving tunnel-related features together with ϕ and θ angles (i.e. *all_feats dists_only*)
- *tun_only* with all the tunnel-related features,
- 2 models corresponding to the analytical backward feature selection starting from all features and from all the tunnel features (Section V-A2): *bs_all* and *bs_tun*, respectively.

The goodness of fit for the aforementioned models can be observed in Figure 8b. First of all, combining all the features together yields the lowest value of RMSE, reaching the value of 15.1 dB; this means that the features investigated in this paper are meaningful in general. Moreover, 3 distance-related features are more powerful in linear prediction than 7 tunnel-related features, however, the latter ones in combination with ϕ and θ angles exhibit better performance. From the linear perspective, 3 of the tunnel features have no meaning in RSRP prediction, as *tun_only* and *bs_tun* models perform almost identically.

VI. BEYOND THE LINEARITY: AUTOMATIC RELEVANCE DETERMINATION ANALYSIS

The analysis included in the previous section compared simple linear models and provided some insights about the feature-to-output correlation, as well as the extent to which the individual parameters explain the variance of the RSRP. In this section we describe the GPR analysis with ARD applied on all the engineered features with the purpose of discovering their relevance when more complex input-output relationships are considered.

1) Per-corridor test: We standardised the inputs and centred them around zero mean, as required by the ARD. We conducted a series ARD tests for squared-exponential kernel space. Since the measurements in the tunnel system were taken in 9 stages (scenarios), we modelled the Gaussian Process 9 times, each time excluding a specific tunnel interval. During each test, the model was subjected to Cross-Validation (CV) with variable number of data splits, so that each time a split consisted of no more than 10 samples. For example, in one of the scenarios, after discarding one corridor the remaining 895 samples were divided into 179 consecutive splits, 5 samples per split. At each of the CV splits, the hyperparameters of the model were adjusted by Adam optimiser, until the convergence of the loss function (marginal likelihood). Compared to Section V, we excluded d_{3D} from the analysis, as its high relevance could be clearly observed; we also disregarded the features related to ϕ and θ angles. Thus, the ARD analysis was performed on the remaining 8 features.

Table V presents the lenghtscales of the features, modelled in the squared-exponential kernel space without any prior assumptions specified³. Since the lenghtscale is reciprocal to the relevance score, each feature in the table is additionally given a relevance rank for a given scenario to facilitate the relative evaluation. It is visible that the most relevant parameter is $d_{in,2D}$, ranked first in 8 out of 9 scenarios; the least useful feature, $d_{wall,z}$ was ranked the worst in the same number of scenarios. Let us consider the four most relevant features: $d_{in,2D}$, $d_{wall,x}$, $d_{pen,3D}$ and $d_{cor,avg}$, respectively. Such results agree with the analytical backward selection model *bs_all* from Section V-A2 (excluding d_{3D}). This would imply, that these features are significant in RSRP modelling both in linear and non-linear realms. Furthermore, both distancerelated and tunnel-geometry-related parameters are necessary to properly explain the behaviour of NB-IoT signal in deepindoor settings.

³Modifying the prior distributions did not affect the results noticeably.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3027829, IEEE Internet of Things Journal



Fig. 8: Comparison of the selected linear models.

TABLE V: ARD results for diffe	rent corridor scenarios
--------------------------------	-------------------------

Scenario	d_{in}	$d_{in,2D}$ $d_{pen,3D}$		$d_{cor,avg}$ n_c		c	d_{corner}		$d_{wall,x}$		$d_{wall,y}$		d _{wall,z}		RMSE[dB]		
1	0.235	1	4.996	5	1.682	3	3.575	4	8.008	7	0.41	2	6.951	6	9.572	8	8.4
2	0.187	1	1.158	3	1.332	4	5.755	6	8.417	7	0.549	2	3.652	5	9.448	8	10.7
3	0.273	1	8.053	7	7.199	5	7.318	6	6.640	4	0.28	2	0.845	3	9.967	8	12.2
4	0.255	1	0.506	3	1.775	4	4.091	5	7.26	7	0.459	2	6.243	6	9.671	8	8.4
5	0.202	1	0.539	2	2.389	4	4.774	5	7.759	7	0.598	3	5.93	6	9.513	8	7.4
6	0.207	1	0.577	3	1.825	4	4.492	5	7.193	7	0.525	2	5.056	6	9.719	8	8.7
7	0.192	1	2.718	4	0.577	3	5.59	5	9.485	7	0.551	2	8.053	6	11.596	8	13.1
8	0.62	3	0.46	2	2.217	4	3.033	5	7.557	7	0.249	1	7.244	6	11.683	8	8.6
9	0.202	1	0.689	3	2.525	4	3.25	5	9.398	8	0.42	2	7.974	6	9.063	7	10.4
AVG RANK:		1.222		3.555		3.888		5.111		6.777		2		5.555		7.888	

TABLE VI: ARD results for 75% of the shuffled data

$d_{in,2}$	D	$d_{pen,31}$	D	$d_{cor,a}$	vg	n_c		d_{corn}	er	d_{wall}	,x	$d_{wall,y}$		$d_{wall,z}$		RMSE[dB]
0.547	3	12.468	7	2.168	4	3.906	5	8.912	6	0.369	2	0.356	1	25.852	8	7

2) Shuffled samples test: In this experiment, we randomised the combined dataset and divided it into training and test sets with size ratio of 75% and 25%, respectively. We trained the GPR model on the training set and perform ARD analysis, summarised in Table VI. One can observe that the relevance ranks assigned to the features are not the same as in the percorridor analysis, and now the 3 most relevant features are: $d_{wall,y}, d_{wall,x}$ and $d_{in,2D}$. The biggest rank change happened to $d_{pen,3D}$ that in this experiment was classified as second least relevant parameter. The discrepancies in the results between Table V and Table VI indicate that each of the measured tunnel corridors represents a complex deep-indoor scenario which environmental characteristics affect the received signal power in a peculiar way. Excluding one of the tunnels always led to different feature relevance determination, thus we relied on the case with the samples shuffled from all the corridors (Table VI) and used the same data split and the selected features for indoor path-loss component prediction.

3) Indoor path-loss component prediction comparison: In order to compare the linear regression and GPR models with the 3GPP 38.901 model, we transformed the RSRP values from the measurements into PL_{in} values as depicted in Figure 2. In our study, we followed the assumption that out of the components contributing to the received signal power at the receiver, i.e.: antenna gains G_{TX} , G_{TX} , noise figures L_{TX} , L_{RX} , path-loss PL_{o2i} and other losses L_{misc} (e.g. antenna polarisation impairments, fading margin) only PL_{o2i} is a variable, while all remaining components are considered invariant. The transmitter and receiver gains and noise figures were assumed (see Table III), and the magnitude of the L_{misc} was empirically found to be equal to 5.6 dB. Moreover, we assumed the outdoor and building penetration losses constant, leaving only the PL_{in} component variable. Consequently, we trained a linear regression model and a GPR model on thus extracted PL_{in} , so that a direct comparison with 3GPP 38.901 model (Equation 2) was possible.

The boxplot presented in Figure 9 compares the Mean Absolute Error (MAE) of the predictions. Note that both the linear regression and GPR models utilise 3 most relevant features from Table VI: $d_{wall,y}$, $d_{wall,x}$ and $d_{in,2D}$. The linear model can be represented as follows:

$$PL_{in}[dB] = 11.0773 - 0.1362 * d_{in,2D} - 6.9658 * d_{wall,x} + 4.055 * d_{wall,y}$$
(10)

With respect to the 3GPP model, the linear model exhibits error improvement of 1.8 dB, whilst the GPR model predicts

^{2327-4662 (}c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 13,2020 at 08:37:05 UTC from IEEE Xplore. Restrictions apply.



Fig. 9: MAE of the 3GPP model, the linear regression and the Gaussian Process predictions (median values inside the boxes).



Fig. 10: Deep-indoor path-loss vs. 2D indoor distance, predictions on the test data.

with 4.1 dB lower error. The non-linear model has additionally 2.3 dB lower MAE than the linear regression model. Based on the given dataset, one may observe that the underlying behaviour of the PL_{in} in the deep-indoor environment might be more complex than linear.

In Figure 10, the predictions of all the considered models are presented with respect to $d_{in,2D}$. The values of the indoor distances have been limited to 25 metres, as assumed in [10]. It can be easily noticed that the 3GPP 38.901 proposal does not offer accurate PL_{in} predictions in the considered underground scenario. The models proposed in this paper exhibit considerably higher accuracy, however, it is visible that certain groups of observations (e.g. those for $d_{in,2D}$ values between 5 and 10 metres) are not predicted correctly by any model. For $d_{in,2D} > 15m$, the non-linear model follows the PL_{in} more faithfully than the linear regression model. One can notice a large spread of measured path loss values, which may imply that certain corridor-specific propagation effects (e.g. wave guiding) were not captured in our approach.

VII. DISCUSSION

Feature analysis reveals that with the aid of the parameters considered in this work it was possible to construct statistical models predicting the measured PL_{in} more accurate than the model defined in 3GPP 38.901 standard. However, one may identify several problems making the validation of the results tricky. First and foremost, the collected dataset is subjected to bias, since all data were measured in several tunnels of the same underground system. Therefore, certain geographical factors characteristic to the measurement area, such as: the location of the eNB in relation to the buildings, the deployment of above-ground obstacles, the size and the orientation of the tunnels, as well as the influence of ventilation ducts, doors and other objects in the tunnelhave affected the observed signal power. On the other hand, the captured data were taken with changing orientation of the corridors and various terrain profile above the ground, which has made the study more versatile. Acquiring the RSRP measurements from another underground tunnels, preferably situated farther from the eNB, would be beneficial for the accuracy of the model; thus, the dataset used in this paper has been published in [29] and can be extended with more underground signal samples. However, in the new area one would need either a high-resolution LIDAR dataset or another solution enabling precise indoor localisation of the measurement points, since otherwise deriving the distance-related features $(d_{3D}, d_{in,2D})$ and $d_{pen,3D}$) would not be possible. In fact, isolating single corridors from the analysis has led to diverse feature relevance scores that were difficult to interpret; further analysis of some corridors we considered the most peculiar did not lead to a meaningful conclusion. One may conclude that the engineered features, though relevant in the majority, do not reflect the entire complexity of signal behaviour in the underground scenario, which limits the prediction accuracy of the derived models. The most prominent feature of the measurement area not considered in this study is the spatial profile of the tunnel: the distribution of obstacles, the area and cross-section of the free space in the corridor, etc. On the other hand, several features analysed in this study have proven their usefulness in NB-IoT signal prediction. Regardless of the statistical tool applied in the linear analysis, d_{3D} was selected as the most relevant parameter. On the other hand, the ARD examination favoured $d_{wall,y}$, $d_{wall,x}$ and $d_{in,2D}$. All in all, these 3 features remain the most useful in understanding the underground NB-IoT signal behaviour. In general, one can notice that knowing the distance to the base station and the placement of the IoT device in terms of the tunnel walls is essential in predicting the received signal power.

Is the study only relevant for NB-IoT deployments?: The analysis described in this article considers deep-indoor propagation of NB-IoT guard-band signal in band 20 (820.5 MHz). However, we believe that the observed signal behaviour might also be useful source of reference for underground deployments of unlicensed sub-GHz IoT technologies, such as LoRaWAN and Sigfox, as their operating frequency is also sub-GHz, 868 MHz in Europe. One has to remember the difference in consequences that signal attenuation may have for the operation of a particular IoT standard; as already mentioned in this paper, NB-IoT is sensitive to the radio conditions, as they affect the number of repetitions, thus both the data rate and the energy consumption. In the case of Sigfox, the message is always repeated up to three times regardless of the received signal power⁴, and in LoRaWAN the RSRP may influence the spreading factor in use, thus the data-rate and time-on-air of the packets [30].

VIII. CONCLUSIONS AND FUTURE WORK

Until now, the area of deep-indoor path-loss modelling in NB-IoT has been unexplored, as both empirical data and accurate models are missing. In this work, we addressed this gap by introducing, describing and analysing a number of distance- and tunnel-related features. 3 parameters: $d_{in,2D}$, $d_{wall,x}$ and $d_{wall,y}$ were distinguished as the most relevant for predicting the measured RSRP. We constructed linear regression and Gaussian Process models that predicted the signal power with the MAE decreased noticeably by 1.8 dB and 4.1 dB, respectively, compared to the 3GPP 38.901 model.

In order to increase the accuracy of the proposed indoor path-loss models, the next step would be to 1) collect more measurement data from diverse deep-indoor scenarios, preferably at various distances from the base station and 2) define more environmental features, such as those related to the spatial profile of the underground tunnel: the distribution of obstacles, the area and cross-section of the free space in the corridor, among others. Nevertheless, the findings of this paper may be a solid base for future deep-indoor signal propagation research and LP-WAN underground deployment studies.

REFERENCES

- Techjury, "Internet of things statistics 2020 [the rise of iot]," accessed: 24-04-2020. [Online]. Available: https://techjury. net/stats-about/internet-of-things-statistics/
- [2] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow band internet of things," *Ieee Access*, vol. 5, pp. 8038776, 20557–20577, 2017.
- [3] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 45.820, 12 2015, version 13.1.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/ Specifications/SpecificationDetails.aspx?specificationId=2719
- [4] P. Andres-Maldonado, P. Ameigeiras, J. Prados-Garzon, J. Navarro-Ortiz, and J. M. Lopez-Soler, "Narrowband iot data transmission procedures for massive machine-type communications," *Ieee Network*, vol. 31, no. 6, pp. 8 120 238, 8–15, 2017.
- [5] R. Harwahyu, R. G. Cheng, W. J. Tsai, J. K. Hwang, and G. Bianchi, "Repetitions versus retransmissions: Tradeoff in configuring nb-iot random access channels," *Ieee Internet of Things Journal*, vol. 6, no. 2, pp. 8 605 340, 3796–3805, 2019.
- [6] M. Lauridsen, I. Z. Kovács, P. Mogensen, M. Sørensen, and S. Holst, "Coverage and capacity analysis of Ite-m and nb-iot in a rural area," *Ieee Vehicular Technology Conference*, vol. 0, p. 7880946, 2016.
- [7] K. M. Malarski, A. Bardram, M. D. Larsen, J. Thrane, M. N. Petersen, L. Moeller, and S. R. Ruepp, "Demonstration of nb-iot for maritime use cases," *Proceedings of 2018 9th International Conference on the Network of the Future*, pp. 106–108, 2018.
- [8] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Z. Kovacs, P. Mogensen, and M. Sorensen, "Coverage and capacity analysis of sigfox, lora, gprs, and nb-iot," *Ieee Vehicular Technology Conference*, vol. 2017-, p. 8108666, 2017.

⁴Source: https://www.sigfox.com/en/what-sigfox/technology#id_radio

- [9] H. Malik, S. Z. Khan, J. L. R. Sarmiento, A. Kuusik, M. M. Alam, Y. Le Moullec, and S. Parand, "Nb-iot network field trial: Indoor, outdoor and underground coverage campaign," 2019 15th International Wireless Communications and Mobile Computing Conference (iwcmc), pp. 537–542, 537–542, 2019.
- [10] 3GPP, "TR 138 901 V14.3.0 5G; Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 14.3.0 Release 14)."
- [11] Itu-r, "Guidelines for evaluation of radio interface technologies for IMT-2020 M Series Mobile, radiodetermination, amateur and related satellite services." [Online]. Available: http://www.itu.int/ITU-R/go/patents/en
- [12] K. M. Malarski, J. Thrane, M. G. Bech, K. Macheta, H. L. Christiansen, M. N. Petersen, and S. R. Ruepp, "Investigation of deep indoor nbiot propagation attenuation," *Proceedings of 2019 Ieee 90th Vehicular Technology Conference*, vol. 2019-, p. 8891414, 2019.
- [13] J. Thrane, K. M. Malarski, H. L. Christiansen, and S. Ruepp, "Experimental evaluation of empirical nb-iot propagation modelling in a deep-indoor scenario," *accepted to: 2020 IEEE Globecom*, 2020.
- [14] M. F. Chen, H. C. Chiu, K. S. Tseng, Y. C. Yang, C. Y. Chou, and J. A. Jiang, "An iot-based temperature monitoring system for underground cable tunnels," 2019 Ieee Pes Gtd Grand International Conference and Exposition Asia, Gtd Asia 2019, pp. 8715849, 960–964, 2019.
- [15] A. Tan, S. Wang, N. Xin, Y. Shi, and Y. Peng, "A multi-channel transmission scheme in green internet of things for underground mining safety warning," *Ieee Access*, vol. 8, pp. 8932383, 775–788, 2020.
- [16] H. Malik, N. Kändler, M. M. Alam, I. Annus, Y. Le Moullec, and A. Kuusik, "Evaluation of low power wide area network technologies for smart urban drainage systems," 2018 Ieee International Conference on Environmental Engineering, Ee 2018 - Proceedings, pp. 1–5, 2018.
- [17] A. Hrovat, G. Kandus, and T. Javornik, "A survey of radio propagation modeling for tunnels," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 658–669, 2014.
- [18] K. I.-K. Wang, S. Wu, A. Ivoghlian, Z. Salcic, A. Austin, and X. Zhou, "Lws: A lorawan wireless underground sensor network simulator for agriculture applications," 2019 Ieee Smartworld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (smartworld/scalcom/uic/atc/cbdcom/iop/sci), pp. 475–482, 475–482, 2019.
- [19] S. Persia, C. Carciofi, M. Barbiroli, M. Teodori, V. Petrini, A. Garzia, and M. Faccioli, "Iot enabling technologies for extreme connectivity smart grid applications," 2019 Ctte-fitce: Smart Cities and Information and Communication Technology, Ctte-fitce 2019, p. 8894819, 2019.
- [20] K. Kousias, G. Caso, O. Alay, A. Brunstrom, L. De Nardis, M.-G. Di Benedetto, and M. Neri, "Coverage and deployment analysis of narrowband internet of things in the wild," 2020.
- [21] S. Gangakhedkar, Ö. Bulakci, and J. Eichinger, "Addressing deep indoor coverage in narrowband-5G," *IEEE Vehicular Technology Conference*, vol. 2017-Septe, no. Mcl, pp. 1–5, 2018.
- [22] L. M. C. Eraldo Damosso, COST Action 231: digital mobile radio towards future generation systems: final report. Brussels: European Comission, 1999.
- [23] X. Su, X. Yan, and C. L. Tsai, "Linear regression," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 4, no. 3, pp. 275–294, 2012.
- [24] T. A. Reddy and T. Agami Reddy, "Estimation of linear model parameters using least squares," *Applied Data Analysis and Modeling for Energy Engineers and Scientists*, pp. 141–182, 2011.
- [25] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in 9th Python in Science Conference, 2010.
- [26] J. Görtler, R. Kehlbeck, and O. Deussen, "A visual exploration of gaussian processes," *Distill*, 2019, https://distill.pub/2019/visual-explorationgaussian-processes.
- [27] C. E. Rasmussen and C. K. Williams, Gaussian processes for machine learning. MIT,, 2006.
- [28] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration," in *Advances in Neural Information Processing Systems*, 2018.
- [29] K. M. Malarski, J. Thrane, H. L. Christiansen, and S. Ruepp, "NB-IoT deep-indoor measurements in an underground tunnel system," IEEE Dataport, http://dx.doi.org/10.21227/xvrx-6r04.
- [30] L. Alliance, "Lorawan 1.1 specification," Tech. Rep., 2017, accessed: 24-04-2020. [Online]. Available: https://lora-alliance.org/resource-hub/ lorawanr-specification-v11