Localizing Basestations From End-User Timing Advance Measurements

LUKAS ELLER\textsuperscript{\textcopyright}, (Student Member, IEEE), VACLAV RAIDA\textsuperscript{\textcopyright}, (Graduate Student Member, IEEE), PHILIPP SVOBODA\textsuperscript{\textcopyright}, (Senior Member, IEEE), AND MARKUS RUPP\textsuperscript{\textcopyright}, (Fellow, IEEE)

Institute of Telecommunications, Technische Universität Wien, 1040 Vienna, Austria

Corresponding author: Lukas Eller (lukas.eller@tuwien.ac.at)

The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

ABSTRACT Although mobile communication has become ubiquitous in our modern society, operators typically treat the underlying networking infrastructure in a secretive manner. However, detailed topology information is a key enabler for operator benchmarking and can serve as ground-truth data for system-level simulations or user equipment localization schemes. Still, existing approaches for base station localization that are either based on received signal strength or on the spatial distribution of measurements are not accurate enough for such use cases. Accordingly, we propose a localization scheme that operates on end-user measurements of the 4G & 5G timing advance parameter, which acts as a quantized distance measure between the user equipment and the base station. By directly incorporating GPS noise, multipath propagation, and quantization into our stochastic system model, we obtain an estimator that offers a reliable measure of confidence and requires only the configuration of two hyperparameters with a dedicated physical interpretation. We evaluate our approach using a set of drive-test measurements consisting of 190 LTE eNodeBs with ground-truth locations confirmed by an Austrian mobile network operator. Our selective estimator can either operate without prior knowledge, resulting in mean distance errors of below 100 m, or in a classification setup, where it correctly identifies up to 95% of eNodeBs from a set of candidate cell tower locations. To allow for reproducibility, we make our dataset and a reference implementation publicly available.

INDEX TERMS Bayesian estimation, base station, cellular networks, eNodeB, inference, localization, LTE, selective prediction, timing advance, 5G.

I. INTRODUCTION

Despite the central role of mobile communications in our society, the infrastructure powering these networks is typically hidden from the everyday consumer. The advent of crowdsourcing in past decades has, however, increasingly exposed some aspects of the network topology, hence allowing researchers, regulators, and industry observers to monitor the performance of networks more closely [1]. In that context, the inference of base station (BS) locations from end-user measurements has been actively studied. Still, widely used platforms, such as OpenCellID, derive their estimates from simplistic algorithms, reporting the centroid of all measurement positions in a given cell [2]. Although the quality of such estimates may be sufficient for coarse user equipment (UE) positioning [3], it does not offer an accurate picture of the network layout, which is a key enabler for a wide range of novel use cases — see Figure 1. Such topology data can, for instance, act as a realistic basis for system-level simulations [4] or can improve network planning schemes through the examination of real-world layouts [5], [6]. This is of particular interest when considering the increased complexity of extremely heterogeneous and possibly non-static future network deployments [7]. Further, we can build upon reliable BS localization schemes by additionally retrieving the sectorization together with the corresponding transmitter orientations [8]. Combined results from localization and orientation inference can then be used to derive coverage or throughput maps or act as a baseline for further optimization of the network configuration [9]. Although performance maps are often generated through the local interpolation of measurements [10]–[12], alternative schemes requiring detailed knowledge of the network topology can significantly improve the estimation results [13], [14]. This is especially relevant for areas with an insufficient number of measurements, where the derived spatio-temporal performance maps can then power...
anticipatory networking schemes [15]. Moreover, we see network layout inference as a crucial stepping stone to fully exploit the potential of digital twins for 5G & 6G networks [16], [17]. With detailed network topology data, researchers can assess potential challenges and risks for novel applications before deployment. Clearly, also UE positioning itself benefits from improved BS localization: In a bootstrap manner, we can use noisy UE position estimates to determine BS locations, which can in turn act as a basis for the positioning of UEs whenever Global Navigation Satellite systems (GNSSs) are unavailable [18]. Overall, we identify two recent developments that render such detailed inference of the network topology feasible. First is the widespread availability of the 4G & 5G timing advance (TA) parameter on crowdsourcing platforms, which can be interpreted as a quantized time delay measure between the UE and the BS [19]. Second is the public accessibility of cell tower candidate locations in many countries [20], [21]. This effectively renders the problem of identifying the correct BS as a classification task. In this work, we introduce a novel scheme for base station localization from end-user TA measurements, which takes both of these developments into account.

We summarize our contribution as follows:

(i) Our estimation scheme relies on a stochastic system model that is grounded in wireless propagation research and requires only the specification of two hyperparameters with a straightforward physical interpretation. Also, it is flexible enough to identify the most likely location from a set of cell tower candidates or to evaluate the area of interest without any prior knowledge.

(ii) We design our estimator in such a way that allows us to make predictions with a reject-option. Hence, we abstain from prediction whenever the expected confidence is insufficient to ensure reliable inference.

(iii) Further, we evaluate our approach in an LTE network through an extensive drive-test campaign, consisting of a rural and an urban subsets to account for varying propagation environments. An Austrian mobile network operator (MNO) validated the ground-truth locations of the 190 considered eNodeBs.

Due to the more widespread adoption of LTE systems at the time of writing, we focus on the localization of eNodeBs in particular. However, the proposed scheme is directly applicable to 5G. In fact, we can expect a higher precision due to the increased spatial resolution of TA in 5G — see Appendix VIII for details. To allow for reproducibility, we also make a reference implementation and an anonymized version of our dataset publicly available — the link is provided in the footnote.

II. RELATED WORK

The idea of using time delay measurements for positioning is not particularly new. In fact, LTE provides a dedicated downlink (DL) signal for UE localization through the positioning reference signal (PRS) [22]. Correspondingly, the UE position can be derived through the time difference of arrival of PRS signals from multiple BSs via multilateration [23] — given that the exact eNodeB locations are known. For BS localization, we do not require signals from different eNodeBs. Instead, we consider a set of measurements—possibly from multiple UEs obtained via crowdsourcing — connected to the eNodeB of interest.

The most commonly found schemes in literature can be split into two categories: (i) approaches based on received signal strength (RSS) and (ii) geometric ones operating on the spatial distribution of measurements [24]. While RSS-based approaches have been successfully applied for related problems such as fingerprinting based small cell discovery [25], full localization schemes require accurate modelling of the expected path-loss over distance [26], [27]. This is especially tedious for urban environments with significant deviations due to shadow fading or blockages. While fusing measurements from multiple sectors can improve the performance, the obtained mean positioning error is still in the range of 900 m [26].

Geometric methods, on the other hand, infer the eNodeB location indirectly from the cell assignment by operating on the spatial distribution of the measurements. While the centroid method used in OpenCellID is a very basic variant of such an approach [2], there exist several extensions, most notably the weighted centroid method, also accounting for the reported reference signal received power (RSRP) at the UE location [28]. Nevertheless, these approaches have been found to be highly unreliable when comparing the results to ground-truth eNodeB locations [28]–[30].

To resolve these limitations, recent work in the field has used supervised machine learning techniques to automatically select the most promising algorithm for a given set of input data [29]. Similarly, the authors in [30] utilize unsupervised methods to identify predictive data instances, providing them with a measure of confidence. Another promising approach is to natively account for the sectorization of
eNodeBs, as the asymmetry of measurements for a sector causes deviations between the centroid and the BS location. In [31], the authors address this and proposes a purely geometric approach that infers BS positions by minimizing a cost matrix implied by a sectorization model, thus reducing the mean positioning error to around 500 m. Although these results are promising, we argue that the use of the now-widely available TA parameter can significantly improve the positioning accuracy as it is in a sense the direct equivalent of dedicated PRS-based localization. In fact, using TA for BS localization has already been proposed in GSM networks, where the low granularity of TA was identified as the main limitation [32]. With increased resolution, TA has gained renewed interest in LTE networks, especially for UE localization [33]–[35]. Most inference schemes do, however, use a deterministic system model or treat the overall error term as a single Gaussian [32]. We find this to be a drastic simplification, considering that the statistics of individual error sources, such as multipath propagation, are well understood in mobile communications [36, Tab. 7.3-6]. Meanwhile, the machine learning approach in [37] treats BS localization as a classification task, with candidates obtained from publicly available cell tower databases. While the image classification reformulation allows for effective fusion of RSS-, geometric-, and TA-based approaches, it requires a vast amount of training data, in contrast to a carefully designed model based approach accounting for propagation conditions. The influence of the wireless channel on location accuracy has also been studied in [22], where the authors state that actively handling multipath effects is expected to increase the performance of UE localization based on the PRS. Likewise, a stochastic scheme that can identify and discard non-line-of-sight (NLOS) measurements is proposed in [38]. For their TA-based UE localization scheme, the authors in [18] extend this notion by explicitly accounting for three different error sources that distort the reported travelled distance between UE and BS. As such, they introduce generic Gaussian measurement noise and a dedicated quantization term modeled as a Gaussian or uniform distribution. They further suggest to model the multipath component via a positively biased random variable.

For our BS localization scheme we further strengthen these assumptions through the following: (i) We discard the quantization noise, which we treat as a deterministic procedure, by incorporating it directly into our system model. (ii) We refrain from including a generic measurement error, as this holds the possibility of obtaining a negative value for the expected distance. Instead, we model the distance as a Rician, which is equivalent to assuming Gaussian noise on top of the reported UE position estimates. (iii) We utilize an exponential distribution for the additional multipath distance, which is in accordance with state-of-the-art propagation models.

III. DISTANCE FROM TIMING ADVANCE

TA is an integer \( k \in K \), \( K = \{0, \ldots, K - 1\} \) that a BS communicates to each served UE in order to compensate for different travel-times of signals in the uplink (UL). In LTE, the maximum TA value is given by \( K - 1 = 1282 \) [19, Section 4.2.3]. In a nutshell, TA tells the UE how much earlier it has to transmit its UL frames relative to the time of DL frames’ arrival, such that all UL frames from all UEs arrive at the BS approximately at the same time. Hence, TA characterizes a quantized round trip time between the BS and the UE. For more details, see [39, Section 8.1].

In our use case, we are interested in a measure of distance between the BS and the UE. According to [34], we can recalculate TA into a quantized one-way delay and—assuming the speed of light in vacuum—into the quantized distance \( s_k \in S \), \( S = \{s_0, \ldots, s_{K - 1}\} \):

\[
s_k = k \cdot d_{TA} \quad \text{with} \quad k \in K, \quad d_{TA} = 78.125 \text{ m}, \quad (1)
\]

where \( d_{TA} \) can be interpreted as the spatial resolution of TA for localization purposes.

We further denote the non-quantized travelled distance between the BS and the UE as \( s \). Note that \( s \) can take any nonnegative value \( s \in \mathbb{R}^+ \). Accordingly, the BS obtains \( k \) from \( s \) via quantization. In our system model, we assume that the BS quantizes the TA by selecting the \( k \) that minimizes the difference between the quantized and non-quantized round trip time (i.e., nearest neighbor). In accordance with (1), we define the quantization \( q : \mathbb{R}^+ \to S \) in terms of the travelled distance \( s \) as

\[
q(s)|_{s \in [e_k, e_{k+1})} = s_k, \quad \forall s \in \mathbb{R}^+, k \in K, \quad (2)
\]

with bin edges \( e_0, \ldots, e_K \):

\[
e_k = \begin{cases} 
0, & k = 0, \\
\frac{s_k + s_{k-1}}{2}, & k \in \{1, \ldots, K - 1\}, \\
\infty, & k = K.
\end{cases}
\]

We stress that this one particular implementation of \( q(\cdot) \) is not the only possible choice. Another alternative could be, for instance, truncation, i.e., the floor operation from [32]. Again, we refer the interested reader to Appendix VIII, for a discussion of the TA granularity in 5G systems — where we have \( d_{TA} \) values of as low as 4.883 m.

IV. SYSTEM MODEL

Fig. 2 illustrates our system model assumptions. Our goal is to estimate the exact BS location \( x_{BS} = (x_{BSx}, x_{BSy})^T \in \mathbb{R}^2 \) via multilateration from UE positions \( x_{UE} \) and their relative distances \( d \) to the BS.\(^2\) In practice, we receive only noisy estimates \( \hat{x}_{UE} \) of the user position via dedicated APIs [40], which requires us to incorporate the expected noise variance into the system model. Further, the relative distance \( d \) is not directly available but we only have access to quantized TA measurements \( k \). We also need to consider that the direct line of sight between the BS and UE may be blocked (Fig. 2). The non-quantized distance \( s \) that the electromagnetic signal

\(^2\)Note, that we limit ourselves to 2D formulation in this work. In principle, the system model can be extended to additionally infer the height of the BS.
locations into the matrix \( \hat{k} \) stacked into the vector \( k \) multilateration.

Accordingly, the reported TA value is only a distorted measure of the (unknown) true distance \( d \) required for correct multilateration.

For a set of \( M \) measurements, with TA measurements stacked into the vector \( k = (k^{(1)}, \ldots, k^{(M)})^T \) and noisy UE locations into the matrix \( \hat{X}_{UE} = (\hat{x}_{UE}^{(1)}, \ldots, \hat{x}_{UE}^{(M)}) \), we can formulate the above problem in the Bayesian framework. The estimated base station location \( \hat{x}_{BS} \) shall then maximize the posterior distribution given the measurements:

\[
\hat{x}_{BS} = \arg \max_x f(x_{BS}|k; \hat{X}_{UE}).
\]

(i) The exponential distribution maximizes entropy [42], (ii) Also, the choice is in accordance with the assumption presented in Figure 2. A block diagram highlighting all the components of the estimator is provided in Fig. 2. Further, we refer the reader to Appendix VIII for the derivation of the noise standard deviation \( \sigma \) corresponding to a particular value of \( \rho \).

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure3}
\caption{Overall estimator block diagram. We intend to obtain an estimate of \( \hat{x}_{BS} \) from measurements of \( k \) and \( \hat{x}_{UE} \).}
\end{figure}

\section{B. MULTIPATH PROPAGATION}

Due to multipath propagation, the travelled distance \( s \) is typically larger than the length of the direct path \( d \) between \( x_{UE} \) and \( x_{BS} \). Hence, we model the distribution of \( s \) as an exponential shifted by \( d \). Accordingly, we have

\[
f(s|d; \lambda) = \lambda e^{-\lambda(s-d)}u(s - d),
\]

where \( u(\cdot) \) denotes the unit step function

\[
u(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}
\]

Again, \( \lambda \) has a straightforward interpretation: its inverse \( \frac{1}{\lambda} \) specifies the mean of the expected distance increase caused by the multipath propagation in meters.

The choice of modelling \( s \) with (9) is motivated as follows: (i) The exponential distribution maximizes entropy [42], given the first moment and support \([0, \infty)\). We possess no additional information that would justify entropy reduction, i.e., more restrictive distribution, or specifying further moments. (ii) Also, the choice is in accordance with the 3GPP 3D channel model [36, Tab. 7.3-6], which employs the exponential distribution for the time-of-arrival delay of multipath components.

To eliminate \( d \) from (9), we first note that \( s \mid d \). I.e., \( s \) and \( d \) are conditionally independent given \( d \). If \( d \)
is known, then $d$ and $\sigma$ convey no additional information about $s$. Accordingly, we have

$$f(s|d, \hat{d}; \sigma, \lambda) = f(s|d; \lambda), \quad (11)$$

and we can express the joint conditional distribution as a product of (8) and (9):

$$f(s, d|\hat{d}; \sigma, \lambda) = f(s|d; \lambda) \cdot f(d|\hat{d}; \sigma). \quad (12)$$

By marginalizing over $d$, we obtain the distribution of $s$ conditioned on $\hat{d}$:

$$f(s|\hat{d}; \sigma, \lambda) = \int_0^\infty f(s, d|\hat{d}; \sigma, \lambda) \, dd = \int_0^\infty f(s|d; \lambda) f(d|\hat{d}; \sigma) \, dd = \int_0^\infty \lambda e^{-\lambda(s-d)} f(d|\hat{d}; \sigma) \, dd, \quad (13)$$

where we use that $u(s-d) = 0$ except for the case of $s \geq d$. Equation (13) can now be directly evaluated for fixed values of $\hat{x}_{UE}$ and $x_{BS}$ via (6).

### C. Timing Advance and Quantization

From $s$, we obtain the one-way delay distance $s_k$ through the quantization step defined in (3). Note that the reported TA values can then be derived from (1). Respectively, we obtain the conditional probability mass function (PMF) for $s_k$ given $\hat{d}$ by integrating (13) over the corresponding bin edges $e_k, e_{k+1}$, such that

$$p(s_k|\hat{d}; \sigma, \lambda) = \int_{e_k}^{e_{k+1}} f(s|\hat{d}, \sigma; \lambda) \, ds, \quad k \in K. \quad (14)$$

By inserting (9), (8), and (13) into (14) we get,

$$p(s_k|\hat{d}; \sigma, \lambda) = \int_{e_k}^{e_{k+1}} f(s|\hat{d}, \sigma; \lambda) \, ds = \frac{\lambda}{\sigma} \int_{e_k}^{e_{k+1}} e^{-\lambda(s-d)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-d)^2}{2\sigma^2}} \, ds.$$

For the case of a uniform categorical prior the maximum a posteriori (MAP) estimate then collapses into a maximum likelihood (ML) scheme. Alternatively, one could bypass evaluating the evidence in (20) by employing Markov Chain Monte Carlo (MCMC) methods, which would also allow us to select continuous priors for $x_{BS}$. In fact, we can adapt our system model in a straightforward manner to obtain an MCMC variant in the form of a PyMC3 implementation [43]. One way to achieve this would be to model the quantization step via a uniform distribution spanned by the bin edges $e_k, e_{k+1}$.

For our categorical approach, we select the prior candidates $\mathcal{X}_i$ from the set of feasible nearby locations $\mathcal{X}_i$, which we define through a distance threshold such that:

$$\mathcal{X}_i = \{x | \|x - x_{UE}^{(i)}\| \leq d_{lim}^{(i)} \text{ for any } i \in \{1, \ldots, N\}\}. \quad (22)$$

with the distance threshold $d_{lim}^{(i)}$ given by

$$d_{lim}^{(i)} = (k^{(i)} + 1) \cdot d_{TA} + 3 \cdot \sigma^{(i)}. \quad (23)$$

In practise, we approximate $\mathcal{X}_i$ via the smallest enclosing rectangle. We then either (i) sample $\mathcal{X}_i$ uniformly in a grid-search manner to obtain $\mathcal{X}_i$, with the granularity specified by the grid distance $D_x$, or (ii) choose the candidates from publicly available cell tower locations $\mathcal{X}_i$ such that

We refer the interested reader to Appendix VIII for a detailed derivation of the approximation. In our implementation, we select the approximation for $\frac{\hat{d}}{\sigma} \gg 10$. For $D_x = 20\,\text{m}$, this is equivalent to the distance threshold $\hat{d} > 133.33\,\text{m}$.

With the expression for the PMF of $s_k$, we can now evaluate

$$p(k|\hat{d}; \sigma, \lambda) = p(k \cdot d_{TA}|\hat{d}; \sigma, \lambda) \quad (18)$$

the likelihood of the TA measurements given the noisy distance $\hat{d}$ between $\hat{x}_{UE}$ and $x_{BS}$.

### V. Evaluation of the Posterior

With the likelihood of the TA measurements as a function of UE and BS positions derived as (15), we can obtain the posterior following Bayes’s rule:

$$f(x_{BS}|k; \hat{x}_{UE}) = \frac{p(k|x_{BS}; \hat{x}_{UE}) \cdot f(x_{BS})}{p(k)}.$$

The evaluation of the evidence $p(k)$ requires us to compute the integral

$$\int_{x_{BS}} p(k|x_{BS}; \hat{x}_{UE}) \cdot f(x_{BS}) \, dx_{BS}. \quad (20)$$

Hence, we limit ourselves to the finite set $\mathcal{X}_c$ containing $C$ candidate locations, i.e., we evaluate (19) for a categorical prior $p(x_{BS}^{(i)})$, which turns the integral (20) into the following summation:

$$p(k) = \sum_{i=1}^{C} p(k|x_{BS}^{(i)}; \hat{x}_{UE}) \cdot p(x_{BS}^{(i)}). \quad (21)$$

For the case of a uniform categorical prior the maximum a posteriori (MAP) estimate then collapses into a maximum likelihood (ML) scheme. Alternatively, one could bypass evaluating the evidence in (20) by employing Markov Chain Monte Carlo (MCMC) methods, which would also allow us to select continuous priors for $x_{BS}$. In fact, we can adapt our system model in a straightforward manner to obtain an MCMC variant in the form of a PyMC3 implementation [43]. One way to achieve this would be to model the quantization step via a uniform distribution spanned by the bin edges $e_k, e_{k+1}$.
\( X = X_t \cap X_f \). In the remainder, we denote the first approach as the grid-search prior and the latter as the cell-identification use case.

**VI. DRIVE-TEST DATASET AND RESULTS**

We evaluate our approach through a set of measurements from a live LTE network collected in a drive-test campaign in the surroundings of Vienna, Austria, during one day in August 2021. We used a Keysight Nemo measurement phone, which allows us to obtain TA measurements and their corresponding UE locations with a sufficiently short temporal interval \[44\]. Besides this, the used measurement phones are typical consumer grade Samsung Galaxy Note 4 devices. During the measurement we fixed the phones in the car trunk with a running download script and an active band lock at 800 Mhz — the equivalent measurement setup is described in more detail in \[45\, Sec. 3B\]. Further, the ground-truth eNodeB locations for all of these measurements have been confirmed by an Austrian MNO. We want to stress, that the drive-test was in no way optimized for eNodeB localization, i.e., the route was not adapted to increase the number or the spatial coverage of measurements per eNodeB. Instead we follow the direct route to the selected destination \(~70\ km\ east of Vienna’s city center, as we intend to evaluate our approach under a realistic end-user data collection scheme.

The characteristics of our dataset, with the measurements split into \textit{Rural} and \textit{Urban} subsets to assess the influence of the environment on the performance of the proposed estimator. It is apparent from Fig. 4 that the cell deployment is significantly denser in the \textit{Urban} subset. The \textit{Rural} subset has a larger cell radius; thus, we have a median number of measurements of 35 compared to 6 for the \textit{Urban} counterpart. At the same time, the area of feasible candidates \( X_f \) and the median number of cell tower candidates from

\[ X_t \cap X_f \) within it are substantially higher. Accordingly, the spatial range of the measurements per cell is also lower in the \textit{Urban} subset due to frequent handovers.

For illustration, Fig. 5 highlights selected measurements from the drive-test campaign. The values of \( s_k \), which are derived from the reported TA integers via (1), are shown over the covered drive-test distance — we further mark the cell identifiers in color to highlight handovers. Interestingly, we observe a triangular pattern for \( s_k \) as we enter and leave the coverage area of individual eNodeBs. In fact, we can even derive the cell radius, which we again identify as substantially smaller in the urban area than in the rural surroundings of Vienna.

**A. VISUAL ANALYSIS OF POSTERIOR**

Before evaluating the performance on the complete drive-test dataset, we first conduct a detailed visual analysis of a single prediction, in order to obtain an intuitive understanding of the inner workings of our estimator.

Fig. 6 showcases the prediction result for an exemplary eNodeB from the \textit{Rural} dataset, generated with \( \lambda = 0.01 \) and \( \rho = 20\ m \). Here, the uniform prior was evaluated in a grid-search manner such that \( X_c = X_g \). For reference, we provide the spatial distribution of the measurements together with the radius of the quantized distance measure \( s_k \). We also plot the cell tower candidates described in Sec. V and show a satellite image of the surroundings to highlight the propagation conditions.\footnote{Satellite images were obtained from Google Maps.}

In the single-measurement case in Figs. 6a and 6b, we obtain the circular symmetric posterior shown in Fig. 6c, where the locations inside the circle of Fig. 6c are more likely due to multipath propagation. Accordingly, the estimator correctly captures the uncertainty in the dataset and communicates it via the posterior distribution. When we increase the number of measurements to a total of \( M = 14 \), we obtain the posterior shown in Fig. 6f, with

\[ X_c \in X_t \cap X_f \] within it are substantially higher. Accordingly, the spatial range of the measurements per cell is also lower in the \textit{Urban} subset due to frequent handovers.

For illustration, Fig. 5 highlights selected measurements from the drive-test campaign. The values of \( s_k \), which are derived from the reported TA integers via (1), are shown over the covered drive-test distance — we further mark the cell identifiers in color to highlight handovers. Interestingly, we observe a triangular pattern for \( s_k \) as we enter and leave the coverage area of individual eNodeBs. In fact, we can even derive the cell radius, which we again identify as substantially smaller in the urban area than in the rural surroundings of Vienna.

**A. VISUAL ANALYSIS OF POSTERIOR**

Before evaluating the performance on the complete drive-test dataset, we first conduct a detailed visual analysis of a single prediction, in order to obtain an intuitive understanding of the inner workings of our estimator.

Fig. 6 showcases the prediction result for an exemplary eNodeB from the \textit{Rural} dataset, generated with \( \lambda = 0.01 \) and \( \rho = 20\ m \). Here, the uniform prior was evaluated in a grid-search manner such that \( X_c = X_g \). For reference, we provide the spatial distribution of the measurements together with the radius of the quantized distance measure \( s_k \). We also plot the cell tower candidates described in Sec. V and show a satellite image of the surroundings to highlight the propagation conditions.\footnote{Satellite images were obtained from Google Maps.}

In the single-measurement case in Figs. 6a and 6b, we obtain the circular symmetric posterior shown in Fig. 6c, where the locations inside the circle of Fig. 6c are more likely due to multipath propagation. Accordingly, the estimator correctly captures the uncertainty in the dataset and communicates it via the posterior distribution. When we increase the number of measurements to a total of \( M = 14 \), we obtain the posterior shown in Fig. 6f, with
the measurement distribution indicated in Figs. 6d and 6e. Clearly, the uncertainty decreases significantly, with the posterior now concentrated near the ground-truth eNodeB position. By directly reporting the MAP estimate, we obtain an error of 47 m, for a grid sampling distance of $D_s = 10$ m. Alternatively, we can treat the estimation problem in the cell-identification framework. When we consider the posterior in Fig. 6c, we note that even a single measurement allows us to effectively rule out a significant portion of the possible cell tower candidates. In the case of $M = 14$ we can already identify the correct eNodeB in Fig. 6f with high confidence. Clearly, the performance of such a scheme depends highly on the density of the cell tower deployment in the area of interest. Hence, we can expect the classification approach to be significantly more challenging in urban areas.

**B. CELL TOWER IDENTIFICATION WITH REJECT OPTION**

We now evaluate the cell-identification approach over the complete drive-test dataset. As such, we want to identify the serving eNodeB from the set of candidates $X_c = X_t \cap X_f$. Motivated by the results from the visual analysis, where our estimator reliably communicated the uncertainty in the prediction, we target a selective classification approach.

As a measure of confidence, we directly utilize the maximum posterior probability among all candidate locations:

$$p_{\text{conf}} = \max_i p(x_{\text{BS}}^{(i)}|k; \hat{X}_{\text{UE}}^i) \quad i \in \{1, \ldots, C\}. \quad (24)$$

We further define the selection function $g(\cdot)$ [47], which rejects the prediction whenever the confidence threshold $t_p$ is not fulfilled:

$$g(p_{\text{conf}}, t_p) = \begin{cases} 1, & p_{\text{conf}} \geq t_p, \\ 0, & \text{else} \end{cases} \quad (25)$$

Hence, the estimator abstains from uncertain predictions. Accordingly, we also make use of the notion of coverage $\phi$, which denotes the ratio of predictions accepted by $g(\cdot)$:

$$\phi(t_p) = \frac{1}{N} \sum_{i=1}^{N} g(p_{\text{conf}}^{(i)}, t_p). \quad (26)$$

The performance of our selective classifier can then be visualized by computing the misclassification rate $(1 - \text{Accuracy})$ over the coverage for all feasible values of $t_p$. 

**FIGURE 6.** Estimates for an exemplary eNodeB from the Rural dataset with different number of measurements $M$. 

(a) Single Measurement, $M = 1$

(b) Map for $M = 1$

(c) Posterior for $M = 1$

(d) Measurements, $M = 14$

(e) Map for $M = 14$

(f) Posterior for $M = 14$
Fig. 7 shows such a risk-coverage curve for all eNodeBs in the drive-test dataset. Clearly, our estimator allows us to trade-off accuracy for coverage. For the Rural dataset with \( \lambda = 0.01 \), we achieve a high baseline accuracy of over 90%. We can further increase this to 95% by abstaining from 10% of the predictions. Although the baseline accuracy of the Urban dataset is significantly lower, this is again reliably communicated through the confidence measure. At \( \lambda = 0.001 \), we observe a consistent increase in accuracy for lower coverage rates. We thus conclude that the estimator works reliably on the given dataset.

For reference, Fig. 7 also includes the risk-coverage curves of two baseline estimators. The Baseline No-Multipath scheme denotes a variant of our estimator, which does not incorporate multipath propagation into the system model. Hence, it only accounts for GPS noise and operates directly on the Rician likelihood defined in \( (8) \), which considerably reduces performance. To further evaluate the effect of multipath propagation we also include the Multipath Oracle estimator. In a sense, this scheme acts as the upper bound for the performance of our estimator; it operates on the TA values derived from our system model that we compensated for the multipath effects. As such, we use ground-truth eNodeB locations to compute discrete TA values \( k \) for the direct path between the UE and the BS in accordance with the quantization step defined in \( (15) \). The performance of the No-Multipath estimator on this new dataset can then be interpreted as the remaining error induced by the quantization and the GPS noise under the absence of multipath. In summary, we can conclude that our estimator can compensate for a significant part of the deviations caused by the propagation environment.

When comparing the different hyperparameter configurations in Fig. 7 we note, that a higher value of \( \lambda = 0.01 \) is beneficial when operating on the Rural subset, compared to \( \lambda = 0.001 \) for the Urban dataset. This is as expected, considering that multipath effects due to NLOS scenarios are typically more common in densely populated areas. We see this as a major strength of our approach — it requires only two dedicated hyperparameters with a straightforward physical interpretation. Accordingly, we can select suitable values based on a priori knowledge of the propagation conditions.

Tab. 1 summarizes the results on both datasets, where we added the mean distance errors and the coverage values corresponding to the selected threshold \( t_p \). Note that the error is calculated as the distance between the ground-truth location and the selected candidate. With \( t_p = 0.8 \), both datasets achieve a comparable accuracy, albeit we have to discard a significantly higher number of predictions for Urban. Correspondingly, the mean distance error is 5 m for Urban and 15 m for Rural. In summary, we find these results promising, considering that even for the lower threshold of \( t_p = 0.5 \), the distance error amounts to 66 and 14 m for Urban and Rural, respectively.

### C. Results for Grid-Search Prior

Analogous to the cell tower identification use case, we can also conduct selective prediction for the uniform prior evaluated in a grid-search manner. Here, we directly utilize the spatial layout of the prior candidates to obtain a measure of uncertainty via the posterior standard deviation.

We thereby treat the standard deviation component-wise in each direction, such that \( \sigma_{\text{post}} = (\sigma_{\text{post}, 0}, \sigma_{\text{post}, 1})^\top \). The first component \( \sigma_{\text{post}, 0} \) can then be computed by evaluating the expression

\[
\sigma_{\text{post}, 0} = \sqrt{\sum_{i=1}^{C} p_i \cdot (\mathbf{X}_{\text{BS, 0}}^{(i)} - \mu_{\text{BS, 0}})^2},
\]

with \( \mathbf{X}_{\text{BS, 0}}^{(i)} \) taken from the \( i \)-th BS candidate location on the grid. Hereby, we obtain the component-wise mean \( \mu_{\text{BS, 0}} \) via

\[
\mu_{\text{BS, 0}} = \sum_{i=1}^{C} p_i \cdot \mathbf{X}_{\text{BS, 0}}^{(i)}.
\]

In both cases, we use \( p_i \) as a shorthand for the posterior, i.e.,

\[
p_i = p \left( \mathbf{X}_{\text{BS}}^{(i)} | \mathbf{k} ; \hat{\mathbf{X}}_{\text{UE}} \right).
\] Clearly, the procedure is equivalent for the second component \( \sigma_{\text{post}, 1} \). We further adapt the selection function, such that it

---

**TABLE 1.** Performance of the proposed localization scheme in the cell tower identification framework with \( \lambda = 0.001 \) for Urban and \( \lambda = 0.01 \) for Rural.

| \( t_p \) | Urban | Rural |
| --- | --- | --- |
| Accuracy | 73 % | 94 % |
| Mean distance error | 66 m | 14 m |
| Coverage | 38 % | 95 % |

---

**FIGURE 7.** The misclassification rate is significantly lower in Rural. In both cases, coverage can be traded off for accuracy.
rejects predictions based on the standard deviation threshold $t_{\sigma_{\text{post}}}$.

$$g(r_{\text{post}}, t_{\sigma_{\text{post}}}) = \begin{cases} 
1, & \max(\sigma_{\text{post},0}, \sigma_{\text{post},1}) \leq t_{\sigma_{\text{post}}} \\
0, & \text{else}.
\end{cases}$$

(30)

Hence, the measure of uncertainty is given by the maximum over the posterior standard deviation vector $\sigma_{\text{post}}$.

Fig. 8 showcases the performance of the selective regressor, with grid granularity $D_g = 10$ m in the form of a risk-coverage curve. Interestingly, the performance for both Rural and Urban is similar over a wide range of the coverage even though the results of the cell-identification use case suggest differently. Hence, we can conclude that the significant difference between the performances of Rural and Urban in Fig. 7 is mainly due to the denser deployment in urban areas, where even a small distance error can result in incorrect identification of eNodeBs. For larger values of the coverage, we do note a drastic increase in the mean distance error for Rural. This is expected considering the result shown in Fig. 6c. In a rural environment, we obtain significantly larger offsets for estimates with high uncertainty. We can, however, identify such cases reliably and subsequently abstain from prediction.

Again, the results show that our estimator works well on the real-world dataset. Clearly, we can trade off prediction error for coverage; for low thresholds of $t_{\sigma_{\text{post}}}$, we reach mean distance errors close to the TA granularity $d_{TA}$. This allows us to conduct reliable eNodeB location inference, even without access to prior cell tower candidates. Meanwhile, the performance of the Multipath Oracle again showcases that multipath propagation is the most significant error source for TA-based localization. Our estimator, which directly incorporates the multipath statistics, again performs better than the No-Multipath baseline. The improvement is especially large for low values of the coverage, where also the performance gap between our estimator and the Multipath Oracle gets smaller.

For comparison, the mean distance error for commonly used baseline approaches is shown in Tab. 2. In particular, for the centroid, weighted centroid and max RSRP estimators discussed in [28]. The comparison with Fig. 8 shows, that the proposed estimator achieves a notably smaller positioning error than these geometric approaches, especially on the Rural subset with significantly larger cell radiuses. Unsurprisingly, these geometric approaches can only operate on crowdsourced data with numerous measurements and fail when applied to sparse drive-test data such as ours. Interestingly, our TA based approach also outperforms estimates from major localization platforms, even though these utilize a large set of crowdsourced measurements. Tab. 2 provides the positioning errors obtained by querying OpenCellID and Google Geolocation for the eNodeBs in our drive-test dataset [2], [48]. In comparison, our estimator seems to be well suited even for a small number of measurements, considering that our estimates are only based on a single non optimized drive-test campaign. In such cases, the uncertainty is reliably captured via the measure of confidence, while providing additional measurements further increases the achieved coverage. Correspondingly, we see the proposed scheme as an appropriate choice for both drive-test and crowdsourced data.

| TABLE 2. Mean distance error for estimates from crowdsourced databases and baseline approaches from the literature. Percentages refer to the ratios of eNodeBs from the drive-test dataset available on the respective platforms. |
|-----------------|-----------|-----------|
|                | Urban     | Rural     |
| OpenCellID     | 499 m     | 1470 m    |
| Google Geolocation | 259 m     | 849 m     |
| Centroid       | 249 m     | 1123 m    |
| Weighted Centroid | 230 m     | 990 m     |
| Max RSRP       | 229 m     | 780 m     |

VII. DISCUSSION

When applying our approach to a set of LTE drive-test measurements in the surroundings of Vienna, Austria, we identify significant differences between the Rural and Urban subsets of our drive-test campaign. Due to the denser deployment in urban areas, identifying the cell towers from candidate locations is significantly more challenging in Urban than in Rural. On the other hand, the close proximity of neighboring eNodeBs in Urban results in a lower mean distance error as compared to Rural. In both cases, the notion of uncertainty provided by our estimator proved to be a critical feature. It allows us to abstain from prediction whenever the confidence of the given estimate is not sufficient. Accordingly, we can achieve an accuracy of up to 95% for eNodeB identification and a mean distance error well below 100 m for position inference in the grid-search prior framework. Thus,
our proposed estimator outperforms all considered baseline approaches as well as positioning estimates from major crowdsourcing platforms. Notably, this is also a considerable improvement over the positioning errors reported for comparable datasets in the literature, with around 500 m for the best performing approach [31]. We also identify the correct modeling of multipath propagation as a crucial requirement for TA based localization, considering that the evaluated Baseline No-Multipath estimator suffers from a severely higher error. Similarly, the reference Multipath Oracle estimator, operating on a dataset compensated for multipath effects, reveals that the additional travelled distance is the dominating error component. In contrast, the errors due to quantization and GPS noise are significantly lower and thus not a limiting factor for such TA-based approaches for inferring the network topology — especially when considering the increased TA resolution for 5G networks, which will provide a spatial granularity of as low as 4.883 m.

VIII. FINAL CONCLUSION

In this work, we introduced a stochastic scheme for inference of BSs from end-user TA measurements. The proposed approach directly incorporates GPS noise, multipath propagation, and quantization into the system model, resulting in an estimator that requires only the configuration of two hyperparameters with a dedicated physical interpretation. Hence, we purposely avoid introducing generic error terms, but explicitly account for the individual components of our system model. Our estimator can then be deployed in two variants: (i) either in a cell-identification setup to identify the serving BS from a set of cell tower candidates or (ii) without any prior knowledge, where we sample the area of interest in a grid-search manner. In both cases, the proposed estimator provides a reliable measure of confidence and significantly outperforms the baseline approaches as well as existing crowdsourced localization platforms, achieving an identification accuracy of up to 95% and a mean positioning error well below 100 m. Overall, these results are promising, considering that our drive-test campaign was purposefully not optimized for the particular use case of BS localization to account for realistic end-user behaviour. Correspondingly, one can apply the proposed algorithm to existing drive-test or crowdsourcing data and evaluate the need for additional dedicated measurements through the provided measure of uncertainty. We see this as a key enabler for deploying the proposed BS localization scheme for large scale, possibly country-wide inference of the network topology.

APPENDIX A

TIMING ADVANCE RESOLUTION IN 5G

Analogous to LTE, TA for 5G is reported as the integer

\[ k \in \mathbb{K}, \quad \mathbb{K} = \{0, \ldots, K - 1 \} \]

— denoted as \( T_A \) in the 3GPP standard with a maximum value of \( K - 1 = 3846 \) [49, Section 4.2].

A particular value of \( k \) corresponds to the time duration

\[ N_{TA} = k \cdot 16 \cdot 64 \cdot 2^{-\mu}, \quad (31) \]

given in the 5G-NR basic time-unit of \( T_c \) [49, Section 4.2], which can be computed as

\[ T_c = \frac{1}{\Delta f_{\text{max}} \cdot N_f} \quad \text{[s]}, \quad (32) \]

with \( \Delta f_{\text{max}} = 480 \cdot 10^3 \text{ Hz} \) and \( N_f = 4096 \) [50, Section 4.1]. In (31), \( \mu \) denotes the subcarrier spacing configuration, which is described in detail in [50, Section 4.3.2]. With the speed of light in vacuum (\( \sim 3 \cdot 10^8 \text{ m/s} \)) and \( k = 1 \), the TA granularity for localization purposes \( d_{TA} \) can then be computed as

\[ d_{TA} = \frac{1}{2} \cdot N_{TA} \cdot T_c \cdot 3 \cdot 10^8 \text{ [m]}, \quad (33) \]

In Tab. 3, the granularity \( d_{TA} \) for the possible configurations of \( \mu \) is presented in meters. Note, that the LTE granularity is obtained as the baseline case with \( \mu = 0 \) — see [34] for comparison.

APPENDIX B

ANDROID LOCATION ACCURACY

For the case of Android, we can derive a suitable value for \( \sigma \) from the reported accuracy measure. The API [40] defines the horizontal accuracy \( \rho \) as a radius of 68% confidence. That is, with probability \( P = 0.68 \), the true location \( \hat{x}_{\text{UE}} \) lies inside the circle with center \( \hat{x}_{\text{UE}} \) and radius \( \rho \). Integrating a two-dimensional Gaussian in polar coordinates yields

\[ P = \int_0^{2\pi} \int_0^{\rho} \frac{1}{2\pi \sigma^2} e^{-\frac{r^2}{2\sigma^2}} r \, dr \, d\theta = -e^{-\frac{\rho^2}{2\sigma^2}} \bigg|_0^\rho = 1 - e^{-\frac{\rho^2}{2\sigma^2}}. \quad (34) \]

Hence, the standard deviation \( \sigma \) can be calculated from the accuracy \( \rho \) as follows:

\[ \sigma = \frac{\rho}{\sqrt{-2 \ln(1 - P)}}. \quad (35) \]

With \( P = 0.68 \), we obtain \( \sigma \approx \rho/1.5 \).

APPENDIX C

GAUSSIAN APPROXIMATION

For the case of small GPS noise \( \sigma \) and large reported distances \( \hat{d} \), we can approximate the Ricean distributed distance \( d \) via a Gaussian distribution. Specifically, we have

\[ f \left( \hat{d} \mid \sigma, \lambda \right) \approx \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\hat{d} - \lambda)^2}{2\sigma^2}} \quad (36) \]

whenever \( \hat{d}/\sigma \gg 1 \). Accordingly, we can approximate the likelihood via the well-known Gaussian error function

\[ p \left( s_k \mid \hat{d}, \sigma, \lambda \right) \approx = \frac{1}{\sigma \sqrt{2\pi}} \int_0^{\infty} e^{-\frac{(d - \lambda)^2}{2\sigma^2}} \left( 1 - e^{-\lambda(d_k - \hat{d})} \right) \, dd \]

TABLE 3. 5G TA localization granularity \( d_{TA} \) in meters, reported for different subcarrier-spacing configurations \( \mu \).

| \( \mu \) | \( d_{TA} \) [m] |
|---|---|
| 0 | 78.125 |
| 1 | 39.063 |
| 2 | 19.531 |
| 3 | 9.766 |
| 4 | 4.883 |
Thus, avoiding the numerical computation of the integral.

**APPENDIX D**

**NOTATION TABLE**

A summary of the used notation is provided in Tab. 4.

| $\mathbf{x_{BS}}$ | Ground-truth BS position from MNO |
| $\hat{\mathbf{x}}_{UE}$ | Position estimate reported by UE |
| $\mathbf{x}_{UE}$ | Unknown true UE position |
| $k$ | TA index as reported by UE |
| $d_{TA}$ | Granularity of TA for localization |
| $s_{k}$ | Quantized travelled distance derived from $k$ |
| $\sigma^{2}$ | Expected noise variance of GPS estimate |
| $\rho$ | Radius of horizonal GPS accuracy |
| $\lambda$ | Inverse of expected multipath component |
| $e_{k}$ | Bin edges of quantizer $q(\cdot)$ |
| $\mathcal{X}_{C}$ | Set of candidate BS positions |
| $\mathcal{X}_{F}$ | Feasible area for BS positions |
| $\mathcal{X}_{g}$ | $\mathcal{X}_{F}$ sampled grid-wise with granularity $D_{g}$ |
| $\mathcal{X}_{l}$ | Known set of unlabeled cell tower locations |
| $\phi(\cdot)$ | Coverage for a given value of $t_{p}$ or $t_{\text{sigma}}$ |
| $t_{p}$ | Lower threshold for posterior confidence |
| $t_{\text{sigma}}$ | Upper threshold for posterior variance |

\[
- \frac{1}{\sigma \sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{(d_{k}^{2} - d_{\lambda}^{2})}{2\sigma^{2}}} \left(1 - e^{-\lambda(d_{k} - d_{\lambda})}\right) \, dd = \\
\frac{1}{2} \left[ \text{erf} \left( \frac{d_{k} - d_{\lambda}}{\sigma\sqrt{2}} \right) + e^{\frac{d_{k}^{2}}{\sigma^{2}}} \text{erf} \left( \frac{d_{k}^{2} - d_{\lambda}}{\sigma\sqrt{2}} \right) \right]_{d_{k}^{2}}^{d_{k}^{2} + d_{\lambda}^{2}} - \\
\frac{1}{2} \left[ \text{erf} \left( \frac{d_{k}^{2}}{\sigma\sqrt{2}} \right) + e^{d_{k}^{2}} \text{erf} \left( \frac{d_{k}^{2} - d_{\lambda}}{\sigma\sqrt{2}} \right) \right]_{d_{k}^{2}}^{d_{k}^{2} + d_{\lambda}^{2}} \right].
\]

(37)

REFERENCES

[1] V. Raida, P. Svoboda, M. Lerch, and M. Rupp, “Crowdsensed performance benchmarking of mobile networks,” *IEEE Access*, vol. 7, pp. 154899–154911, 2019.

[2] OpenCellID. Accessed: Dec. 2021. [Online]. Available: http://wiki.opencellid.org/wiki

[3] M. Ulm, P. Widhalm, and N. Brändle, “Characterization of mobile phone localization errors with OpenCellID data,” in *Proc. 4th Int. Conf. Adv. Logistics Transp. (iCALT)*, May 2015, pp. 100–104.

[4] M. K. Müller, F. Ademaj, T. Dittrich, A. Fastenbauer, B. R. Elbal, A. Nabi, N. Nagel, S. Schwarz, and M. Rupp, “Flexible multi-node simulation of cellular mobile communications: The Vienna 5G system level simulator,” *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–17, Dec. 2018.

[5] A. Tautfieke, M. Jaber, A. Imran, Z. Dawy, and E. Yacoub, “Planning wireless cellular networks of future: Outlook, challenges and opportunities,” *IEEE Access*, vol. 5, pp. 4821–4845, 2017.

[6] R. M. Dreifuß, S. Daumton, Y. Qian, P. Varkey, M. Balandat, S. Kasturia, A. Tomar, A. Yazdan, V. Ponnampalam, and R. W. Heath, Jr., “Optimizing coverage and capacity in cellular networks using machine learning,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 8138–8142.

[7] P. Yang, Y. Xiao, M. Xiao, and S. Li, “6G wireless communications: Vision and potential techniques,” *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul./Aug. 2019.

[8] L. Eller, P. Svoboda, and M. Rupp, “Bayesian inference of source orientation in LTE networks based on end-user measurements,” in *Proc. IEEE 94th VTC-Fall*, Sep. 2021, pp. 1–7.

[9] E. Tekgul, T. Novlan, S. Akoum, and J. G. Andrews, “Sample-efficient learning of cellular antenna parameter settings,” in *Proc. IEEE Inf. Theory Workshop (IFW)*, Oct. 2021, pp. 1–6.

[10] L. Eller, P. Svoboda, and M. Rupp, “Propagation-aware Gaussian process regression for signal-strength interpolation along street canyons,” in *Proc. IEEE 93rd VTC-Spring*, Apr. 2021, pp. 1–7.

[11] S. Tripkovic, P. Svoboda, V. Raida, and M. Rupp, “Cluster density in crowdsourced mobile network measurements,” in *Proc. IEEE 93rd VTC-Spring*, Apr. 2021, pp. 1–7.

[12] J. Riihijarvi and P. Mahonen, “Machine learning for performance prediction in mobile cellular networks,” *IEEE Comput. Intell. Mag.*, vol. 13, no. 1, pp. 51–60, Feb. 2018.

[13] J. Thane, B. Sliwa, C. Wietfeld, and H. L. Christiansen, “Deep learning-based signal strength prediction using geographical images and expert knowledge,” in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[14] O. Rozenblit, Y. Haddad, Y. Mirsky, and R. Azoulay, “Machine learning methods for SIR prediction in cellular networks,” *Phys. Commun.*, vol. 31, pp. 239–253, Dec. 2018.

[15] B. Sliwa and C. Wietfeld, “Data-driven network simulation for performance analysis of anticipatory vehicular communication systems,” *IEEE Access*, vol. 7, pp. 172638–172653, 2019.

[16] H. X. Nguyen, R. Trestian, D. To, and M. Tatipamula, “Digital twin for 5G and beyond,” *IEEE Commun. Mag.*, vol. 59, no. 2, pp. 10–15, Feb. 2021.

[17] R. Dong, C. She, W. Hardjwana, Y. Li, and B. Vucetic, “Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4692–4707, Oct. 2019.

[18] J. D. Roth, M. Tummala, J. C. McEachen, J. W. Scrofani, and R. A. DeGabriele, “Maximum likelihood geolocation in LTE cellular networks using the timing advance parameter,” in *Proc. 10th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2016, pp. 1–10.

[19] Physical Layer Procedures, document TS 36.213, Version 14.11.0, 3GPP, 2019.

[20] Senderkataster. Accessed: Oct. 2021. [Online]. Available: https://www.senderkataster.at

[21] EMF-Karte Bundesnetzagentur. Accessed: Oct. 2021. [Online]. Available: https://www.bundesnetzagentur.de/DE/EN/portal/T2/TK/Funktechnik/EMF/ start.html

[22] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci, “Evaluation of the LTE positioning capabilities under typical multipath channels,” in *Proc. 6th Adv. Satell. Multimeda Syst. Conf. (AMSIS), 12th Signal Process. Space Commun. Workshop (SPSC)*, Sep. 2012, pp. 139–146.

[23] A. Kangas, I. Siomina, and T. Wigren, “Positioning in LTE,” in *Handbook of Position Location: Theory, Practice, and Advances*. Hoboken, NJ, USA: Wiley, 2011, pp. 1081–1127.

[24] J. A. N. Rusvik, “Localizing cell towers from crowdsourced measurements,” M.S. thesis, Dept. Inform. Algorithms, Univ. Bergen, Bergen, Norway, 2015.

[25] A. Prasad, P. Lundén, O. Tirkkonen, and C. Wijting, “Energy efficient small-cell discovery using received signal strength based radio maps,” in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–5.

[26] S. Heinberg and J. Garcia, “Localization performance for eNodeBs using solitary and fused RSS-modeling approaches,” in *Proc. Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2019, pp. 127–132.

[27] Y. Zhang, L. Bao, M. Welling, and S.-H. Yang, “Base station localization in search of empty spectrum spaces in cognitive radio networks,” in *Proc. 5th Int. Conf. Mobile Ad-Hoc Sensor Netw.*, 2009, pp. 94–101.

[28] E. Neidhardt, A. Uzun, U. Bareth, and A. Küpper, “Estimating locations and coverage areas of mobile network cells based on crowdsourced data,” in *Proc. 6th Joint IFIP Wireless Mobile Netw. Conf. (WMNC)*, Apr. 2013, pp. 1–8.

[29] M.-R. Fida and M. K. Marina, “Uncovering mobile infrastructure in developing countries with crowdsourced measurements,” in *Proc. 10th Int. Conf. Inf. Commun. Technol. Develop.*, Jan. 2019, pp. 1–11.

[30] Z. Li, A. Nika, X. Zhang, Y. Zhu, Y. Yao, B. Y. Zhao, and H. Zheng, “Identifying value in crowdsourced wireless signal measurements,” in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 607–616.
[31] S. Sundberg and J. Garcia, “Locating eNodeBs through sectorization inference—Sector fitting evaluated on a railway use case,” Comput. Netw., vol. 190, May 2021, Art. no. 107945.

[32] M. Raitoharju, S. Ali-Löytty, and L. Wirola, “Estimation of base station position using timing advance measurements,” Proc. SPIE, vol. 8285, Mar. 2011, Art. no. 82855.

[33] L. Jarvis, J. McEachen, and H. Loomis, “Geolocation of LTE subscriber stations based on the timing advance ranging parameter,” in Proc. MIL- COM, Nov. 2011, pp. 180–187.

[34] T.-H. Ngo and Y. Kim, “Using timing advance to support proximity discovery in network-assisted D2D communication,” in Proc. 7th Int. Conf. Ubiquitous Future Netw., Jul. 2015, pp. 926–928.

[35] T. Hiltunen, J. Turkka, R. Mondal, and T. Ristaniemi, “Performance evaluation of LTE radio fingerprint positioning with timing advancing,” in Proc. 10th Int. Conf. Inf., Commun. Signal Process. (ICICS), Dec. 2015, pp. 1–5.

[36] Study on 3D Channel Model for LTE, document TR 36.873, Version 12.7.0, 3GPP, 2018.

[37] A. Ghasemi and J. Parekh, “Deep learning based localization of LTE eNodeBs from large crowdsourced smartphone datasets,” in Proc. IEEE 93rd VTC-Spring, Apr. 2021, pp. 1–7.

[38] Y.-T. Chan, W.-Y. Tsui, H.-C. So, and P.-C. Ching, “Time-of-arrival based localization under NLOS conditions,” IEEE Trans. Veh. Technol., vol. 55, no. 1, pp. 17–24, Jan. 2006.

[39] Physical Channels and Modulation, document TS 36.211 Version 14.11.0, 3GPP, 2019.

[40] Android Developers. (2021). Location. [Online]. Available: https://developer.android.com/reference/android/location/Location#getAccuracy

[41] V. Raida, P. Svoboda, M. Lerch, and M. Rupp, “Repeatability for spatiotemporal throughput measurements in LTE,” in Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring), Apr. 2019, pp. 1–5.

[42] H. Hofmann, H. Wickham, and K. Kafadar, “Letter-value plots: Boxplots for large data,” J. Comput. Graph. Statist., vol. 26, no. 3, pp. 460–477, 2017.

[43] Y. Gelfman and R. El-Yaniv, “Selective classification for deep neural networks,” May 2017, arXiv:1705.08500.

[44] Keysight Nemo Measurement Solution. Accessed: Oct. 2021. [Online]. Available: https://www.keysight.com/buttons/product/NTH00000B/nemo-handheld-measurement-solution.html

[45] V. Raida, P. Svoboda, M. Lerch, and M. Rupp, “Repeatability for spatiotemporal throughput measurements in LTE,” in Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring), Apr. 2019, pp. 1–5.

[46] Google Geolocation. Accessed: Dec. 2021. [Online]. Available: https://developers.google.com/maps/documentation/geolocation

[47] 5G/NR:Physical Layer Procedures for Control, document TS 138.213, Version 15.13.0, 3GPP, 2021.

[48] Google Geolocation. Accessed: Dec. 2021. [Online]. Available: https://developers.google.com/maps/documentation/geolocation

[49] 5G/NR:Physical Channels and Modulation, document TS 138.211, Version 16.2.0, 3GPP, 2020.