

Evaluation of a Gaussian Mixture Model-based Channel Estimator using Measurement Data

Nurettin Turan, Benedikt Fesl, Moritz Grundei, Michael Koller, and Wolfgang Utschick
Professur für Methoden der Signalverarbeitung, Technische Universität München, 80333 Munich, Germany
Email: {nurettin.turan,benedikt.fesl,moritz.grundei,michael.koller,utschick}@tum.de

Abstract—In this work, we use real-world data in order to evaluate and validate a machine learning (ML)-based algorithm for physical layer functionalities. Specifically, we apply a recently introduced Gaussian mixture model (GMM)-based algorithm in order to estimate uplink channels stemming from a measurement campaign. For this estimator, there is an initial (offline) training phase, where a GMM is fitted onto given channel (training) data. Thereafter, the fitted GMM is used for (online) channel estimation. Our experiments suggest that the GMM estimator learns the intrinsic characteristics of a given base station’s whole radio propagation environment. Essentially, this ambient information is captured due to universal approximation properties of the initially fitted GMM. For a large enough number of GMM components, the GMM estimator was shown to approximate the (unknown) mean squared error (MSE)-optimal channel estimator arbitrarily well. In our experiments, the GMM estimator shows significant performance gains compared to approaches that are not able to capture the ambient information. To validate the claim that ambient information is learnt, we generate synthetic channel data using a state-of-the-art channel simulator and train the GMM estimator once on these and once on the real data, and we apply the estimator once to the synthetic and once to the real data. We then observe how providing suitable ambient information in the training phase beneficially impacts the later channel estimation performance.

Index Terms—Gaussian mixture models, measurement data, machine learning, channel estimation, ambient information

I. INTRODUCTION

Modern communications systems increasingly utilize ML algorithms to meet the compound requirements of high-dimensional channel estimation (CE) in massive multiple-input multiple-output (MIMO) orthogonal frequency-division multiplexing (OFDM) applications [1]. The channel characteristics of the whole propagation environment of a base station (BS) cell can be described by means of a probability density function (PDF) f_h . This PDF f_h describes the stochastic nature of all channels in the whole coverage area of a BS and therefore captures ambient information. Every channel of any mobile terminal (MT) within the BS cell is a realization of a random variable with PDF f_h . The main problem is that this PDF is typically not available analytically. For this reason, many classical channel estimation approaches cannot be applied or

this ambient information is ignored and replaced with Gaussian assumptions which may only hold locally around a given user. In this setting, ML approaches play an increasingly important role. These aim to (implicitly) learn the underlying PDF from data samples such that the ambient information is taken into account in ML channel estimation algorithms [2]–[4].

According to this development, many new channel models have been designed to capture the complex channel characteristics of a whole BS environment. Modern channel simulators based on ray tracing or stochastic-geometric models allow for the generation of large synthetic datasets that can be used for training and testing, e.g., [5]–[7]. Even though such increasingly complex simulators generate ever more realistic BS scenarios, it is crucial to also evaluate the performance of physical layer (PHY) algorithms on real-world data, i.e., on data collected in a measurement campaign. This is especially important and interesting for ML algorithms which mainly depend on the underlying data (PDF), e.g., [2], [4], [8].

An evaluation of an ML-based CE algorithm on measurement data was done in [9], where a neural network-based estimator is analyzed. However, the estimator in [9] was originally derived via assumptions on the channel model and on the antenna configuration which might not hold in practice. In this work, we evaluate the recently proposed GMM-based channel estimator from [4] on data originating from the same measurement campaign. The estimator first approximates the PDF f_h of the whole radio propagation environment with a GMM. This is done offline and only once. Thereafter, the estimator utilizes this ambient information for CE in the online phase. The estimator is proven to asymptotically converge to the optimal conditional mean estimator (CME) (which would be calculated using the unknown PDF f_h) but so far was only evaluated on synthetic data.

Our experiments indicate that the GMM estimator captures the ambient information well because it outperforms state-of-the-art CE algorithms evaluated (and trained) on the same measurement data. In particular, we achieve lower MSEs and higher spectral efficiencies. In addition, we generate synthetic channel data using a state-of-the-art channel simulator and train the GMM estimator once on these and once on the measurement data, and we apply the estimator once to the synthetic and once to the measurement data for evaluation. We observe that providing suitable ambient information in the training phase beneficially impacts the CE performance.

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the program of “Souverän. Digital. Vernetzt.”. Joint project 6G-life, project identification number: 16KISK002

©This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

The remainder of this work is organized as follows. In Section II, the GMM channel estimator is explained and in Section III, the measurement campaign is described and a channel simulator is introduced for comparison. Section IV provides simulation results and Section V concludes this work.

II. GAUSSIAN MIXTURE MODEL CHANNEL ESTIMATOR

We consider CE in the uplink from a single-antenna MT located within the cell to a BS. The BS is equipped with N antennas. After correlating with the commonly known pilot sequence, we obtain the noisy observation

$$\mathbf{y} = \mathbf{h} + \mathbf{n} \in \mathbb{C}^N, \quad (1)$$

where $\mathbf{h} \in \mathbb{C}^N$ is the uplink-channel of a certain MT located within the coverage area of the BS and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{C}_n = \sigma^2 \mathbf{I})$ denotes the additive white Gaussian noise (AWGN). The goal is then to estimate the channel \mathbf{h} given \mathbf{y} , i.e., to denoise the observation \mathbf{y} . The stochastic nature of all channels in the whole coverage area of the BS is assumed to be described by means of a continuous PDF $f_{\mathbf{h}}$. Every channel \mathbf{h} of any MT located within the BS cell is a realization of a random variable with PDF $f_{\mathbf{h}}$. For such a system model, the MSE-optimal channel estimator is given by the CME

$$\hat{\mathbf{h}} = \mathbb{E}[\mathbf{h} | \mathbf{y}] = \int \mathbf{h} f_{\mathbf{h}|\mathbf{y}}(\mathbf{h} | \mathbf{y}) d\mathbf{h}, \quad (2)$$

which can generally not be computed analytically. Further, $f_{\mathbf{h}}$ is typically not available in an analytic form.

However, in [4] a method to approximate (2) with the help of GMMs was proposed. To this end, assuming to have access to a set $\mathcal{H}_M = \{\mathbf{h}_m\}_{m=1}^M$ of training channel samples, which represent the radio propagation environment (ambient information), and motivated by universal approximation properties of GMMs [10], we fit a GMM $f_{\mathbf{h}}^{(K)}$ with K components to \mathcal{H}_M in order to approximate the unknown channel PDF $f_{\mathbf{h}}$.

A GMM is a PDF of the form [11]

$$f_{\mathbf{h}}^{(K)}(\mathbf{h}) = \sum_{k=1}^K p(k) \mathcal{N}_{\mathbb{C}}(\mathbf{h}; \boldsymbol{\mu}_k, \mathbf{C}_k), \quad (3)$$

where every summand is one of its K components. It is characterized by the means $\boldsymbol{\mu}_k \in \mathbb{C}^N$, the covariances $\mathbf{C}_k \in \mathbb{C}^{N \times N}$, and the mixing coefficients $p(k)$. Maximum likelihood estimates of these parameters can be computed using an expectation-maximization (EM) algorithm and the training data set \mathcal{H}_M , see [11].

The idea in [4] is to compute the MSE-optimal estimator $\hat{\mathbf{h}}_{\text{GMM}}^{(K)}$ for channels distributed according to $f_{\mathbf{h}}^{(K)}$ and to use it to estimate the channels distributed according to $f_{\mathbf{h}}$. This estimator $\hat{\mathbf{h}}_{\text{GMM}}^{(K)}$ converges pointwise to the optimal estimator $\hat{\mathbf{h}}$ from (2) as $K \rightarrow \infty$, cf. [4].

Once the (offline) GMM fitting process is done, the (online) channel estimates can be computed in closed form:

$$\hat{\mathbf{h}}_{\text{GMM}}^{(K)}(\mathbf{y}) = \sum_{k=1}^K p(k | \mathbf{y}) \hat{\mathbf{h}}_{\text{LMMSE},k}(\mathbf{y}), \quad (4)$$

with the responsibilities

$$p(k | \mathbf{y}) = \frac{p(k) \mathcal{N}_{\mathbb{C}}(\mathbf{y}; \boldsymbol{\mu}_k, \mathbf{C}_k + \mathbf{C}_n)}{\sum_{i=1}^K p(i) \mathcal{N}_{\mathbb{C}}(\mathbf{y}; \boldsymbol{\mu}_i, \mathbf{C}_i + \mathbf{C}_n)}, \quad (5)$$

and

$$\hat{\mathbf{h}}_{\text{LMMSE},k}(\mathbf{y}) = \mathbf{C}_k (\mathbf{C}_k + \mathbf{C}_n)^{-1} (\mathbf{y} - \boldsymbol{\mu}_k) + \boldsymbol{\mu}_k. \quad (6)$$

The weights $p(k | \mathbf{y})$ are the probabilities that component k generated the current observation \mathbf{y} , cf. [4].

A. Complexity Analysis and Low Cost Adaptations

The inverse in (6) can be precomputed offline for various signal-to-noise ratios (SNRs) because the GMM covariance matrices \mathbf{C}_k do not change once the GMM fitting process is done. Accordingly, evaluating (6) online is dominated by matrix-vector multiplications and has a complexity of $\mathcal{O}(N^2)$. It remains to calculate the responsibilities (5) by evaluating Gaussian densities. A Gaussian density with mean $\boldsymbol{\mu} \in \mathbb{C}^N$ and covariance matrix $\mathbf{C} \in \mathbb{C}^{N \times N}$ can be written as

$$\mathcal{N}_{\mathbb{C}}(\mathbf{h}; \boldsymbol{\mu}, \mathbf{C}) = \frac{\exp(-(\mathbf{h} - \boldsymbol{\mu})^H \mathbf{C}^{-1} (\mathbf{h} - \boldsymbol{\mu}))}{\pi^N \det(\mathbf{C})}. \quad (7)$$

Again, since the GMM covariance matrices and mean vectors do not change between observations, the inverses and the determinants of the densities can be pre-computed offline. Therefore, the online evaluation is also in this case dominated by matrix-vector multiplications and has a complexity of $\mathcal{O}(N^2)$. Overall, evaluating (4) has a complexity of $\mathcal{O}(KN^2)$ [4]. Since $\hat{\mathbf{h}}_{\text{GMM}}^{(K)}$ converges pointwise to the MSE-optimal estimator $\hat{\mathbf{h}}$ from (2) as $K \rightarrow \infty$, a trade-off between the performance of the estimator and the complexity can be achieved by adjusting the number K of GMM components.

The complexity of the estimator from (4) can be reduced by introducing structural constraints to the GMM covariance matrices \mathbf{C}_k . For example, in case of a uniform linear array (ULA) employed at the BS, it is common to assume Toeplitz covariance matrices, see, e.g., [12]. Further, for large numbers of antenna elements, a Toeplitz matrix is well approximated by a circulant matrix [13]. Motivated by these common assumptions we enforce structural constraints onto the GMM covariances. Since we consider exclusively an environment, where at the BS a uniform rectangular array (URA) with N_v vertical and N_h horizontal ($N = N_v \times N_h$) elements is employed, the structural assumptions result in block-Toeplitz matrices with Toeplitz blocks, or block-circulant matrices with circulant blocks, respectively [14]. In general the structured covariances can be expressed as

$$\mathbf{C}_k = \mathbf{Q}^H \text{diag}(\mathbf{c}_k) \mathbf{Q}, \quad (8)$$

where on the one hand, when assuming a Toeplitz structure, $\mathbf{Q} = \mathbf{Q}_{N_v} \otimes \mathbf{Q}_{N_h}$, where \mathbf{Q}_J contains the first J columns of a $2J \times 2J$ discrete Fourier transform (DFT) matrix, and $\mathbf{c}_k \in \mathbb{R}_+^{4N_h N_v}$ [14], [15]. On the other hand, when assuming circular structure, we have $\mathbf{Q} = \mathbf{F}_{N_v} \otimes \mathbf{F}_{N_h}$, where \mathbf{F}_J is the $J \times J$ DFT-matrix, and $\mathbf{c}_k \in \mathbb{R}_+^{N_h N_v}$. In both cases, the structural constraints allow to store only the \mathbf{c}_k 's of a GMM

which drastically reduces the memory overhead and the number of parameters to be learned, similar as in [4], [16].

Further, in case of circular covariances, the complexity of evaluating (4) reduces to $\mathcal{O}(KN \log(N))$, where 2D-DFT transforms are exploited when evaluating (5) and (6), cf. [4].

III. MEASUREMENT CAMPAIGN AND SYNTHETIC DATA

The measurement campaign was conducted at the Nokia campus in Stuttgart, Germany, in October/November 2017. As can be seen in Fig. 1, the receive antenna with a down-tilt of 10° was mounted on a rooftop about 20 m above the ground and comprises a URA with $N_v = 4$ vertical and $N_h = 16$ horizontal single polarized patch antennas. The horizontal spacing is λ and the vertical spacing equals $\lambda/2$, where the geometry of the BS antenna array was adapted to the urban microcell (UMi) propagation scenario, which exhibited a larger horizontal than vertical angular spread. The carrier frequency is 2.18 GHz. The BS transmitted time-frequency orthogonal pilots using 10 MHz OFDM waveforms. In particular, 600 sub-carriers with 15 kHz spacing were used, which resembles typical Long Term Evolution (LTE) numerology. The pilots were sent continuously with a periodicity of 0.5 ms and were arranged in 50 separate subbands, with 12 consecutive subcarriers each, for channel sounding purposes. For the duration of one pilot burst the propagation channel was assumed to remain constant.

The single monopole receive antenna, which mimics the MT, was mounted on top of a moving vehicle at a height of 1.5 m. The maximum speed was 25 kmph. Synchronization between the transmitter and receiver was achieved using GPS. The data was collected by a TSMW receiver and stored on a Rohde & Schwarz IQR hard disk recorder.

In a post-processing step, by the correlation of the received signal with the pilot sequence a channel realization vector with $N = N_v \times N_h$ coefficients per subband was extracted. The measurement was conducted at a high SNR, which ranged from 20 dB to 30 dB. Thus, the measured channels are regarded as ground truth. Further, we assume fully calibrated antennas and thus channel reciprocity is assumed. In this work, we will therefore consider a system, where we artificially corrupt the measured channels with AWGN at specific SNRs and thereby obtain noisy observations $\mathbf{y} = \mathbf{h} + \mathbf{n}$. The task is then to denoise the observations and obtain an estimated channel $\hat{\mathbf{h}}$. We want to highlight that we investigate a single-snapshot scenario, i.e., the coherence interval of the covariance matrix and of the channel is identical (the channel covariance matrix changes at the same time scale as the channel).

A. Synthetic Data Generation using QuaDRiGa

Version 2.6.1 of the QuaDRiGa channel simulator [5], [17] was used to generate channel state information (CSI) in a UMi scenario. The environment for which synthetic data is generated was adapted as closely as possible to the circumstances of the measurement environment. For this reason, the carrier frequency was set to 2.18 GHz. The base station is placed at a height of 20 meters. The minimum and maximum distances between MTs and the BS are 35 m and 315 m, respectively. The

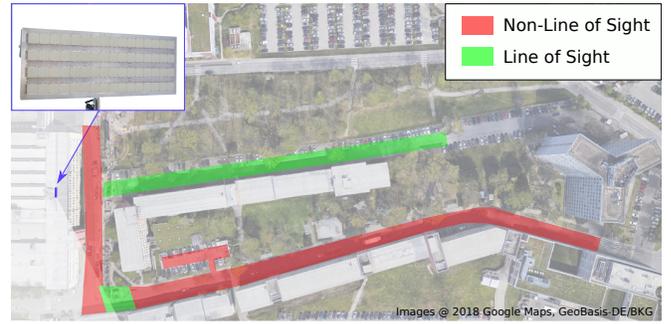


Fig. 1: Measurement setup on the Nokia campus in Stuttgart, Germany.

MTs are located outdoors at a height of 1.5 m. A QuaDRiGa channel is given by $\mathbf{h} = \sum_{\ell=1}^L \mathbf{g}_\ell e^{-2\pi j f_c \tau_\ell}$ with ℓ the path number, L the number of multi-path components (MPCs), f_c the carrier frequency, and τ_ℓ the ℓ th path delay. The number L depends on whether there is line of sight (LOS) or non-line of sight (NLOS) propagation, cf. [17]. The coefficients vector \mathbf{g}_ℓ consists of one complex entry for each antenna pair and comprises the attenuation of a path, the antenna radiation pattern weighting, and the polarization. As described in the QuaDRiGa manual [17], the generated channels are post-processed to remove the path gain.

IV. EXPERIMENTS AND RESULTS

Normalizing the data so that $\mathbb{E}[\|\mathbf{h}\|^2] = N$ allows us to define an SNR in our simulations as $\frac{1}{\sigma^2}$. Given T test samples $\{\mathbf{h}_t\}_{t=1}^T$ and obtaining corresponding channel estimates $\{\hat{\mathbf{h}}_t\}_{t=1}^T$, we use the normalized MSE $\frac{1}{NT} \sum_{t=1}^T \|\mathbf{h}_t - \hat{\mathbf{h}}_t\|^2$ as performance measure.

The following baseline estimators are considered. In the system at hand, the least squares estimate is simply given by the noisy observations $\hat{\mathbf{h}}_{\text{LS}} = \mathbf{y}$. Another baseline is the sample covariance matrix based approach, where we construct a sample covariance matrix $\mathbf{C}_s = \frac{1}{M} \sum_{m=1}^M \mathbf{h}_m \mathbf{h}_m^H$ given a set of training samples and calculate linear minimum mean square error (LMMSE) channel estimates $\hat{\mathbf{h}}_{\text{s-cov}} = \mathbf{C}_s (\mathbf{C}_s + \mathbf{C}_n)^{-1} \mathbf{y}$.

Compressive sensing approaches commonly assume that the channel exhibits a certain structure: $\mathbf{h} \approx \mathbf{D}\mathbf{t}$, where $\mathbf{D} \in \mathbb{C}^{N \times L}$ is a dictionary. We used an oversampled DFT matrix as dictionary, with $L = 4N$ (cf., e.g., [18]). A compressive sensing algorithm like orthogonal matching pursuit (OMP) [19] can then be used to obtain a sparse vector \mathbf{t} , and the estimated channel is calculated as $\hat{\mathbf{h}}_{\text{OMP}} = \mathbf{D}\mathbf{t}$. Since the sparsity order of the channel is not known but the algorithm's performance crucially depends on it, we use a genie-aided approach to obtain a bound on the performance of the algorithm. In particular, we use the true channel to choose the optimal sparsity order.

We further compare to a convolutional neural network (CNN)-based channel estimator, which was introduced in [12]. There, the authors exploit assumptions which stem from a spatial channel model (3GPP, cf. [20]) in order to derive the CNN architecture. The CNN is then trained on measurement

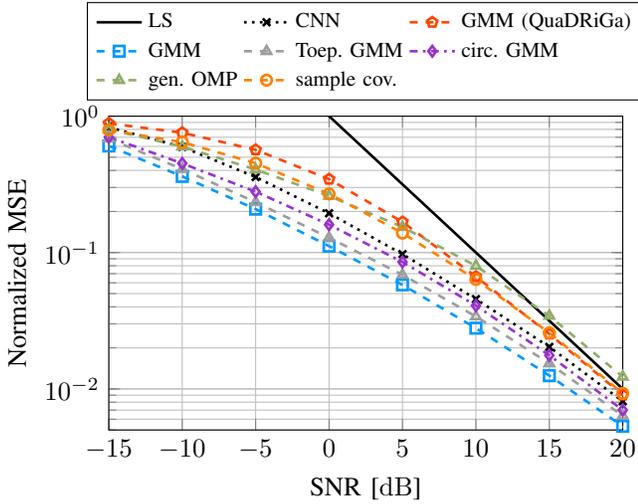


Fig. 2: Normalized MSE for various estimators over the SNR (evaluated on measurement data, with $T = 10,000$ samples). Each GMM approach is constructed using $K = 64$ components.

data to compensate the mismatch of the assumptions and the real world data. We use the rectified linear unit as activation function and the input transform is based on the $2N \times 2N$ DFT matrix, cf. [12, Equation (43)].

In Fig. 2, we use $T = 10,000$ channel samples stemming from the measurement campaign for evaluating the performances of the different channel estimators. In particular, we compare the GMM estimator with full covariances, denoted by “GMM”, and block Toeplitz (“Toep. GMM”) or block circulant (“circ. GMM”) covariances, to the state-of-the-art estimators described above. In total we use $M = 300,000$ channel samples stemming from the measurement campaign as training data in the fitting process of the GMM approaches, each with $K = 64$ components. Also the learning process of the CNN estimator (“CNN”) and the construction of a sample covariance matrix for the sample covariance LMMSE estimation approach (“sample cov.”) use these M samples.

The GMM estimator with full covariance matrices performs best over the whole SNR range from -15 dB to 20 dB, followed by the “Toep. GMM” and the “circ. GMM” approaches. As expected, the GMM estimator’s performance suffers from introducing structural covariance constraints but it still outperforms the other channel estimation approaches. With “GMM (QuaDRiGa)” we depict the GMM approach where synthetic training data ($M = 300,000$ samples) is used to fit the GMM. Despite using synthetic data of an environment, which was adapted as closely as possible to the circumstances of the measurement campaign’s environment, we can observe a severe performance degradation in the estimation performance of the GMM estimator.

In Fig. 3, we replace the test data and use $T = 10,000$ synthetic channel samples for comparing the “GMM” estimator (fitted with measurement data), the “CNN” approach (trained on measurement data), the “sample cov.” approach (sample

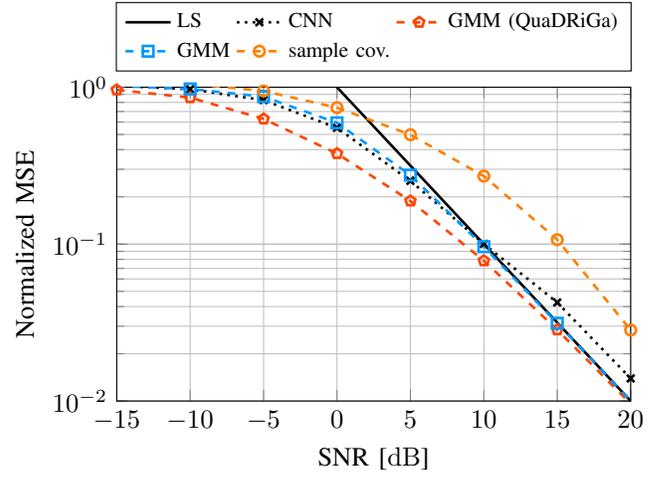


Fig. 3: Normalized MSE for various estimators over the SNR (evaluated on synthetic data, with $T = 10,000$ samples). Each GMM approach is constructed using $K = 64$ components.

covariance obtained using measurement data) with the “GMM (QuaDRiGa)” (fitted with synthetic data) approach. We can observe that the “GMM (QuaDRiGa)” approach now performs best since the learned ambient information now matches the synthetic test data on which the estimator is evaluated. We conclude that using synthetic channel data is not suitable to replicate the ambient information of the campus where the measurement campaign was conducted, and vice versa. Thus, this validates the claim that ambient information is learnt by the GMM when provided suitable training data. We further want to highlight that the CNN estimator which constitutes a data based approach as well, should also be able to capture the underlying ambient information of the considered propagation environment. Up to some extent this seems to be the case since the CNN approach exhibits the best performance right after the GMM approaches. Nevertheless, even the GMM approach with block circulant covariances, which has the same order of complexity as the CNN approach, yields a better overall estimation performance.

In Fig. 4 we consider the same simulation parameters as in Fig. 2, and analyze a performance upper bound for the achievable spectral efficiency

$$\bar{r} = E \left[\log_2 \left(1 + \frac{|\hat{\mathbf{h}}^H \mathbf{h}|^2}{\sigma^2 \|\hat{\mathbf{h}}\|^2} \right) \right], \quad (9)$$

when applying a matched filter $\frac{\hat{\mathbf{h}}^H}{\|\hat{\mathbf{h}}\|}$ in the uplink [12], which may also be interpreted as a measure of the accuracy of the estimated channel subspace [12]. Note, that there is no one-by-one relation between the spectral efficiency and the MSE in general, cf. [9]. In essence, we can observe that the GMM yields the best performance while there is only a minor degradation when using structured covariances.

In Fig. 5, we depict the behavior of the GMM estimator with full covariances for a varying number of components K and a varying number of training data used to fit the GMM.

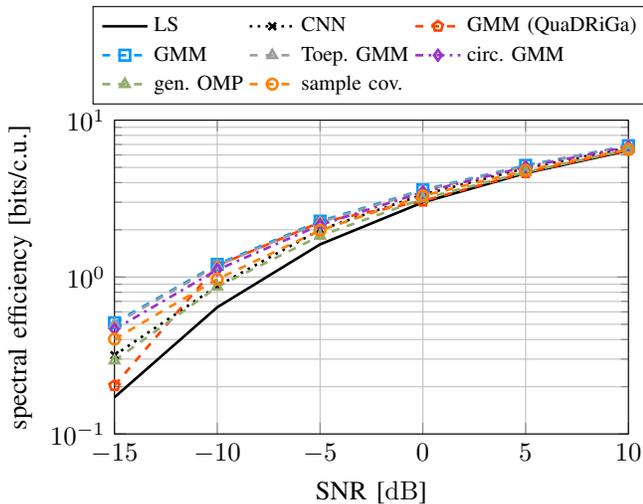


Fig. 4: Achievable spectral efficiencies using various estimators over the SNR (evaluated on measurement data, with $T = 10,000$ samples). Each GMM approach is constructed using $K = 64$ components.

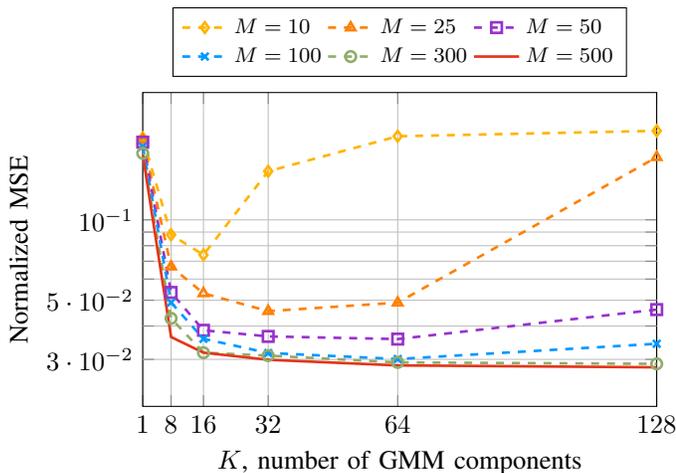


Fig. 5: Normalized MSE of the GMM estimator with full covariance matrices over the number of components K (evaluated on measurement data, with $T = 10,000$ samples). The GMM estimator is fitted using $M \cdot 10^3$ samples. The SNR is 10 dB.

The SNR is 10 dB. The number of parameters of a GMM increases with an increasing K , which requires more training data to achieve a good fit. As the figure suggests, as long as there are enough training data, increasing K leads to a better performance.

In contrast, in Fig. 6, where we consider GMMs with fewer parameters by assuming either block Toeplitz (top) or block circulant (bottom) covariances, we can observe that the estimator already achieves a good performance with a low to moderate number of available training samples. In particular, with structured covariances, increasing K leads to a better performance already with more than 100,000 available training samples. Overall, in Fig. 5 and Fig. 6, even a small to moderate

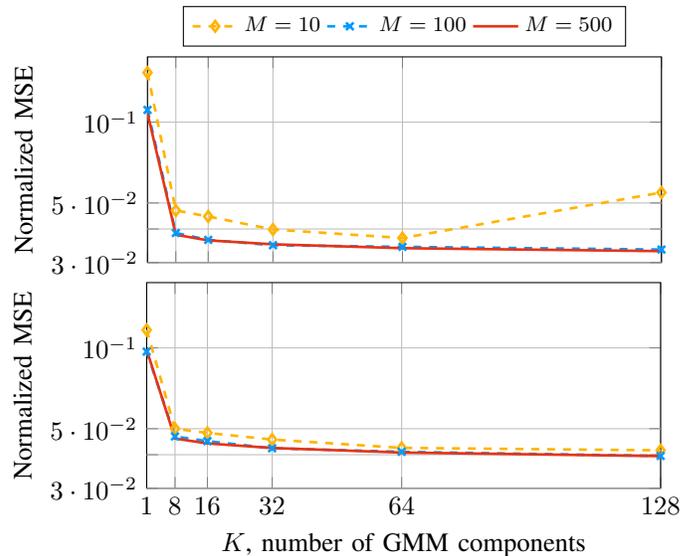


Fig. 6: Normalized MSE of the GMM estimator with block Toeplitz (top) or block circulant (bottom) covariances over the number of components K (evaluated on measurement data, with $T = 10,000$ samples). The respective GMM estimator is fitted using $M \cdot 10^3$ samples. The SNR is 10 dB.

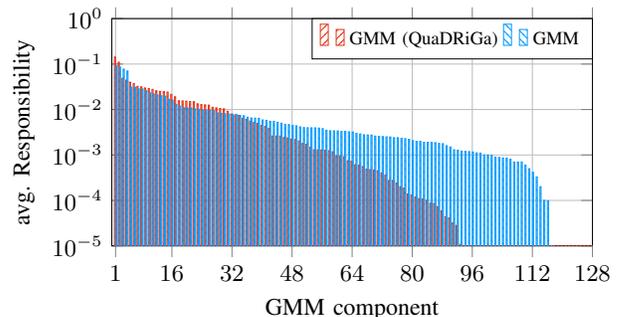


Fig. 7: Average responsibilities of the GMM components (evaluated on measurement data, with $T = 10,000$ samples). The SNR is 10 dB.

number of components ($K = 16$ or $K = 32$) performs well. Accordingly, a suitable number of components K should be determined based on the amount of available training data and the desired overall estimation complexity.

In Fig. 7 and Fig. 8, we aim to investigate the differences in the distributions of the synthetic and the measurement data from a different perspective: We plot the average responsibilities, cf. (5), of the GMM when fitted on either synthetic (“GMM (QuaDRiGa)”) or on measurement (“GMM”) data. To this end, we evaluate $p(k | \mathbf{y})$ for each observation \mathbf{y} and average these responsibilities with respect to the samples in the evaluation set with $T = 10,000$ samples and an SNR of 10 dB. Afterwards, we sort the components from highest to lowest average responsibility. In Fig. 7, the evaluation set

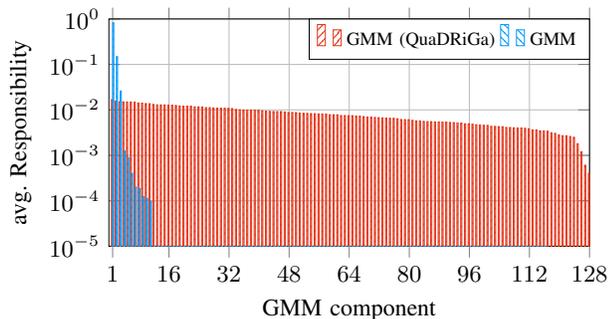


Fig. 8: Average responsibilities of the GMM components (evaluated on synthetic data, with $T = 10,000$ samples). The SNR is 10 dB.

contains only measurement data. It can be observed that the average responsibilities of the GMM fitted on synthetic or on measurement data are similar up to some extent. That is, the channel simulator is able to capture general information about the underlying UMi scenario—but not the details of the measurement environment in its full extent. In contrast, in Fig. 8, the same GMMs are evaluated with synthetic data. The average responsibilities can be clearly distinguished since a large mismatch can be observed. A possible explanation for this observation is that the GMM fitted onto the measurement data is specifically designed for the environment of the measurement campaign with unique immanent characteristics. In contrast to the channel simulator with a stochastic nature (hence aiming to model general UMi scenarios), this GMM does not generalize to different UMi scenarios. This behavior is desirable since the distinctive design allows for performance gains as discussed earlier. Fig. 7 and Fig. 8 show above all that the attempt to represent channel data with the false GMM only works to a limited extent, which can be seen, among other performance metrics, from the fact that fewer components of the false GMM are identified as representative than would be the case with the correct GMM. This underlines the importance of an evaluation with measurement data.

V. CONCLUSION AND OUTLOOK

In this work, we used real-world data stemming from a measurement campaign in order to evaluate and validate a recently introduced GMM-based algorithm for uplink channel estimation. Our experiments suggest that the GMM estimator learns the intrinsic characteristics of a given base station’s whole radio propagation environment. To validate the claim that ambient information is learnt, we conducted experiments, where we used test data either stemming from the measurement campaign or synthetic data. We observed that providing suitable ambient information, which is implicitly contained within the data, in the training phase (offline), beneficially impacts the channel estimation performance in the online phase. We further showed that structurally constrained covariances of the GMM, which are motivated by model-based insights, also work well when using real-world data. In particular, one can drastically

reduce the computational complexity and memory overhead with only small performance losses. An immediate additional advantage is that less training data is needed due to the lower number of GMM parameters, which need to be fitted, when assuming structural constraints. Future work might consider a more accurate and involved emulation of the propagation environment using a digital twin. For example, a digital representative of the propagation environment can be generated using a ray tracing tool, where the measurement campus with all of the buildings and streets, which are characteristic for certain propagation properties, is recreated virtually. Given the digital twin of the propagation environment, the performance of the data based channel estimators might be evaluated under these more accurate digital representatives.

REFERENCES

- [1] H. Ye, G. Y. Li, and B.-H. Juang, “Power of deep learning for channel estimation and signal detection in OFDM systems,” *IEEE Wireless Commun. Letters*, vol. 7, no. 1, pp. 114–117, 2018.
- [2] J. Guo, C.-K. Wen, M. Chen, and S. Jin, “Environment knowledge-aided massive MIMO feedback codebook enhancement using artificial intelligence,” *IEEE Trans. Commun.*, 2022.
- [3] M. Koller, B. Fesl, N. Turan, and W. Utschick, “An asymptotically optimal approximation of the conditional mean channel estimator based on Gaussian mixture models,” in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, 2022, pp. 5268–5272.
- [4] —, “An asymptotically MSE-optimal estimator based on Gaussian mixture models,” *IEEE Trans. Signal Process.*, pp. 1–14, 2022.
- [5] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, “Quadriga: A 3-d multi-cell channel model with time evolution for enabling virtual field trials,” *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, 2014.
- [6] A. Alkhateeb, “DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications,” in *Proc. of Inf. Theory and Appl. Workshop (ITA)*, San Diego, CA, Feb 2019.
- [7] Remcom, “Wireless InSite,” <http://www.remcom.com/wireless-insite>.
- [8] N. Turan, M. Koller, B. Fesl, S. Bazzi, W. Xu, and W. Utschick, “GMM-based codebook construction and feedback encoding in FDD systems,” 2022, arXiv preprint: 2205.12002.
- [9] C. Hellings, A. Dehmani, S. Wesemann, M. Koller, and W. Utschick, “Evaluation of neural-network-based channel estimators using measurement data,” in *Proc. Int. ITG Workshop on Smart Antennas (WSA)*, 2019, pp. 1–5.
- [10] T. T. Nguyen, H. D. Nguyen, F. Chamroukhi, and G. J. McLachlan, “Approximation by finite mixtures of continuous density functions that vanish at infinity,” *Cogent Math. Statist.*, vol. 7, no. 1, p. 1750861, 2020.
- [11] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [12] D. Neumann, T. Wiese, and W. Utschick, “Learning the MMSE channel estimator,” *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, 2018.
- [13] R. M. Gray, “Toeplitz and circulant matrices: A review,” *Found. and Trends® in Commun. and Inf. Theory*, no. 3, pp. 155–239, 2006.
- [14] M. Koller, C. Hellings, and W. Utschick, “Learning-based channel estimation for various antenna array configurations,” in *Proc. IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019, pp. 1–5.
- [15] G. Strang, “A proposal for Toeplitz matrix calculations,” *Stud. in Appl. Math.*, vol. 74, no. 2, pp. 171–176, 1986.
- [16] B. Fesl, M. Joham, S. Hu, M. Koller, N. Turan, and W. Utschick, “Channel estimation based on Gaussian mixture models with structured covariances,” 2022, arXiv preprint: 2205.03634.
- [17] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, “Quadriga: Quasi deterministic radio channel generator, user manual and documentation,” Fraunhofer Heinrich Hertz Institute, Tech. Rep., v2.2.0, 2019.

- [18] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *2015 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2015, pp. 2909–2913.
- [19] M. Gharavi-Alkhansari and T. Huang, "A fast orthogonal matching pursuit algorithm," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, vol. 3, 1998, pp. 1389–1392.
- [20] 3GPP, "Spatial channel model for multiple input multiple output (MIMO) simulations," 3rd Generation Partnership Project (3GPP), Tech. Rep. 25.996 (V16.0.0), Jul. 2020.