Estimating Multi-Modal Dense Multipath Components using Auto-Encoders

S. Schieler*, M. Döbereiner†, S. Semper*, M. Landmann†,
*Electronic Measurements and Signal Processing Research Group,
Technische Universität Ilmenau, Ilmenau, Germany, steffen.schieler@tu-ilmenau.de
†Fraunhofer Institute for Integrated Circuits, Ilmenau, Germany

Abstract—

We present a maximum-likelihood estimation algorithm for radio channel measurements exhibiting a mixture of independent Dense Multipath Components. The novelty of our approach is in the algorithms initialization using a deep learning architecture. Currently, available approaches can only deal with scenarios where a single mode is present. However, in measurements, two or more modes are often observed. This much more challenging multi-modal setting bears two important questions: How many modes are there, and how can we estimate those?

To this end, we propose a Neural Net-architecture that can reliably estimate the number of modes present in the data and also provide an initial assessment of their shape. These predictions are used to initialize for gradient- and model-based optimization algorithm to further refine the estimates.

We demonstrate numerically how the presented architecture performs on measurement data and analytically study its influence on the estimation of specular paths in a setting where the single-modal approach fails.

*Index Terms—***DMC, Channel Estimation, Parameter Estimation, Autoencoders, Deep Learning**

I. Introduction

Radio channels, especially their accurate description, have been a vividly studied topic for the past decades. Often, one seeks underlying parameters of the propagating waves in a measured environment. Usually, the assumption is that these waves travel as specular rays/paths, i.e., Specular Components (SC), through the environment in the form of plane waves. However, it has been acknowledged [1]–[4] that the the data model applied for High Resolution Parameter Estimation (HRPE) as a superposition of resolvable SC is not sufficient to completely account for the bandlimited data collected by the receiver. One possible physical interpretation of Dense Multipath Components (DMC) is as a superposition of a large number of SC which can not be resolved with bandwidth and Signal-to-Noise Ratio (SNR) of the measurement system.

The authors in [5] demonstrate that the mentioned model mismatch can be overcome by accounting for this remaining energy as a colored Gaussian noise. The underlying process has to be parametrized in a sophisticated manner such that it does capture the behavior of the non-resolvable components

sufficiently well, does not interfere with the estimation of resolvable specular paths while also the complexity of the stochastic model allows for an efficient HRPE process.

No matter if the process is modeled as both spatially and temporally correlated process [2], or just temporally correlated [1], [6], the previously proposed models usually account for single mixture models, i.e., one DMC mode, either in space- and delaytime or solely in delaytime. However, for several propagation scenarios in different bands and bandwidths, see for instance [7] and Figure 6, we observed that the DMC process must be modeled as a superposition of *several* independent modes, which are separated in delaytime and/or space. If this multitude of the DMC is neglected during the estimation process, the wrongful estimation of a single mode ultimately also deteriorates the HRPE of the SC due to this inherent model mismatch and subsequently biased estimation of SC.

In this paper, we focus on a multi-modal temporal distribution of the DMC and assume a spatially uncorrelated process, for which [1] presented a model-based maximum-likelihood approach, which is split in an initialization stage and an iterative refinement stage based on gradient descent. We give a summary of the algorithm in Section II-A. While the employed model itself can easily be extended to incorporate an arbitrary number of distinct modes, the initialization step implicitly assumes the presence of just one mode.

To allow for robust multi-modal estimation, we propose to use a so-called Auto-Encoder using convolutional networks [8], which jointly infers the number and shapes of the DMC modes directly from measured transfer-functions. Moreover, the proposed Neural Net (NN) can separate these, such that the initialization from [1] can be applied separately. Along the way, we also propose a numerically more suitable parametrization of the DMC.

II. Algorithm

We briefly review the setting for the currently available algorithms and explain their approach. A radio channel observation consisting of $N_f \in \mathbb{N}$ frequency samples and M uncorrelated snapshots can be modeled as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{n}(\boldsymbol{\delta}) + \boldsymbol{n} \in \mathbb{C}^{N_f \times M}, \quad (1)$$

where $\boldsymbol{f} : \Theta \subset \mathbb{R}^s \to \mathbb{C}^{N_f \times M}$ describes the resolvable specular components and $\boldsymbol{n} : \Delta \subset \mathbb{R}^d \to \mathbb{C}^{N_f \times M}$ is a complex zero-mean Gaussian but colored noise process we use to model
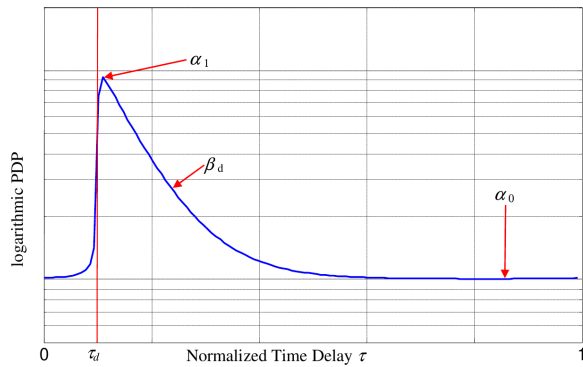
Fig. 1. Example for a single DMC mode and its parameters[1, p. 39]. Here, we have that $\boldsymbol{\delta} = [\alpha_1, \beta_d, \tau_d]$.

the DMC. Additionally, the columns $\boldsymbol{n}_i$ of $\boldsymbol{n}$ are modeled as independent zero-mean Gaussian random vectors with covariance $\alpha_0 \boldsymbol{I}_{N_f}$ representing the measurement noise.

If we already have a sufficiently good $\hat{\boldsymbol{\theta}}$ to describe the specular components, it is justified to consider only the residual

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{n}(\boldsymbol{\delta}) + \boldsymbol{n}. \qquad (2)$$

Parametrizing $\boldsymbol{n}(\boldsymbol{\delta}) \sim \mathcal{N}(\boldsymbol{0}_{N_f \times M}, \boldsymbol{\Sigma}(\boldsymbol{\delta}) \otimes \boldsymbol{I}_M)$, where $\otimes$ denotes the outer product, we see that the covariance of the zero-mean residual satisfies

$$\mathbf{E}\left[\boldsymbol{r}(\boldsymbol{\theta})^* \boldsymbol{r}(\boldsymbol{\theta})\right] = (\boldsymbol{\Sigma}(\boldsymbol{\delta}) + \alpha_0 \boldsymbol{I}_{N_f}) \otimes \boldsymbol{I}_M, \qquad (3)$$

where $\boldsymbol{z}^*$ denotes Hermitian transposition and $\boldsymbol{\theta}$ being the true parameter that is the origin of the observation $\boldsymbol{y}$.

A. State of the Art Algorithm

The core idea is to consider the negative log-likelihood of $\boldsymbol{y}$ with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}_0$, which essentially reads as

$$\lambda(\boldsymbol{r}, \boldsymbol{\delta}, \alpha_0) = \log \det(\boldsymbol{\Sigma}(\boldsymbol{\delta}) + \alpha_0 \boldsymbol{I}) + \mathrm{tr}(\boldsymbol{r}^* \boldsymbol{\Sigma}(\boldsymbol{\delta})^{-1} \boldsymbol{r}), \quad (4)$$

where we omitted constant summands, factors and the dependence on $\boldsymbol{\theta}$. If we now consider [1, eq. 2.61] we find the mapping for $\boldsymbol{\Sigma} : \mathbb{R}^3 \to C^{N_f \times N_f}$ to be a Toeplitz matrix-valued function defined as

$$\boldsymbol{\Sigma}(\boldsymbol{\delta})_{i,j} = \frac{\delta_1}{\delta_2 + \jmath 2\pi(f_i - f_j)} \cdot \exp(-\jmath 2\pi(f_i - f_j) \cdot \delta_3), \quad (5)$$

where f_i denotes the sampling frequencies used to obtain $\boldsymbol{y}$, where we renamed $\boldsymbol{\delta} = [\alpha_1, \beta_d, \tau_d]$. See Figure 1 for an interpretation of these.

a) Initialization: In order to find a good estimate, one approach [1, Sec. 6.1.8] is to first define $\hat{\boldsymbol{r}} = \sum_{i=1}^{M} |\boldsymbol{F}^* \boldsymbol{r}_i|^2$ using the Fourier matrix $\boldsymbol{F}$ and then to calculate

$$\hat{\alpha}_0 = \min \hat{\boldsymbol{r}}; \quad \hat{\delta}_1 = \max \hat{\boldsymbol{r}} - \hat{\alpha}_0$$

together with

$$\hat{\delta}_2 = \frac{\hat{\delta}_1}{N_f(\|\hat{\boldsymbol{r}}\|_1 - \hat{\alpha}_0)},$$

where $\|\cdot\|_1$ denotes the ℓ_1-norm. Finally, a simple yet usually sufficient estimate for δ_3 is given by

$$\hat{\delta}_3 = \frac{\mathrm{argmax}_i(\hat{\boldsymbol{r}}_{i+1} - \hat{\boldsymbol{r}}_i) - 1}{N_f - 1}$$

due to the shape of the assumed Power Delay Profile (PDP).

b) Refinement: Once a suitable $\hat{\boldsymbol{\eta}}^0 = [\hat{\boldsymbol{\delta}}^T, \hat{\alpha}_0]$ is found, we are free to perform any gradient-based algorithm. For example, for our simulation and analysis we are going to use

$$\boldsymbol{\eta}^{k+1} = \boldsymbol{\eta}^k + (\boldsymbol{H}(\boldsymbol{\eta}^k) + \mu_k \boldsymbol{I}_4)^{-1} \cdot \boldsymbol{J}(\boldsymbol{\eta}^k), \qquad (6)$$

where $\boldsymbol{H}(\boldsymbol{\eta}^k)$ is the negative Fisher Information Matrix (FIM) defined via λ and $\boldsymbol{J}(\boldsymbol{\eta}^k)$ is the score function of λ. Simply plugging (5) into (4) and using the definition for $\boldsymbol{H}$ and $\boldsymbol{J}$ allows to efficiently carry out the update step in (6).

B. The Limitations

If we consider the approach outlined above in the setting depicted in Figure 4, we see that the model mismatch in terms of number of DMC modes is detrimental. The only parameter correctly estimated is δ_3, while the other account for the existence of two modes instead of the assumed single mode. Not only is the estimate of $\boldsymbol{\Sigma}$ inherently biased, also any sensible estimation routine for the specular path parameters $\boldsymbol{\theta}$ is influenced by $\boldsymbol{\Sigma}$. For instance, for specular paths arriving at a normalized delay in $[0.3, 0.5]$, the SNR is estimated to be worse than it actually is, which will have an influence on how reliable one considers the estimated parameter of this path. In Section IV-A we provide a more detailed analysis of this phenomenon.

C. Proposed Algorithm

The modifications and extensions to the above state-of-the-art algorithm are three-fold.

a) Change of Variables: Originally in [1], α_0 and α_1 were defined on a linear scale, which forced any optimization with respect to these parameters to obey the side constraints $\alpha_0, \alpha_1 > 0$. However, these constraints can easily be alleviated by means of a change of variables as

$$\alpha_0^{\mathrm{prev}} = \exp(\alpha_0), \quad \alpha_1^{\mathrm{prev}} = \exp(\delta_1),$$

which not only renders the optimization over $\boldsymbol{\delta}$ an unconstrained problem, but also the first- and second-order derivatives are much better behaved in terms of the condition number of the Hessian matrix.

b) Extension of the Model: As a first step, we generalize the parametric model for $\boldsymbol{\Sigma}$ to the multi-modal version $\boldsymbol{\Sigma}_m : \mathbb{R}^{3 \times m} \to \mathbb{C}^{N_f \times N_f}$ via

$$\boldsymbol{\Sigma}_m(\boldsymbol{\Delta}) = \sum_{i=1}^{m} \boldsymbol{\Sigma}(\boldsymbol{\Delta}_i), \qquad (7)$$

which is a simple linear-combination of m covariances corresponding to the single-mode setting. This means implicitly that $\boldsymbol{\Delta}$ is the parameter of the multi-modal version of the DMC process. Hence, the quantities $\boldsymbol{H}$ and $\boldsymbol{J}$ used in (6) can straightforwardly be extended to this more general model. Consequently, the challenging task is robustly estimating the quantity m as the number of modes present in $\boldsymbol{y}$ and their correct initialization.

c) Initialization via an Auto-Encoder: When looking at a PDP like the one in Figure 4, it is intuitively clear how many modes are present and where they are located. Hence, we approach this learning problem like a supervised-learning, 1D
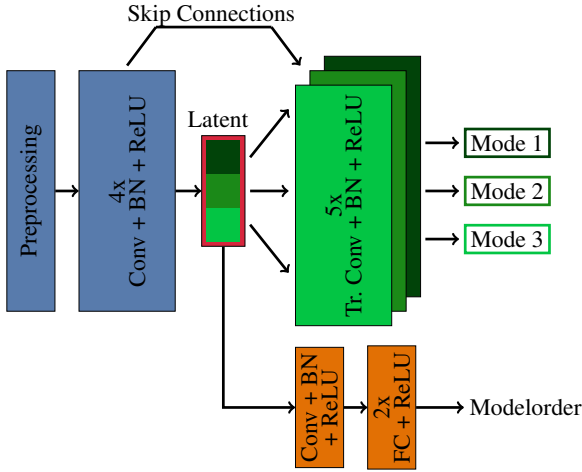
Fig. 2. The architecture of the neural network is a 1D adaptation of U-Net. Additional decoders are added to predict multiple DMC modes and a model-order subnet (orange) to predict the input model order based on the latent space also used for the reconstruction task.
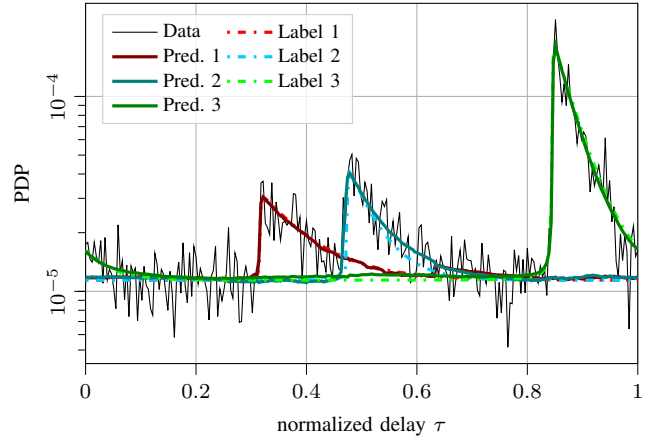

Fig. 3. Data, labels and predictions of the proposed learning algorithm applied to synthetic data. The predictions are processed individually afterwards to estimate $\mathbf{\Delta}$. Periodicity is an artifact of the DFT-based preprocessing.

imaging problem. To this end, we designed an Auto-Encoder neural network targeted to solve two different tasks:

1) Predict the correct model-order for the given data sample in the range $m = 0, \ldots, 3$. Note the maximum value of 3 is arbitrary and the architecture can be extended to predict more modes straightforwardly.
2) Recover, separate, and denoise up to $m = 3$ DMC modes present in $\mathbf{y}$, such that each component can be estimated separately with the already existing methods.

With the denoised and separated modes obtained from the decoders, the actual parameters $\mathbf{\delta}$ of each mode are estimated using a Levenberg-Marquardt-algorithm based estimator. In the following section, we outline the network architecture and the structure of the training data in order to accomplish these tasks.

III. Learning Architecture

On a very high level, the architecture of the neural network consists of an encoder, which downsamples the complex-baseband input data into a latent space. Then, up to three separate decoders are tasked with the reconstruction of the independent DMC modes, based on the latent space.

a) Labels and Data: As we consider using real measurement data to train our algorithm infeasible, because it is not practical to manually generate a critical amount of labeled data, we decided to use a synthetic dataset for the training by means of the model $\mathbf{\Sigma}$. This bears the advantage that we can generate an almost arbitrarily large dataset to train the network to obtain a well generalizing estimator.

Hence, our synthetic dataset consists of 10^8 randomly generated instances of $\mathbf{\Sigma}_m(\mathbf{\Delta}) + \alpha_0 \mathbf{I}_{N_f}$ and the corresponding random realizations $\mathbf{r} = \mathbf{n}(\mathbf{\Delta}) + \mathbf{n} \in \mathbb{C}^{N_f \times M}$ as input data, ensuring the network never actually sees the exact same sample twice, to strengthen generalization and robustness to noise.

As labels, we use the corresponding, separated DMC components denoted by $\mathbf{\Sigma}(\mathbf{\delta}_m)$. The relationship between input

data, labels and the resulting predictions for the supervised learning task is illustrated by Figure 3.

b) Preprocessing: Before being passed to our neural network, the input data $\mathbf{r}$ are preprocessed. The preprocessing involves applying an inverse DFT along the frequency dimension and using the mean of the magnitude-square

$$\mathbf{d} = \frac{\sqrt{N_f}}{M} \sum_{k=1}^{M} |\mathbf{F}^* \mathbf{r}_k|^2 \in \mathbb{C}^{N_f}, \tag{8}$$

where $\mathbf{F}^* \in \mathbb{C}^{N_f \times N_f}$ denotes the inverse DFT matrix and $\mathbf{r}_k$ the $k-$th snapshot, where the contribution of SCs has already been removed.

The second preprocessing step regards appropriate normalization of the input data $\mathbf{d}$ to limit the dynamic range of the input values fed into the non-linear neural network. This step is necessary, as dynamic ranges in real-world measurement data can vary over several orders of magnitude. To address this, we normalize the values in $\mathbf{d}$ to get the normalized version $\mathbf{d}_n$ by means of applying

$$\mathbf{d}_n = \log \mathbf{d} - \log \max(\mathbf{d}) \tag{9}$$

element-wise.

This normalization scheme is calculated based on the inputs only and enables reconstruction of correctly scaled predictions after forward propagation using stored values of $\max(\mathbf{d})$ and rescaling the forward propagation output accordingly.

c) Network Architecture: The general design of the applied neural network architecture is lent from U-Net [9]. As our task is based on 1-dimensional input data, the 2D convolutional layers in U-Net are replaced by appropriate 1-D convolutions. Before being passed to the encoder, $\mathbf{d}_n$ is passed through two convolutional layers (with batchnorm and ReLU activation function), which increases the number of channels from 1 to 32. In the encoder four downsampling blocks are used, each consisting of a convolutional layer followed by batch normalization and ReLU activation function. The number of filters is doubled by the convolution layers in each downsampling block, which are parametrized with a kernel size of 3, stride of 2 (to achieve the downsampling), and circular padding of 1.
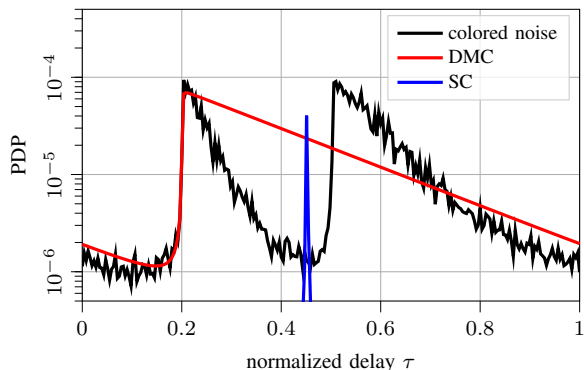
Fig. 4. Setup for the simulations of the CRB with two dense components. Initialization results for a single mode are shown in red. The simulated accuracy results are given in Figure 5.
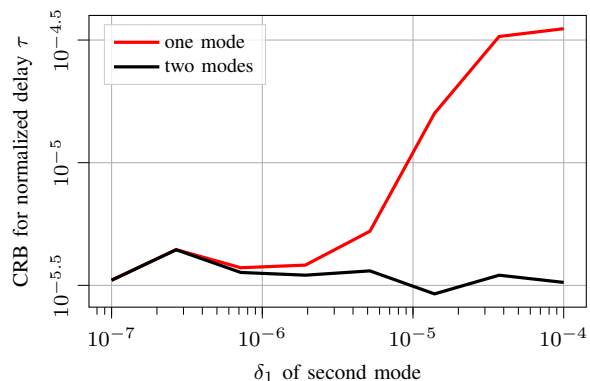

Fig. 5. CRB of the delay of specular component shown in Figure 4 (blue) for one and two DMC modes. If the estimator assumes only one mode present (red), increasing the energy in the second mode (δ_1), degrades the estimators accuracy. Using two modes (black), the energy of the second mode can be increased without noteworthy changes in accuracy.

After the downsampling, the latent space has a dimension of 16 features (with 512 channels). A convolutional layer is used to upscale the latent to 24 features, enabling passing 8 features to each of the three decoders. The decoder consists of three individual, structural identical, decoders (as we attempt to reconstruct $M \leq 3$ modes from each $\boldsymbol{d}_n$), each with 5 upsampling blocks. Each upsampling block consists of a transposed convolution, followed by two convolutional layers. Skip Connections between the decoder and encoders are added to the all but the first upsampling block after each transposed convolution layer. Every decoder receives a third of the latent space to reconstruct its respective target mode.

The model order is prediction is performed by a small subnet attached to the latent space. A convolutional layer is followed by three fully-connected layers to predict a one-hot encoded integer number between 0 to 3.

d) Loss and Training: To ensure proper convergence of the neural network weights the used loss function is a weighted sum of two different components.

1) the mode reconstruction loss uses mean squared error (MSE)
2) the modelorder loss uses Binary Crossentropy (BCE)

For the computation of the mode reconstruction loss, the predicted model-order $\widetilde{m}$ is taken into account to mask mode predictions of non-existent modes in the decoders. For example, if the predicted model-order is 2, the mode prediction of the third decoder is excluded from the computation of the MSE loss. This results in the following loss function $l : \mathbb{R}^{N_f \times 3} \times \mathbb{R}^{N_f \times 3} \times \mathbb{N} \times \mathbb{N} \to \mathbb{R}_0^+$ defined as

$$l(\boldsymbol{x}, \tilde{\boldsymbol{x}}, m, \widetilde{m}) = w_x \sum_{k=0}^{\widetilde{m}} \|\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_k\|_2^2 \qquad (10)$$
$$+ w_m \left[\widetilde{m} \log(m) + (1 - \widetilde{m}) \log(1 - m) \right]$$

where $\boldsymbol{x}$, $\tilde{\boldsymbol{x}}$ denote the label and prediction for the modes from the decoders, m and $\tilde{m}$ denote the label and prediction for the model-order, and w_x and w_m denote scaling weights for the two different prediction parts. The loss weights were fixed to $w_x = 1$, $w_m = 100$, based on observations during training, such that both parts have similar magnitude.

IV. SIMULATIONS

A. Influence of the DMC on the Estimation Accuracy

To showcase the importance of the correct DMC model order for the estimation process, we want to study the resulting best-case accuracy any unbiased estimator can achieve by means of the CRB. To this end, we create a simple synthetic setup, where the simulated Impulse Response (IR) contains a single specular component with fixed path weights and fixed normalized delay $\tau = 0.45$ is enclosed by two DMC modes as depicted in Figure 4. The analysis of the CRB is not only beneficial for theoretic purposes. The estimator proposed in [1] also uses the inverse FIM to determine which specular paths are reasonable estimates and possibly removes specific specular paths if the expected estimation variance indicated by the CRB is too high.

We simulate two different cases. First, the estimation of a single mode, where one mode is spanning the area of the two modes, to account for both at the same time. Second, we carry out the proposed method, which should correctly detect the two modes. We run this scenario for varying intensity of the second mode by means of varying δ_1 of this DMC, i.e. $\Delta_{2,1}$. Then, we evaluate the deterministic CRB for a peak at the depicted position, and we average these values over 200 realizations per level of δ_1 of the second mode.

As we can see in Figure 5, the resulting estimation accuracy differs significantly for the two settings. For higher intensities of δ_1, the scenario increasingly resembles the one depicted in Figure 4. Hence, the assumed colored noise distribution renders the specular path harder to distinguish from random fluctuations in the measurement. However, if we correctly impose a model with two modes, the intensity of the modes does not significantly influence the predicted accuracy for the specular paths delay.

B. Joint Estimation of DMC and SC

To also apply the proposed DMC estimation routine to a more realistic setup, we used measurement data collected in a channel-sounding campaign, which clearly shows at least two
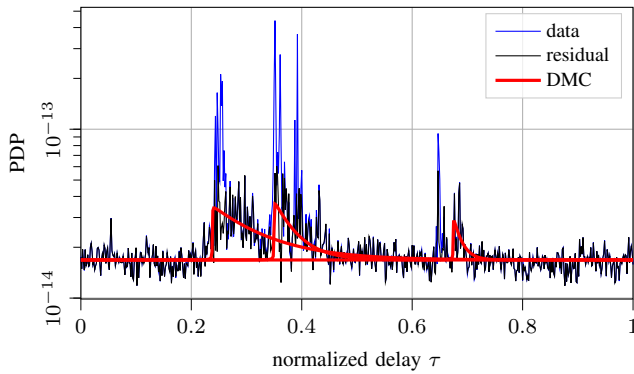
Fig. 6. Estimation of three DMC modes in measurement data [7] from a channel sounding campaign. The DMC are estimated from the residual using the proposed neural network previously trained on synthetic data.

DMC modes in some of the snapshots [7]. However, the data still contains specular components that need to be removed from the observation, such that the assumption in (2) is valid enough such that the PDP is not biased by the remaining specular components. Note, in practice the estimation of both the SC and the DMC is heavily intertwined, since the estimates of θ and δ are influencing each other. This effect is also visible in the presented data. The mode starting at normalized delay $\tau \approx 0.7$ might not be estimated, if more specular components had been accounted for during the estimation prioir to the DMC estimation.

In Figure 6 we plot the acquired data y, the residual $r(\theta)$ and 3 estimated DMC modes. We used an algorithm similar to the one presented in [1] to get an estimate for θ and hence $r(\theta)$. To estimate δ, we used the model-order selection provided by the proposed NN and then the initialization and refinement explained in Section II-A. The first thing we notice is that the prediction and estimation work on measurement data, which is not to be taken for granted since the architecture was trained on synthetic data without any specular components. Second, the results indicate that the proposed architecture is also able to estimate overlapping modes quite robustly. However, depending on the use-case of the estimation data, it could be debatable if the third mode is necessary to be estimated.

V. CONCLUSION

Proper estimation of multiple DMC is required to obtain an unbiased estimate of both the dense as well as the specular components. We show that the introduced autoencoder-based neural net design allows estimating the separated DMC modes together with model-order. We show that although the network is trained on an artificially generated, synthetic dataset, the results suggest that it can also be applied to measurement data. Our results show, the model-based neural network approach is suitable to denoise and separate up to three DMC and suitable for subsequent estimator initialization, surpassing other state-of-the-art approaches.

To further extend the current approach, two different directions seem the most useful for the problem at hand, i.e., direct estimation of DMC parameters and extension to the angular

domain. First, including the angular domain in the estimates provides a valuable framework extension for use cases in antenna array applications. With additional angular information of the DMC better separation between dense components can likely be achieved as the individual DMC can also be separated with regard to their respective directivity angle in the array. Second, to further improve the network performance on measurement data, it is also possible to tune the parameter distribution in the training set to better match the characteristics of the specific measurement dataset. This could be accomplished by utilizing a Generative Adversarial Network (GAN) to generate training data which is even more similar to the measurement data encountered in the inference task.

REFERENCES

[1] A. Richter, "Estimation of Radio Channel Parameters," 2005. [Online]. Available: https://www.db-thueringen.de/receive/dbt_mods_00004815.

[2] M. Käske and R. Thomä, "Maximum-likelihood based estimation of angular parameters of Dense-Multipath-Components," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, 2015.

[3] J. Poutanen, J. Salmi, K. Haneda, *et al.*, "Angular and shadowing characteristics of dense multipath components in indoor radio channels," *IEEE Transactions on Antennas and Propagation*, no. 1, 2011. DOI: 10.1109/TAP.2010.2090474.

[4] E. M. Vitucci, F. Mani, V. Degli-Esposti, *et al.*, "Polarimetric properties of diffuse scattering from building walls: Experimental parameterization of a ray-tracing model," *IEEE Transactions on Antennas and Propagation*, no. 6, 2012. DOI: 10.1109/TAP.2012.2194683.

[5] M. Landmann, M. Käske, and R. S. Thomä, "Impact of Incomplete and Inaccurate Data Models on High Resolution Parameter Estimation in Multidimensional Channel Sounding," *IEEE Transactions on Antennas and Propagation*, no. 2, 2012. DOI: 10.1109/TAP.2011.2173446.

[6] C. Oestges, N. Czink, P. De Doncker, *et al.*, "Radio channel modeling for 4g networks," in *Pervasive Mobile and Ambient Wireless Communications: COST Action 2100*, R. Verdone and A. Zanella, Eds. Springer London, 2012.

[7] C. Schneider, J. Breuer, J. Beyer, *et al.*, "Characterization of urban radio channels and base station antenna correlation in the 3.75 ghz band," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021. DOI: 10.1109/VTC2021-Spring51267.2021.9448718.

[8] Y. LeCun, B. Boser, J. S. Denker, *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, no. 4, 1989. DOI: 10.1162/neco.1989.1.4.541.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," en, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science, Springer International Publishing, 2015. DOI: 10.1007/978-3-319-24574-4_28.