Aalborg Universitet

Maximum likelihood calibration of stochastic multipath radio channel models

Hirsch, Christian Pascal; Bharti, Ayush; Pedersen, Troels; Waagepetersen, Rasmus

Maximum likelihood calibration of stochastic multipath radio channel models

Christian Hirsch, Ayush Bharti, Troels Pedersen, and Rasmus Waagepetersen

Abstract—We propose Monte Carlo maximum likelihood estimation as a novel approach in the context of calibration and selection of stochastic channel models. First, considering a Turin channel model with inhomogeneous arrival rate as a prototypical example, we explain how the general statistical methodology is adapted and refined for the specific requirements and challenges of stochastic multipath channel models. Then, we illustrate the advantages and pitfalls of the method on the basis of simulated data. Finally, we apply our calibration method to wideband signal data from indoor channels.

Index Terms—multipath channels, Monte Carlo methods, maximum likelihood estimation, point processes, radio propagation, shot noise.

I. Introduction

Stochastic multipath models are indispensable for simulating and analyzing radio systems for communication and localization. In a stochastic multipath model, the received signal is modeled as a superposition of attenuated and delayed signal components, each corresponding to one propagation path [1]. Such a model can be described by a marked point process where a marked point represents a delay and its associated path gain. Provided that the model is calibrated, i.e. its parameters have been estimated from measurement data, realizations of the channel can then be simulated from the model and used in system design or performance analysis, thus alleviating the need for further measurements. Calibration of stochastic multipath models is a non-trivial task for several reasons. In particular, due to the finite measurement bandwidth and the presence of additive noise, the marked point process is not observed directly, but should be considered as a hidden variable. Not least in the context of point processes, estimating parameters in models with hidden variables is often a highly involved endeavor [2].

The calibration approach most widely used in the literature is a two-step procedure dating back to Turin [3], outlined in Fig. 1. First, the measurement data is reduced to a set of multipath components, such as delays and path gains. Then, the parameters of the underlying point process are estimated from the obtained multipath components. Although this data reduction step was employed chiefly due to technical limitations of the measurement equipment and data processing used by Turin at that time, many works have since adopted

Christian Hirsch is with the Bernoulli Institute, University of Groningen, Netherlands (e-mail: c.p.hirsch@rug.nl).

Ayush Bharti and Troels Pedersen are with the Department of Electronic Systems, Aalborg University, Denmark (e-mail: {ayb, troels}@es.aau.dk).

Rasmus Waagepetersen is with the Department of Mathematical Sciences, Aalborg University, Denmark (e-mail: rw@es.aau.dk).
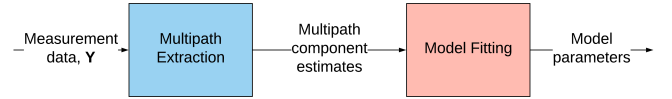

Fig. 1. Calibration procedure usually followed where estimates of multipath components are used as summaries.

and expanded upon this calibration method [4]–[10]. The estimation of the multipath components from measurement data involves high-resolution multipath extraction methods such as CLEAN [11], SAGE [12], and RiMAX [13]. Depending on whether the stochastic model is cluster-based or not, an additional step of clustering the multipath components may also be employed. Clustering is either done manually, e.g. in [4], [5], [14], [15], leading to subjective and non-reproducible results, or using automated algorithms such as [16]–[18], that further increase the complexity of the calibration process. Implementation of these multipath extraction and clustering algorihms is typically non-trivial, and requires a number of arbitrary choices to be made. Moreover, various ad hoc methods are utilized for obtaining the model parameters after multipath extraction. Another potential weakness of such two-step procedures is that the resulting parameter estimates are highly sensitive to the estimation accuracy of the particular set of extracted multipath components. Calibration techniques that do not require multipath extraction but rely on summarizing the data into a set of statistics have been introduced recently in the literature [19]–[23]. However, these methods call for definition of appropriate summary statistics that are informative regarding the model parameters. Moreover, the approximation arising due to summarizing the data maybe difficult to quantify.

In this paper, we propose to use the principled and recognized statistical methodology of maximum likelihood estimation (MLE) to calibrate stochastic channel models with inhomogeneous intensity function. Thus, our parameter estimates are the parameter values maximizing the probability density of the received signals given the transmitted signals. However, we face a missing data problem where it is not possible to evaluate the likelihood function by analytical marginalization with respect to the hidden quantities. We therefore use importance sampling to compute an approximation of the likelihood function using a large Markov Chain Monte Carlo (MCMC) sample from the conditional distribution of the multipath components given the observed data [24]. Thus, in contrast to the previously mentioned methods, our method does not rely on the validity of just one particular set of multipath

components. We believe this will reduce bias and variance of the resulting parameter estimates. Moreover, approximated likelihoods for different models can be used in a natural way for model selection. Considering a parametric Turin model for simulated and real data, we demonstrate the feasibility of Markov Chain Monte Carlo maximum likelihood estimation (MCMC MLE) for model calibration. The MCMC method can also be adapted in a straightforward manner to sample from the posterior distribution of the parameters in case informative priors are available for the parameters.

The rest of the paper is organized as follows. Section II introduces the inhomogeneous Turin model as the stochastic multipath model studied in this paper. Next, in Section III, we describe our proposed procedure for approximate MLE using MCMC. We also explain what properties of the stochastic multipath model require us to develop problem-specific adaptations to standard MCMC and optimization methods. Section IV illustrates our calibration procedure using simulated datasets. While this already provides an intuition on the strengths and peculiarities of the calibration procedure, they become even more apparent in Section V, where we analyze a real dataset of indoor channel measurements originally considered in [25]. Finally, Section VI concludes the paper with a discussion and indications to avenues of future research.

II. STOCHASTIC MULTIPATH MODEL

A. Signal model

Consider frequency domain measurements of a single-input single-output linear, time-invariant radio channel in the band $[-B/2, B/2]$ obtained by a vector network analyzer [26]. In each measurement run, the transfer function is sampled at K equispaced frequencies. The measurement data is modeled as a random vector $\mathbf{Y} = (Y_1, \ldots, Y_K)$ with entries

$$Y_k = H_k + N_k, \quad k = 1, \ldots, K, \tag{1}$$

where H_k is the transfer function sampled at the kth frequency and N_k denotes the measurement noise. The noise samples $(N_1, \ldots, N_K)$ are assumed to be independent and identically distributed circular symmetric Gaussian random variables each with variance σ^2. We denote a realization of the measurement vector $\mathbf{Y}$ by $\mathbf{y} = (y_1, \ldots, y_K)$. Repeating the measurements M times yields the sequence of independent realizations $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}$.

Taking the discrete-frequency, continuous-time inverse Fourier transform of the measurement vector gives the time domain measurement (with a misuse of notation)

$$Y(t) = \frac{1}{K} \sum_{k \leq K} Y_k \exp(i2\pi k \Delta f t), \tag{2}$$

where $\Delta f = B/(K-1)$ is the frequency spacing between two measurement points, giving the period of the time domain signal as $\tau_{\max} = 1/\Delta f$. We denote the imaginary unit by i. The power delay spectrum of $Y(t)$ is defined as

$$P_y(t) = \mathbb{E}\left[|Y(t)|^2\right] = (P_h * |s|^2)(t), \tag{3}$$

where $s(t)$ denotes the transmitted signal in the time domain[1]. The power delay spectrum $P_h(t)$ may be informally interpreted as $P_h(t) = \mathbb{E}\left[|H(t)|^2\right]$, with $|H(t)|^2$ being the instantaneous power delay profile of the channel, see e.g. [27] or [28].

B. Stochastic multipath model

The channel transfer function of a multipath model is of the form

$$H_k = \sum_{\tau \in Z} \alpha_\tau \exp(-i2\pi \Delta f k\tau), \quad k = 1, \ldots, K \tag{4}$$

where Z is a point process on the positive real line $\mathbb{R}_+$ containing the *propagation time delays* τ. A complex-valued *gain* α_τ is associated to each delay $\tau \in Z$. Thereby the process $Z_{\mathsf{m}} = \{(\tau, \alpha_\tau)\}_{\tau \in Z}$ constitutes a *marked* point process on $\mathbb{R}_+ \times \mathbb{C}$. Hence, we refer to a pair (τ, α_τ) as a marked point. The support of the point process Z is the interval $I = [\tau_0, \tau_{\mathsf{max}}]$, where τ_0 is the delay of the line-of-sight (LOS) path.

Particular stochastic multipath models are obtained upon specifying the marked point process Z_{m}. A multitude of such models have been proposed in the literature. Here, we follow the approach by Turin [3], and let Z_{m} be an independently marked Poisson process. This model is completely specified by the intensity function (arrival rate) and its mark density.

The power delay spectrum is connected to the arrival rate and mark density. Assuming the complex gains to be uncorrelated given the delay variables, (corresponding to the familiar uncorrelated scattering assumption), the power delay spectrum factorizes as [28]

$$P_h(\tau) = \lambda(\tau)\mathbb{E}[|\alpha_\tau|^2|\tau] \tag{5}$$

where $\lambda(\cdot)$ denotes the intensity function (or arrival rate) for the delays Z. The power delay spectrum is well studied as it is easy to measure and model. For in-room environments, the power delay spectrum is well modeled by an exponential decaying function,

$$P_h(t) = \begin{cases} G_0 \exp(-t/T), & t > 0 \\ 0, & t \leq 0, \end{cases} \tag{6}$$

where T is the reverberation time and the gain factor G_0 is a positive constant. See [28], [29] and references therein. We first define the arrival rate and thereafter specify the mark density so that its second moment fulfills (5) and (6).

While in his original work, Turin determined the arrival rate empirically in a non-parametric manner, a number of parametric models have occurred in the literature [28]. We consider here the flexible two-parameter model for the arrival rate proposed in [28]

$$\lambda(t) = ct^{\kappa_1}, \quad t \geq 0 \tag{7}$$

with $c > 0$ and $\kappa_1 \in \mathbb{R}$. This model class includes both the constant rate model $\lambda(t) = c$, $t \geq 0$, which is widely used in

---

[1]In the case considered here, $s(t)$ is the inverse discrete Fourier transform of the rectangular frequency window applied in the measurements.

the literature due to its simplicity [4], [6], and the quadratic rate model $\lambda(t) = ct^2$ obtained by mirror source analysis of an empty rectangular room.[2] The quadratic rate model is able to represent the experimentally observed specular-to-diffuse transition [28], [30]–[32]. In the present study, we use the model (7). For computational convenience, we reparametrize the model as

$$\lambda(t) = \exp(\kappa_0 + \kappa_1 \log(t)) \tag{8}$$

where $\kappa_0 = \log(c)$.

Given Z, the path gains α_τ for $\tau \in Z$ are modeled as independent zero-mean complex Gaussian random variables. Therefore, conditioned on the delays, the magnitude, $|\alpha_\tau|$, is Rayleigh distributed, with the corresponding phase being modeled as a uniform distribution on $[0, 2\pi)$. To satisfy (5), (6) and (8), we set the second conditional moment of the magnitude as

$$\mathbb{E}[|\alpha_\tau|^2 | \tau] = G_0 \exp(-\kappa_0) \exp[-\tau/T - \kappa_1 \log(\tau)]$$
$$= \exp[\gamma_0 + \gamma_1 \tau - \kappa_1 \log(\tau)]$$

where we have introduced the reparametrization $\gamma_0 = \log(G_0) - \kappa_0$ and $\gamma_1 = -1/T$. Note that $G_0, T, \kappa_1, \kappa_0$ can be recovered uniquely from $\gamma_0, \gamma_1, \kappa_0, \kappa_1$ and vice-versa. In other words, conditional on τ, the real and imaginary parts of α_τ are independent zero-mean normal, each with variance $\exp[\gamma_0 + \gamma_1 t - \kappa_1 \log(t)]/2$.

The reparametrization using κ_0, γ_0 and γ_1 is not of critical importance but leads to somewhat nicer expressions for derivatives when using Newton-Raphson updates later on, see Section III-C2.

C. Estimation Problem and Likelihood Function

To calibrate the channel model, the parameter vector $\theta = [\kappa, \gamma, \sigma^2]^\top$ with the shorthand notations $\kappa = (\kappa_0, \kappa_1)$ and $\gamma = (\gamma_0, \gamma_1)$ should be estimated from the data $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}$. Following the maximum likelihood principle, the estimate is obtained as

$$\hat{\theta} = \arg\max_\theta \prod_{m=1}^M L(\theta; \mathbf{y}^{(m)}) \tag{9}$$

where $L(\theta; \mathbf{y}^{(m)}) = p(\mathbf{y}^{(m)}; \theta)$ is the likelihood based on one realization $\mathbf{y}^{(m)}$.

Denote by $Z_{\mathsf{m}}^{(m)}$ the point process associated to the measurement vector $\mathbf{Y}^{(m)}$. Suppose for a moment that in addition to the measurement data $\mathbf{y}^{(m)}$ also the corresponding point process realization $z_{\mathsf{m}}^{(m)}$ is observed. Then the likelihood function based on $(\mathbf{y}^{(m)}, z_{\mathsf{m}}^{(m)})$ is

$$L(\theta; \mathbf{y}^{(m)}, z_{\mathsf{m}}^{(m)}) = p(\mathbf{y} | z_{\mathsf{m}}^{(m)}; \sigma^2) p(z_{\mathsf{m}}^{(m)}; \kappa, \gamma) \tag{10}$$

where $p(\mathbf{y}^{(m)} | z_{\mathsf{m}}^{(m)}; \sigma^2)$ is the complex Gaussian density

$$p(\mathbf{y}^{(m)} | z_{\mathsf{m}}^{(m)}; \sigma^2) =$$
$$(2\pi\sigma^2)^{-K} \exp\left(-\frac{1}{2} \sum_{k=1}^K \left(|y_k^{(m)} - H(f_k)|/\sigma\right)^2\right)$$

---

[2] [27] alternatively derived the intensity function for a propagation graph model for the in-room scenario with diffusely reflecting walls. This gives rise to a two-parameter exponential rate model $\lambda(t) = c \exp(\kappa_1 t)$.

and $p(z_{\mathsf{m}}^{(m)}; \kappa, \gamma)$ is the point process density of $Z_{\mathsf{m}}^{(m)}$. The notion of a point density is non-standard as the number of points varies from realization to realization. Technically speaking, the density of the delays is a density with respect to a unit-intensity Poisson process distribution, [2, Section 6.1]. Specifically, the point process density of $Z_{\mathsf{m}}^{(m)}$ can be written [2, Section 3.3] as

$$p(z_{\mathsf{m}}^{(m)}; \kappa, \gamma) = \prod_{\tau \in z^{(m)}} f(\alpha_\tau; \gamma)$$
$$\times \exp\left(-\int_I \exp(\kappa_0 + \kappa_1 \log(t)) \mathrm{d}t\right)$$
$$\times \prod_{\tau \in z^{(m)}} \exp(\kappa_0 + \kappa_1 \log(\tau)).$$

The first factor is the product of the complex Gaussian densities $f(\alpha_\tau; \gamma)$ for the marks α_τ and the product of the last two factors is the Poisson point density for the delays $\tau \in z^{(m)}$.

In practice, $z_{\mathsf{m}}^{(m)}$ is not available and the likelihood is obtained by marginalizing with respect to $Z_{\mathsf{m}}^{(m)}$. More precisely, according to the law of total probability,

$$L(\theta; \mathbf{y}^{(m)}) = \mathbb{E}_\theta[p(\mathbf{y}^{(m)} | Z_{\mathsf{m}}^{(m)}, \sigma^2)]. \tag{11}$$

The likelihood function (11) is unfortunately not available in closed form because it is an expectation of a conditional probability that depends on $Z_{\mathsf{m}}^{(m)}$ in a complicated way and $Z_{\mathsf{m}}^{(m)}$ moreover does not have a fixed dimension. Consequently, the MLE cannot be obtained in a straightforward manner.

The estimation problem is complicated due to the missing data: the maximization of the likelihood would be straightforward if only the point process Z_{m} could be observed. Thus it is tempting to resort to a two-step procedure by first estimating Z_{m} using well known high-resolution path extraction techniques and thereafter to estimate the model parameters. However, such two-step procedures are problematic. Commonly, such high-resolution estimators work under the assumption that the number of points in Z_{m} is known. This number is particularly challenging to estimate when the arrival rate is high compared to the inverse of the measurement signal bandwidth. In the light of the arrival rate model (7) this situation is very relevant in our study.

III. MCMC MLE

To obtain the MLE, we propose an MCMC approach to obtain the maximum likelihood estimate of θ. Our approach is inspired by the method proposed in [2]. Thus, we rely on the observation that maximization of the likelihood function is equivalent to maximization of the likelihood ratio $L(\theta; \mathbf{y})/L(\theta_0; \mathbf{y})$ for a fixed reference parameter value θ_0. This ratio can, as discussed in Subsection III-A, be evaluated using conditional samples of Z_{m} given $\mathbf{y}$. Furthermore, these samples can be generated using an MCMC algorithm detailed in Section III-B. Finally, the approximated likelihood ratio is maximized with respect to θ as discussed in Subsection III-C. We comment in Subsection III-D on approaches for model selection.

For ease of exposition, we focus in the following on approximation of the likelihood ratio in case of one realization, i.e. $M = 1$ and denote by $\mathbf{y}$ the observed measurement data. The derived methodology is straightforwardly extendable to the case $M > 1$, since by (9), the likelihood ratio for multiple realizations is simply obtained by multiplying the likelihood ratios for each separate realization.

The proposed MCMC maximum likelihood approach bears some resemblance to Monte Carlo EM in that it uses samples from the conditional distribution of the missing data given the observed data. However, directly maximizing the Monte Carlo approximation of the likelihood is more efficient than using EM steps for maximization, see the discussion in [33].

A. Monte Carlo approximations of likelihood

Using a result from [2, Section 8.6.1], the likelihood ratio can be expressed as[3]

$$\frac{L(\theta; \mathbf{y})}{L(\theta_0; \mathbf{y})} = \mathbb{E}_{Z_\mathsf{m}|\mathbf{y};\theta_0}\left[\frac{L(\theta; \mathbf{y}, Z_\mathsf{m})}{L(\theta_0; \mathbf{y}, Z_\mathsf{m})}\right], \quad (12)$$

where $\mathbb{E}_{Z_\mathsf{m}|\mathbf{y};\theta_0}$ denotes conditional expectation with respect to the hidden multipath components Z_m given the data $\mathbf{y}$ under the parameter θ_0 and the full data likelihoods on the right hand side are given by (10). The right-hand side of (12) is an importance sampling formula allowing us to use the conditional distribution of Z_m given $\mathbf{y}$ to integrate out Z_m from the full data likelihood ratio $L(\theta; \mathbf{y}, Z_\mathsf{m})/L(\theta_0; \mathbf{y}, Z_\mathsf{m})$. Thus the right-hand side of (12) can be approximated by an empirical average based on samples[4] $Z_{\mathsf{m},1}, \ldots, Z_{\mathsf{m},N}$ from the conditional distribution of Z_m given $\mathbf{y}$,

$$\frac{L(\theta; \mathbf{y})}{L(\theta_0; \mathbf{y})} \approx \frac{1}{N}\sum_{n=1}^{N} W_n, \quad (13)$$

with the notation

$$W_n = L(\theta; \mathbf{y}, Z_{\mathsf{m},n})/L(\theta_0; \mathbf{y}, Z_{\mathsf{m},n}). \quad (14)$$

A delicate issue of the estimator (13) is its Monte Carlo variance. If θ differs substantially from θ_0, then only a small number of terms contribute significantly to the Monte Carlo estimator in (13), which in turn leads to a very high variance of the estimator. The magnitude of the degeneracy is quantified by the *effective sample size* [34]

$$\mathsf{ESS} = \frac{\left(\sum_n W_n\right)^2}{\sum_n W_n^2}. \quad (15)$$

The effective sample size equals N in the extreme case where $W_1 = W_2 \cdots = W_N$. In the other extreme where one term dominates, then the effective sample size approaches unity. When applied to dependent samples, such as those obtained by MCMC samples, the effective sample size can be somewhat optimistic as it does not take into account correlation between samples. However, we still find it useful for gauging of the

---

[3]Replacing the right-hand side of (11) by a direct Monte Carlo approximation using samples from the marginal distribution of Z_m under θ is possible, but gives an unacceptably high variance of the estimated likelihood.

[4]An MCMC algorithm for sampling the conditional distribution of Z_m is discussed in Section III-B.

Algorithm 1 Birth/death MCMC sampler (`rand`/`randn` means draw independent standard uniform/normal variate, $|I|$ is length of I, and $n(z)$ is number of points in z)

Input. $\theta, z_\mathsf{m}^{(0)}$
Output. $z_\mathsf{m}^{(1)}, z_\mathsf{m}^{(2)}, \ldots$ with stat. distribution $p(z_\mathsf{m}|\mathbf{y}, \theta)$
for $k = 1, 2, \ldots$ **do**
  $z_\mathsf{m}' \leftarrow z_\mathsf{m}^{(k-1)}$ [z_m' proposal for next state of Markov chain]
  if `rand` $< p_{\mathrm{move}}$ **then**
    Pick (τ, α_τ) uniformly at random from z_m'
    if `rand` $< p_{\mathrm{delay}}$ **then**
      $\tau \leftarrow \tau + \sigma_{\mathrm{delay}} \cdot$ `randn`
      mhr $\leftarrow L(\theta; \mathbf{y}, z_\mathsf{m}')/L(\theta; \mathbf{y}, z_\mathsf{m}^{(k-1)})$
    else
      if `rand` $< p_{\mathrm{phase}}$ **then**
        phase$(\alpha_\tau) \leftarrow$ phase$(\alpha_\tau)+\sigma_{\mathrm{phase}} \cdot$ `randn`
        mhr $\leftarrow L(\theta; \mathbf{y}, z_\mathsf{m}')/L(\theta; \mathbf{y}, z_\mathsf{m}^{(k-1)})$
      else
        $v :=$ `randn`
        magn$(\alpha_\tau) \leftarrow$ magn$(\alpha_\tau) \cdot \exp(\sigma_{\mathrm{magnitude}} \cdot v)$
        mhr $\leftarrow \exp(v)L(\theta; \mathbf{y}, z_\mathsf{m}')/L(\theta; \mathbf{y}, z_\mathsf{m}^{(k-1)})$
      end if
    end if
  else
    if `rand` $< p_{\mathrm{birth}}$ **then**
      Add marked point (τ, α_τ) to z_m' with τ uniform in I and $\alpha_\tau \sim f$
      mhr $\leftarrow \dfrac{L(\theta; \mathbf{y}, z_\mathsf{m}')(1-p_{\mathrm{birth}})|I|}{L(\theta; \mathbf{y}, z_\mathsf{m}^{(k-1)})p_{\mathrm{birth}}f(\alpha_\tau)(n(z_\mathsf{m}')+1)}$
    else
      Pick (τ, α_τ) uniformly at random from z_m' and delete it from z_m'.
      mhr $\leftarrow \dfrac{L(\theta; \mathbf{y}, z_\mathsf{m}')p_{\mathrm{birth}}f(\alpha_\tau)n(z_\mathsf{m}^{(k-1)})}{L(\theta; \mathbf{y}, z_\mathsf{m}^{(k-1)})(1-p_{\mathrm{birth}})|I|}$
    end if
  end if

  if `rand` $< \min\{1, \mathrm{mhr}\}$ **then**
    $z_\mathsf{m}^{(k)} \leftarrow z_\mathsf{m}'$ [go to proposed state]
  else
    $z_\mathsf{m}^{(k)} \leftarrow z_\mathsf{m}^{(k-1)}$ [remain at current state]
  end if
end for

---

quality of the Monte Carlo estimator. Alternatively, the variance of the Monte Carlo estimator could be estimated using time series methods [35]. In case of multiple measurements, $M > 1$, several MCMC samplers, one for each measurement vector, would be run in parallel. In that case it would be natural to consider the minimal ESS over the M samplers.

B. Birth/death MCMC sampling with parallel tempering

1) Birth/Death MCMC: The challenging task of sampling from $Z_\mathsf{m}|\mathbf{y}, \theta$ can be tackled using specialized MCMC samplers for point processes [2]. Here, we rely on Algorithm 1 which is a variant of the birth-death MCMC algorithm from [2, Chapter 7]. A diagram of one iteration of the birth-death algorithm is shown in Fig. 2.

The MCMC updates of Algorithm 1 are births, deaths and moving of marked points (τ, α_τ). We first elucidate the mechanisms behind the birth and death steps. A birth proposal attempts to add a marked point (τ, α_τ) where τ is drawn uniformly at random in the sampling window I and α_τ is drawn from the circular symmetric Gaussian distribution
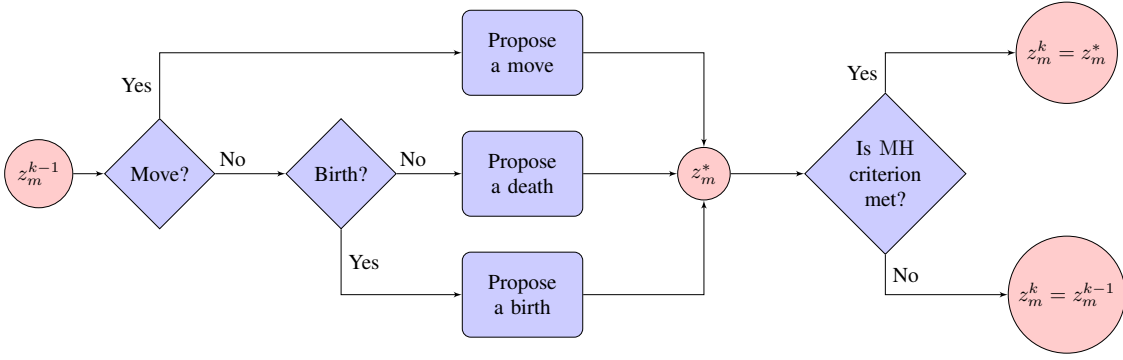
Fig. 2. Diagram of one iteration of the birth-death MCMC algorithm.

described in Section II. A death proposal attempts to remove a marked point selected from the uniform distribution on the current marked points. The proposals are accepted or rejected according to Metropolis-Hastings ratios appropriate for the set-up of a varying number of points, see [2, Chapter 7].

The birth-death MCMC algorithm, unfortunately, suffers from the problem of slow mixing. The reason for this is that if a marked point (τ, α_τ) is borne close to the true location of a delay, this may increase the likelihood of the data under the model substantially, even if the mark α_τ is not entirely correct. Since such a point dies with small probability, it likely remains in the MCMC algorithm for a long time, thus leading to slow mixing of the Markov Chain.

To improve the mixing, we introduce updates that only change the mark α_τ for a uniformly selected point τ. For instance, if a large mark is changed to a smaller, this may increase the chance that a death of the associated marked point becomes accepted later on. In addition, if the originally proposed mark was too small, the mark change allows for correcting this by proposal of a larger mark.

To further improve the mixing, we use a parallel tempering scheme which combines several birth-death Markov chains. Parallel tempering is a versatile technique to reduce autocorrelation in slowly-mixing Markov chain samplers by running in parallel several variants of the chain that mix substantially faster [24]. Occasionally, the faster chains swap states with the slower ones, thereby reducing the mixing time of the slower chains. Here, we can construct faster chains by increasing the noise level whereby the conditional distribution of the point process given the data becomes more dispersed, so that the chain does not get stuck as easily. The swaps between different chains are controlled by a Metropolis-Hastings criterion as follows. A chain in state z_m with parameters θ swaps states with a chain in state z'_m with parameters θ' with probability

$$\min\left\{1, \frac{L(\theta'; \mathbf{y}, z_\mathsf{m})L(\theta; \mathbf{y}, z'_\mathsf{m})}{L(\theta; \mathbf{y}, z_\mathsf{m})L(\theta'; \mathbf{y}, z'_\mathsf{m})}\right\}. \tag{16}$$

In the swap phase, we order the parameters according to their noise level σ and then sequentially attempt a swap move for every pair of successive parameters. That is, if parallel tempering considers K noise levels $\sigma_1 < \cdots < \sigma_K$, then we attempt swap moves between σ_i and σ_{i+1} for $i < K$.

Large spans of noise variances and large number of parallel chains generally reduces the mixing time at the cost of

parallelization overhead. We found that working with only six temperature levels reduces the autocorrelation substantially while maintaining a reasonable complexity. The distances between the six noise levels are chosen such that we achieve the recommended acceptance rates between 20% and 50% [36].

2) Initialization: In principle, the MCMC sampler converges for any choice of initial configuration. The number of iterations required to reach the target equilibrium, called the burn-in, can be reduced by careful initialization. We proceed in two steps. First, we run the MCMC sampler for a number of steps starting in a random initial configuration. This generally leads to a configuration with too many points and a long burn-in would be needed to eliminate the excessive points. Therefore, we remove points that are too close together to obtain a better initial configuration. More precisely, we achieved good results by removing delays that are less than $1\,\mathsf{ns}$ apart.

3) Thinning and swapping: For each of the parallel chains we apply 400,000 basic birth-death MCMC steps. After each 200th step pairwise swaps of chains are proposed. The first 100,000 samples are discarded as a burn-in. Subsequently, to reduce autocorrelation and save storage, we only retain each 200th state of the MCMC sampler. Moreover, the states of the chains with increased thermal noise are discarded. This yields in total a sample of 1,500 realizations of the conditional distribution. For the simulation study presented in Section IV below, the autocorrelation plots in Figure 3 for the sequence of logarithms of the full data likelihood $\log L(\theta; \mathbf{y}, Z_{\mathsf{m},n})$ illustrate that after thinning and parallel tempering, the autocorrelation remains under control.

C. Optimization methods

To maximize the Monte Carlo approximation of the likelihood (13), we use the cross-entropy method (CEM) [37] which is a gradient free method that works robustly in settings where the objective function is subject to Monte Carlo errors. To ensure that the MCMC approximations of the likelihood within the CEM remain valid, we define a trust region, see Section III-C1. After convergence of the CEM/trust region procedure we fine-polish the estimate by applying a few Newton-Raphson updates. The computation of the gradient
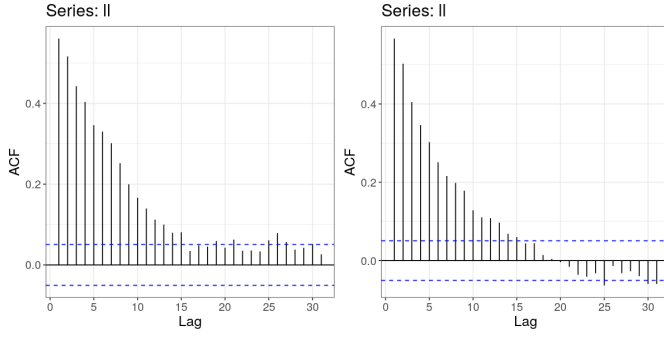
Fig. 3. Estimated autocorrelation for MCMC samples of the log full data likelihood in case of the constant (left) and quadratic (right) rate model. The dashed blue lines indicate lag-wise 95% probability intervals for the estimated auto-correlations under the null-hypothesis of zero autocorrelation.

Algorithm 2 CEM Maximization of $L(\theta; \mathbf{y})$

Input. data $\mathbf{y}$, initial guess $\theta_{\max}$, initial CEM proposal distribution.
Output. ML estimate $\hat{\theta}_{\mathsf{ML}}$
repeat
    $\theta_0 := \theta_{\max}$
    Draw sample $\{Z_{\mathsf{m},n}\}_{n \leq N}$ from $p(z_{\mathsf{m}}|\mathbf{y}; \theta_0)$ using Algorithm 1
    repeat
        repeat
            Generate CEM proposal parameter sample inside trust region of size > 5 times elite sample size.
            Get elite sample from CEM proposal sample.
            Fit new CEM proposal distribution.
            $\theta_{\max} :=$ parameter vector in elite sample with highest likelihood.
        until $\theta_{\max}$ satisfies stopping criteria (see main text)
    until $\theta_{\max}$ in interior of trust region (see main text)
    $\hat{\theta}_{\mathsf{ML}} :=$ output of Newton-Raphson initialized in $\theta_{\max}$

---

and Hessian matrix required for this is discussed in Section III-C2. Algorithm 2 summarizes the resulting procedure and Fig. 4 depicts the procedure in terms of a block diagram.

The optimization procedure is also applicable in case of multiple measurements, $M > 1$. In this case, given θ_0, we would run M MCMC samplers in parallel, one for each measurement vector $\mathbf{y}^{(m)}$, $m = 1, \ldots, M$, and approximate the likelihood for each measurement vector using (13). Finally these approximations are multiplied to get the approximation of the full likelihood to be maximized.

1) CEM with trust region: In the CEM method, a Gaussian proposal distribution is iteratively adapted so as to concentrate it to a small neighborhood of the maximum. More precisely, we first draw a number of parameter vectors independently from the current proposal distribution and evaluate the corresponding likelihoods. Then, a new proposal distribution is fitted to the elite sample, that is, the parameter vectors with the highest likelihoods. This process is repeated until the increase in the highest evaluated likelihood over the elite sample is below some user-specified threshold. The CEM requires evaluations of the likelihood function. However, the cost of these is minor relative to the cost of running the MCMC sampler and the operation can be run in parallel.

If θ and θ_0 are too distant, the approximation (13) of the likelihood ratio (12) becomes unreliable. This situation

TABLE I
SETTINGS OF THE CEM ALGORITHM

Parameter	Value
Initial standard deviation for CEM proposal	1
Size of elite sample	10
ESS-threshold for the trust region	400
Threshold for likelihood convergence	0.1
Final ESS-threshold for interior	750

is indicated by a small ESS value. Therefore, we restrict the CEM maximization to a trust region [38] around the current value θ_0 determined so that the ESS is above a certain threshold for all θ in the trust region. If the CEM maximization terminates at a value $\theta_{\max}$ well inside the interior of the trust region, this value is used as an initial value for some final Newton-Raphson updates to fine-polish the estimate. By well inside we mean that ESS at $\theta_{\max}$ is bigger than a second threshold exceeding the first threshold used to define the trust region. Otherwise we set $\theta_0 = \theta_{\max}$, draw a new MCMC sample and run the CEM maximization procedure once again over a trust region centered around the new value of θ_0 with the original initial proposal standard deviation.

2) Gradient and Hessian for Newton-Raphson updates: Let

$$V_\theta(z) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\log p(\mathbf{y}|z_{\mathsf{m}}, \sigma^2) + \log p(z_{\mathsf{m}}; \kappa, \gamma)\right]$$

denote the gradient of the log joint density of $(\mathbf{y}, z_{\mathsf{m}})$. Following [2, Section 8.6.2], the score function and observed information are

$$u(\theta) = \mathbb{E}_\theta[V_\theta(Z_{\mathsf{m}})|\mathbf{Y} = \mathbf{y}]$$

and

$$j(\theta) = -\mathbb{E}[\mathrm{d}V_\theta(Z_{\mathsf{m}})/\mathrm{d}\theta|\mathbf{Y} = \mathbf{y}] - \mathbb{V}\mathrm{ar}_\theta[V_\theta(Z_{\mathsf{m}})|\mathbf{Y} = \mathbf{y}].$$

The conditional expectations and variances can in general not be evaluated in closed form. However, it is feasible to approximate these quantities by importance sampling. For instance, following [2, (8.43)],

$$u(\theta) \approx \sum_{n \leq N} V_\theta(Z_{\mathsf{m},n})\bar{W}_n$$

where $\bar{W}_n = W_n/\sum_{n \leq N} W_n$ and W_n is defined in (14).

D. Model selection based on likelihood ratios and bridge sampling

In addition to model calibration, a second application of the likelihood-ratio computation concerns model selection. Considering two models A and B with parameter vectors θ_A and θ_B, respectively, we wish to select the model which yields the highest likelihood value. In terms of the likelihood ratio, we select model B if $L(\theta_B; \mathbf{y})/L(\theta_A; \mathbf{y}) > 1$. An appealing property of this criterion is that the ratio on the left-hand side is precisely of the form appearing in equation (13) and therefore amenable to computation via importance sampling. We have here described the likelihood-ratio approach in the case where the compared models belong to the same model class. It is, however, also possible to use the MCMC approach to compute
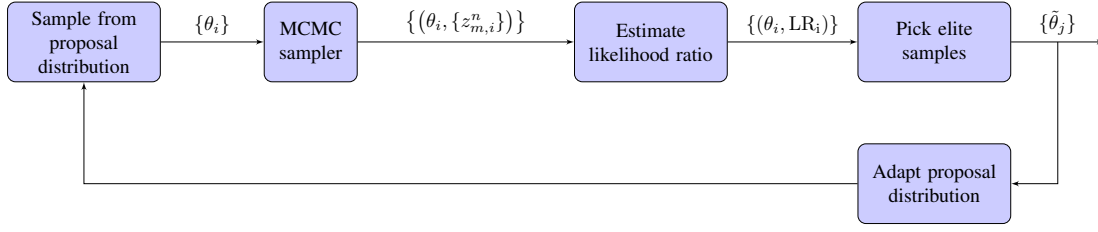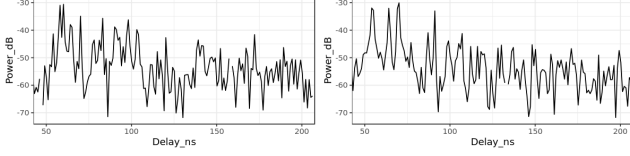
Fig. 4. An overview of the maximization procedure.

Fig. 5. Power-delay profile for realizations of the constant (left) and quadratic (right) rate model.

TABLE II
LOG LIKELIHOOD-RATIOS AND EFFECTIVE SAMPLE SIZES FOR BRIDGE SAMPLING ON SIMULATED DATA.

i	0	1	2	3	4
$\log\left(L(\theta_{i+1};\mathbf{y})/L(\theta_i;\mathbf{y})\right)$	0.11	-1.62	-2.69	-4.61	-4.21
ESS	761	1170	396	1182	106

ratios of likelihoods corresponding to different model classes, possibly with different number of parameters.

If θ_A and θ_B are far apart, then the estimate (13) is unreliable since the weights are almost degenerate. This problem is tackled via *bridge sampling*. Let for ease of notation $\theta_0 = \theta_A$ and $\theta_M = \theta_B$ for some $M \geq 1$. Then, the likelihood ratio is expanded as

$$\frac{L(\theta_M;\mathbf{y})}{L(\theta_0;\mathbf{y})} = \frac{L(\theta_1;\mathbf{y})}{L(\theta_0;\mathbf{y})} \frac{L(\theta_2;\mathbf{y})}{L(\theta_1;\mathbf{y})} \cdots \frac{L(\theta_M;\mathbf{y})}{L(\theta_{M-1};\mathbf{y})}$$

for intermediate parameters $\theta_1, \ldots, \theta_{M-1}$ bridging figuratively the large difference between θ_0 and θ_M. Subsequently, we apply Monte Carlo estimation to each of the ratios on the right-hand side.

IV. SIMULATION STUDY

In this section, we analyze how well MCMC MLE performs on simulated data. We consider two parameter configurations θ_{const} and θ_{quad}. For both, we use the same parameters driving the distribution of the thermal noise and the path gains:

$$\log(\sigma) = -10.5, \ \gamma_0 = -20, \ \gamma_1 = -0.029.$$

Here, the parameters are chosen to resemble the characteristics from the measurement data discussed in Section V. In the parameter set θ_{const}, the arrival rate of the Turin model is constant, i.e.,

$$\kappa_0 = -0.75, \ \kappa_1 = 0.$$

In the parameter set θ_{quad}, the rate increases quadratically, i.e.,

$$\kappa_0 = -10.5, \ \kappa_1 = 2.$$

Here we choose κ_0 such that the expected number of multipath components agrees approximately in both models. Moreover, within the simulation study, we fix the size of the observation window $|I| = 150$ and the delay $\tau_0 = 50$ associated with the LOS path. Figure 5 shows the power-delay profiles simulated from the two models. The settings of the CEM algorithm are given in Tab. I.

A. Bridge Sampling

First we illustrate how to justify selecting either the constant or the quadratic rate model via bridge sampling. We draw a realization $\mathbf{y}$ from the constant rate model θ_{const} and then compare the likelihoods of $\mathbf{y}$ under θ_{const} and θ_{quad}. For this purpose, we interpolate linearly between $\kappa_1 = 0$ and $\kappa_1 = 2$ with a step size of 0.4. The corresponding values of κ_0 are fixed as $(-0.75, -2.7, -4.7, -6.6, -8.6, -10.5)$. In particular, the expected total number of points does not fluctuate substantially among consecutive values. The remaining parameters $\log(\sigma)$, $\gamma_0 = -20$ and $\gamma_1 = -0.029$ agree in θ_{const} and θ_{quad} and are therefore kept fixed. Setting $\theta_0 = \theta_{\mathsf{const}}$ and $\theta_5 = \theta_{\mathsf{quad}}$, Tab. II shows the estimated log likelihood-ratios $\log\left(L(\theta_{i+1};\mathbf{y})/L(\theta_i;\mathbf{y})\right)$ together with the effective sample sizes based on 2,000 nominal samples. The resulting log likelihood ratio $\log\left(L(\theta_{\mathsf{quad}};\mathbf{y})/L(\theta_{\mathsf{const}};\mathbf{y})\right)$ becomes -13.02 identifying θ_{const} as the correct model.

In order to assess how robust the model selection method is with respect to taking a different sample, we took 100 samples from the constant rate model and compared the likelihood with that under the quadratic rate model. In all of the 100 samples, bridge sampling indeed indicates a higher likelihood for the constant rate model. Conversely, when taking 100 samples from the quadratic rate model and comparing it with the constant rate model, bridge sampling indicates in all of the considered samples a higher likelihood for the true quadratic rate model.

B. MLE with known multipath components

The first test case for MLE is the setting of known multipath components. Although measurements in the field do not reveal this kind of information directly, such test cases help to build intuition on how well maximum likelihood estimation can work in an idealized setting.

As an illustration, we provide profile plots of the log-likelihood. That is, we fix all but one of the parameters at their true values and then trace how the log-likelihood changes when varying the considered parameter.
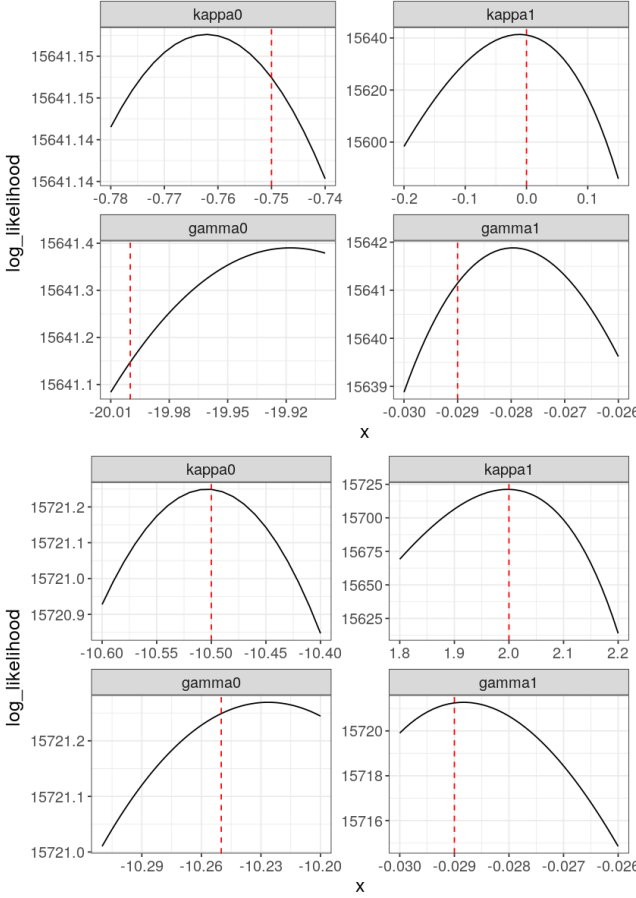
Fig. 6. Profile plots of the log-likelihood for known multipath components in the constant (top) and quadratic (bottom) rate model. Red dashed lines indicate the true parameters.

Fig. 7. Fixed κ_1. Box plots for estimated parameters in 50 realizations of the constant (left) and quadratic (right) rate model. Red dashed lines indicate the true parameters.

In general, Figure 6 suggests that the log-likelihood is maximized close to the true parameters. Still, even in this idealized setting, we do see deviations of the parameter estimates from the true values, and they become more pronounced as we now consider the case of unknown multipath components.

C. Unknown multipath components – fixed κ_1

After these initial findings, we now rely on Algorithm 2 to estimate the model parameters from simulated data sets where for each of the parameter configurations θ_{const} and θ_{quad} we conduct 50 simulations. In this section we regard κ_1 to be known and fix it at its true value. For the remaining parameters, we initialize the optimization at parameter values obtained by perturbing the true values of the parameters. Then, we maximize the log-likelihood via Algorithm 2.

In the present setting, we found CEM to perform robust optimization. In the first steps, the likelihood improvements are large, but effective sample sizes at the new parameter values are small. In other words, although the new parameters indicate a substantially higher likelihood, the Monte Carlo approximation of the likelihood could be quite imprecise. However, as the optimization proceeds, the improvements become smaller, while the effective sample sizes increase. In
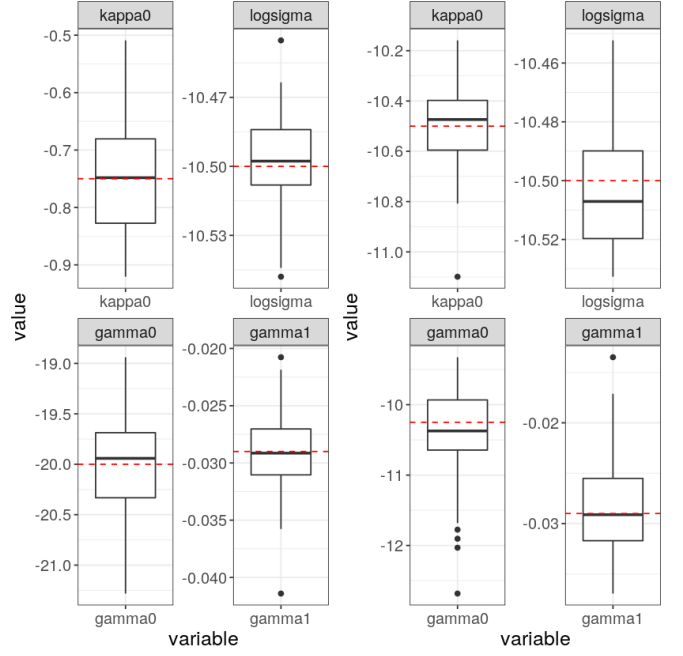
particular, the final decision of identifying the maximum is based on a high effective sample size.

Figure 7 shows boxplot of the parameter estimates under both true parameter configurations. The boxplots illustrate that the medians of the estimates are almost identical to the true values both in the case $\kappa_1 = 0$ and $\kappa_1 = 2$. Overall the MCMC MLE seems to work well.

D. Unknown multipath components – variable κ_1

Next, we optimize with respect to the full parameter vector by including also κ_1 in the optimization process. When estimating the parameters from 50 realizations, Figure 8 reveals that while the medians are still close to the true values, the estimates of both κ_0 and κ_1 now fluctuate substantially. In particular for κ_0 it is evident that the estimation variance is much smaller when κ_1 is fixed compared to when κ_1 is included in the estimation. This is caused by a strong entanglement of the parameters κ_0 and κ_1 that we now explore in further detail.

E. Issues with parameter idenfiability

Due to the complex interplay between the effects of the parameters κ_0 and κ_1, it is difficult to optimize the likelihood jointly with respect to these parameters. For instance, both κ_0 and κ_1 influence the total intensity of points (similarly, both γ_0 and γ_1 affect the general magnitude of the path gains). The contour plot of the log likelihood-ratio for a simulated data set under θ_{const} in Figure 9 illustrates this issue. Indeed, this plot shows two local maxima, both exhibiting a ridges of (κ_0, κ_1)-combinations where the log likelihood-ratio is close to the local optima.
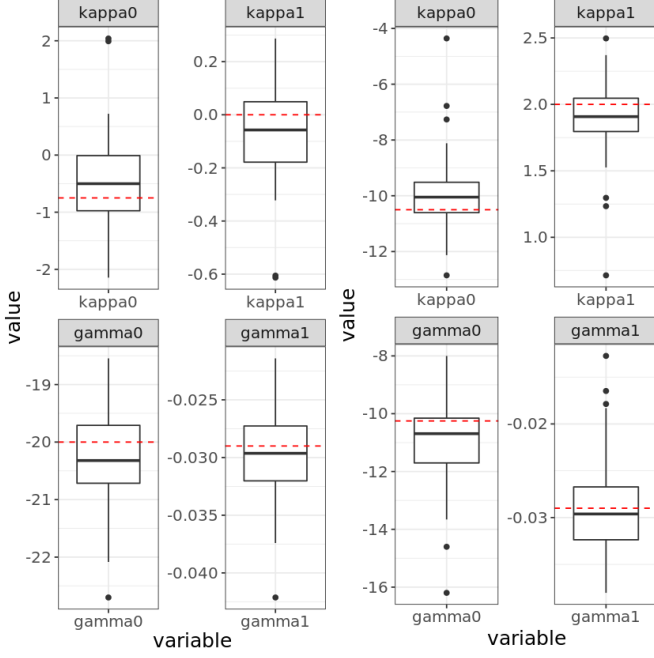
Fig. 8. Variable κ_1. Box plots for estimated parameters in 50 realizations of the constant (left) and quadratic (right) rate model. Red dashed lines indicate the true parameters. Box plot for $\log \sigma^2$ omitted due to space constraints.
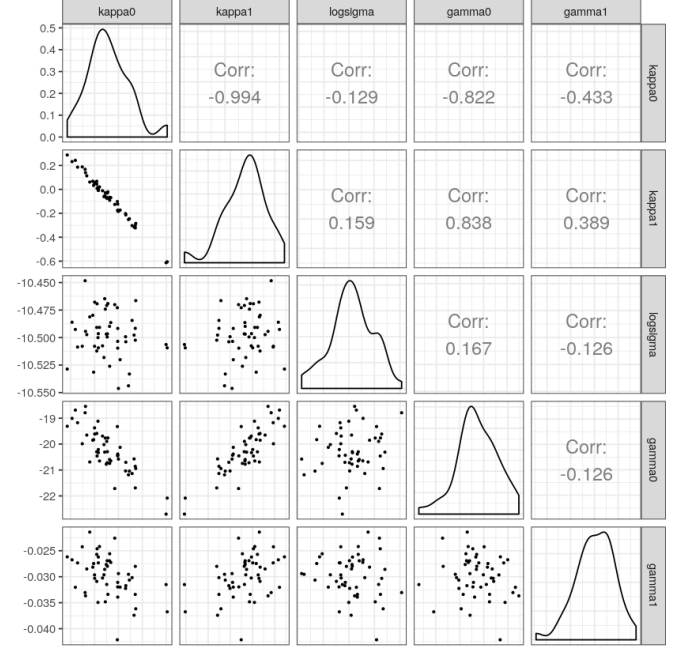

Fig. 10. Pairwise scatter plots for the parameter estimates of 50 realizations of the constant rate model.
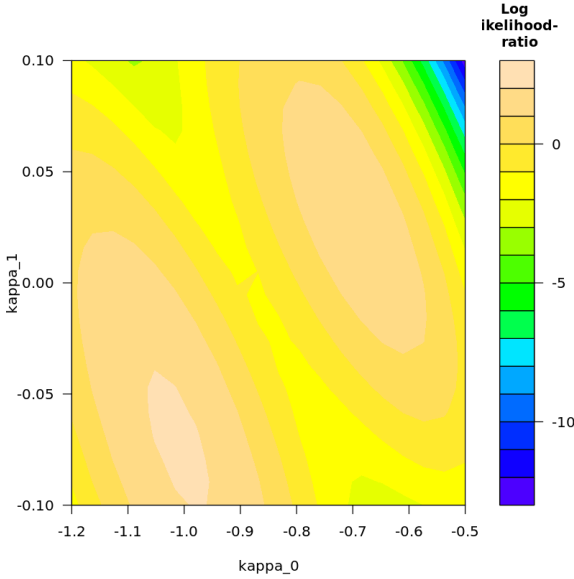

Fig. 9. Contour plot of the log likelihood-ratios for varying κ_0 and κ_1 based on samples from an MCMC run of length 2,000 in the constant rate model.

The parameters κ_0 and κ_1 are thus not jointly well identified by the likelihood. This entails strong correlation between the estimates of κ_0 and κ_1 as well as high variance of each of the estimates. This is further illustrated by the scatterplots of estimates in Figure 10 for each pair of parameters. The estimates shown are those obtained from the 50 simulations under θ_{const}. Supporting the previous findings, we detect a strong negative linear relation between the estimates of the intensity parameters κ_0 and κ_1. Similarly, the estimate of γ_0 is nicely approximated by an affine function of the estimates

of κ_0 and κ_1.

The optimization procedure CEM relies on a Gaussian proposal distribution with diagonal covariance structure. For tightly entangled parameters, the optimization therefore explores the parameter space poorly. Hence, in order to account for the correlations, we transform the parameters linearly with a preconditioning matrix prior to applying the CEM. In the setting of the simulation study we obtain an appropriate linear transformation from preliminary parameter estimates for different simulations. It may not be easy to determine such a transformation in a given application.

V. APPLICATION TO MEASUREMENT DATA

Having analyzed simulated data, we now turn to indoor channel data originally considered in [25]. The data contains the channel response for 750 equally-spaced measurements in the range $[2\,\mathrm{GHz}, 3\,\mathrm{GHz}]$. In particular, the impulse response lies in the interval $[0\,\mathrm{ns}; 750\,\mathrm{ns}]$ and decays rapidly after a strong peak close to $\tau_0 = 50$ ns. Therefore, we henceforth work with a window of size $|I| = 150$.

A. MCMC MLE and bridge sampling

As we saw in the simulation section, the optimization is prone to become unstable when when estimating κ_0 and κ_1 jointly. Hence, we consider two fixed κ_1 values of particular interest: $\kappa_1 = 0$ (constant rate model) and $\kappa_1 = 2$ (quadratically increasing intensity).

In our MCMC set-up, we found that when starting from reasonably chosen initial parameters, the optimization converges both for the constant and for the quadratic rate model. Since random fluctuations were stronger than for the synthetic data, we stabilized the optimization by increasing the number of
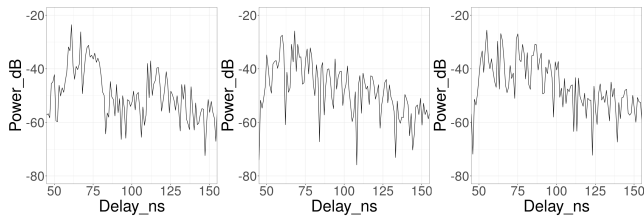
Fig. 11. Power-delay profile for measurement data (left), a realization of the constant rate model (center) and a realization of the quadratic rate model (right).
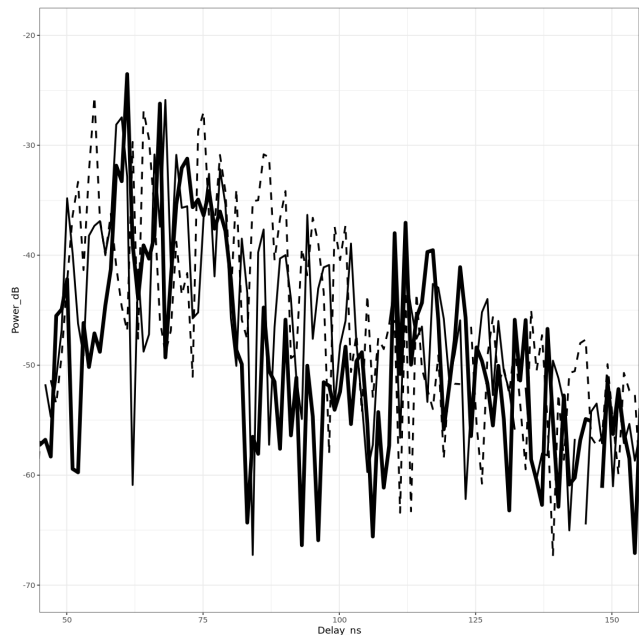

Fig. 13. Log likelihood ratios $\log(L(\theta_{i+1}, \mathbf{y})/L(\theta_i, \mathbf{y}))$ plotted against κ_1.

for the considered data set, the constant rate model seems more appropriate than the quadratic rate model.

VI. CONCLUSIONS

The developed calibration method for stochastic multipath radio channel models is based on the well-established method of maximum likelihood estimation. Thus, we have shown that it is possible to approach the calibration problem in a statistically sound manner without the need to resort to heuristic techniques. In particular, our approach breaks the line of the widespread approach originating from the seminal works of Turin in the 1970s where the calibration problem is broken down into (arbitrarily defined) sub-problems that are tackled by estimators developed separately. Although the developed method has been used to calibrate stochastic channel models using indoor radio channel measurements, it is applicable to measurements from other scenarios as well. Another potential use of the proposed method is for calibrating stochastic channel models based on the output of a ray tracer instead of measurement data.

We find that, despite the intractability of the likelihood function, maximum likelihood estimation is possible by estimating the likelihood function using a birth/death MCMC sampler and then optimizing it using our CEM algorithm. Obviously, being an Monte Carlo approach, it necessitates repeated sampling from a Markov Chain, which entails a significant computational complexity. It was not the objective of the present work to optimize the estimator for computational complexity, and thus we envision that more efficient samplers can be made, in particular considering the availability of more measurement data. Nevertheless, we find that the proposed method is indeed viable provided the necessary computational power.

We observed for the considered model, that only a linear combination of the parameters κ_0 and κ_1 is well identified by the data but not κ_0 and κ_1 separately. This is apparent from plots of the likelihood function. On the contrary, we suspect that such lack of identifiability may be hidden by the current step wise methods based on initial identification of delays and gains. The optimization difficulties due to the poor identifiability can be somewhat mitigated using a reparametrization. However, we resolve the issue by a more robust discrete approach where optimization is first performed over a discrete subset of the parameter space and afterwards the estimate is refined by the method of bridge sampling.

With a view towards avenues of future research, we note that although parametric Turin models can represent inhomogeneities in the arrival distribution, they are still based on the

Fig. 12. Power-delay profile for measurement data (thick), a realization of the constant rate model (thin) and a realization of the quadratic rate model (dashed).

MCMC samples from 1,500 to 2,500. Figure 11 illustrates the power-delay profile of realizations for the fitted constant and quadratic rate models. Although the plots already provide an indication on the different structure of the arrival rates, we stress that the power-delay profile can vary substantially from one realization to another. To see clearly how the model fits to the data, we also plot the measured and estimated power delay profiles on the same figure, see Figure 12.

Let $\hat{\theta}_{\mathrm{const}}$ and $\hat{\theta}_{\mathrm{quad}}$ denote the parameter estimates under the constant and quadratic intensity models. In order to compare the fitted constant rate model with the fitted quadratic rate model, we estimate the likelihood ratio $L(\hat{\theta}_{\mathrm{quad}}, \mathbf{y})/L(\hat{\theta}_{\mathrm{const}}, \mathbf{y})$ via bridge sampling, see Section II. For this purpose, we bridge linearly between $\kappa_1 = 0$ and $\kappa_1 = 2$ in steps of size 0.125. Also for the other parameters, we perform linear interpolation. Figure 13 shows the estimated log likelihood ratios $\log(L(\theta_{i+1}, \mathbf{y})/L(\theta_i, \mathbf{y}))$ for the intermediate parameter vectors together with the associated effective sample sizes. Aggregating the estimates yields a negative value for the estimated log likelihood ratio for the quadratic rate in comparison to the constant rate model. Hence,
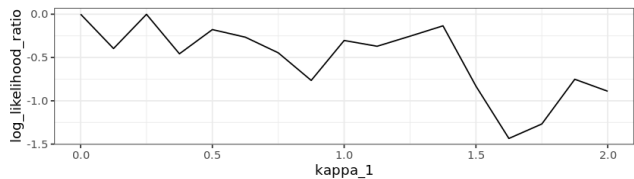
basic assumption of a Poisson point process. Therefore, they do not allow for interactions between the different arrivals. However, a commonly held belief is that arrivals tend to appear in clusters, which would call for a more flexible point-process model. By extending the model selection methods from the present paper, MCMC-based maximum likelihood estimation makes it possible to analyze this belief on the grounds of a statistically well-established methodology.

REFERENCES

[1] H. Hashemi, "Simulation of the urban radio propagation channel," *IEEE Trans. Veh. Technol.*, vol. 28, no. 3, pp. 213–225, 1979.

[2] J. Møller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, vol. 100 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[3] G. L. Turin, F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry, "A statistical model of urban multipath propagation," *IEEE Transactions on Vehicular Technology*, vol. 21, no. 1, pp. 1–9, 1972.

[4] A. A. M. Saleh and R. A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, 1987.

[5] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, "Modelling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 347–360, 2000.

[6] K. Haneda, J. Järveläinen, A. Karttunen, M. Kyrö, and J. Putkonen, "A statistical spatio-temporal radio channel model for large indoor environments at 60 and 70 GHz," *IEEE Trans. Antennas Propag.*, vol. 63, no. 6, pp. 2694–2704, 2015.

[7] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1445–1455, 2014.

[8] J. Poutanen, K. Haneda, L. Liu, C. Oestges, F. Tufvesson, and P. Vainikainen, "Parameterization of the COST 2100 MIMO channel model in indoor scenarios," in *EuCAP*, pp. 3606–3610, 2011.

[9] P. Kyösti, *WINNER II channel models, deliverables D1.1.2 V1.2, part I: Channel models*.

[10] L. Raschkowski, P. Kyösti, K. Kusume, and E. T. Jämsä, *METIS channel models, deliverable D1.4 V3*.

[11] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl. Ser.*, vol. 15, pp. 417–426, 1974.

[12] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. Pedersen, "Channel parameter estimation in mobile radio environments using the sage algorithm," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 434–450, 1999.

[13] A. Richter, M. Landmann, and R. Thomä, "Rimax - a maximum likelihood framework for parameter estimation in multidimensional channel sounding," in *Proc. ISAP, Sendai, Japan*, pp. 53–56, 2004.

[14] R. J.-M. Cramer, R. A. Scholtz, and M. Z. Win, "Evaluation of an ultra-wide-band propagation channel," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 561–570, 202.

[15] C.-C. Chong and S. K. Yong, "A generic statistical-based UWB channel model for high-rise apartments," *IEEE Trans. Antennas Propag.*, vol. 53, no. 8, pp. 2389–2399, 2005.

[16] N. Czink, P. Cera, J. Salo, E. Bonek, J. P. Nuutinen, and J. Ylitalo, "A framework for automatic clustering of parametric MIMO channel data including path powers," in *Proc. IEEE 64th Veh. Technol. Conf.-Fall*, pp. 1–5, 2006.

[17] C. Gentile, "Using the kurtosis measure to identify clusters in wireless channel impulse responses," *IEEE Trans. Antennas Propag.*, vol. 61, no. 6, pp. 3392–3396, 2013.

[18] R. He, W. Chen, B. Ai, A. F. Molisch, W. Wang, Z. Zhong, J. Yu, and S. Sangodoyin, "On the clustering of radio channel impulse responses using sparsity-based methods," *IEEE Trans. Antennas Propag.*, vol. 64, no. 6, pp. 2465–2474, 2016.

[19] R. O. Adeogun, "Calibration of stochastic radio propagation models using machine learning," *IEEE Antennas and Wireless Propagation Letters*, pp. 1–1, 2019.

[20] A. Bharti, R. Adeogun, and T. Pedersen, "Estimator for stochastic channel model without multipath extraction using temporal moments," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications, submitted*, 2019.

[21] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter estimation for stochastic channel models using temporal moments," in *IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, 2019.

[22] A. Bharti and T. Pedersen, "Calibration of stochastic channel models using approximate bayesian computation," in *Global Commun. Conf., IEEE*, 2019.

[23] A. Bharti, R. Adeogun, and T. Pedersen, "Learning parameters of stochastic radio channel models from summaries," *IEEE Open Journal of Antennas and Propagation*, vol. 1, pp. 175–188, 2020.

[24] C. J. Geyer, "Markov chain Monte Carlo maximum likelihood," in *Computing Science and Statistics: Proceedings of 23rd Symposium on the Interface Interface Foundation, Fairfax Station, 1991*, pp. 156–163, Interface Foundation of North America, 1991.

[25] W. Fan, I. Carton, J. Ø. Nielsen, K. Olesen, and G. F. Pedersen, "Measured wideband characteristics of indoor channels at centimetric and millimetric bands," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 58, 2016.

[26] S. Salous, P. Filippidis, R. Lewenz, I. Hawkins, N. Razavi-Ghods, and M. Abdallah, "Parallel receiver channel sounder for spatial and MIMO characterisation of the mobile radio channel," *IEE Proceedings-Communications*, vol. 152, no. 6, pp. 912–918, 2005.

[27] G. Steinböck, T. Pedersen, B. H. Fleury, W. Wang, and R. Raulefs, "Distance dependent model for the delay power spectrum of in-room radio channels," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 8, pp. 4327–4340, 2013.

[28] T. Pedersen, "Modelling of path arrival rate for in-room radio channels with directive antennas," vol. 66, pp. 4791–4805, Sept. 2018.

[29] T. Pedersen, "Stochastic multipath model for the in-room radio channel based on room electromagnetics," vol. 67, pp. 2591–2603, Apr. 2019.

[30] J. Kunisch and J. Pamp, "Measurement results and modeling aspects for the UWB radio channel," in *2002 IEEE Conference on Ultra Wideband Systems and Technologies*, pp. 19–23, 2002.

[31] T. Pedersen and B. H. Fleury, "A realistic radio channel model based in stochastic propagation graphs," in *5th Conf. on Mathematical Modelling (MATHMOD 2006)*, pp. 324–331, 2006.

[32] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of reverberant radio channels using propagation graphs," *IEEE Trans. Antennas Propag.*, vol. 60, no. 12, pp. 5978–5988, 2012.

[33] C. J. Geyer, "Likelihood inference for spatial point processes," in *Stochastic Geometry: Likelihood and Computation* (O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, eds.), pp. 79–140, Boca Raton, Florida: Chapman & Hall/CRC, 1999.

[34] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statistics and Computing*, vol. 6, no. 2, pp. 113–119, 1996.

[35] C. J. Geyer, "Practical Markov chain Monte Carlo," *Statist. Sci.*, vol. 7, pp. 473–483, 11 1992.

[36] G. O. Roberts, A. Gelman, and W. R. Gilks, "Weak convergence and optimal scaling of random walk Metropolis algorithms," *Ann. Appl. Probab.*, vol. 7, no. 1, pp. 110–120, 1997.

[37] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, and P. L'Ecuyer, "The cross-entropy method for optimization," in *Handbook of Statistics* (C. R. Rao and V. Govindaraju, eds.), pp. 35 – 59, Elsevier, 2013.

[38] C. J. Geyer and E. A. Thompson, "Constrained Monte Carlo maximum likelihood for dependent data," *J. Roy. Statist. Soc. Ser. B*, vol. 54, no. 3, pp. 657–699, 1992.
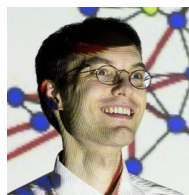
Christian Hirsch received the diploma degree in mathematics from the LMU Munich, Germany in 2010, and the PhD degree in mathematics from Ulm University, Germany in 2014. He is currently Assistant Professor for topological data analysis at the University of Groningen. Before that he held post-doctoral positions at the WIAS Berlin, the LMU Munich and at Aalborg University. He was also a Juniorprofessor at the University of Mannheim. His research interests include spatial statistics, topological data analysis, statistical signal processing and applied probability.

Ayush Bharti received the B.E. degree in Electrical and Electronics Engineering in 2015 from Birla Institute of Technology and Sciences, Pilani, India. He received his M.Sc. degree in Signal Processing and Computing from Aalborg University, Denmark, in 2017. Currently, he is a Ph.D. Fellow at the Department of Electronic Systems, Aalborg University. His research interests include statistical signal processing, radio channel modeling, and Bayesian inference.

Troels Pedersen received the M.Sc. degree in digital communications and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2004 and 2009, respectively. In 2005, he was a Guest Researcher with the FTW Telecommunications Research Center Vienna, Vienna, Austria. He joined the Department of Electronic Systems, Aalborg University, as an Assistant Professor, in 2009, and became an Associate Professor in 2012. In 2012, he was a Visiting Professor with the Institut dlectronique et de Tlcommunications de Rennes, University of Rennes 1, Rennes, France. His current research interests include statistical signal processing and communication theory, including sensor array signal processing, radio geolocation techniques, radio channel modeling, and radio channel sounding. Dr. Pedersen received the Teacher of the Year Award from the Study Board for Electronics and IT, Aalborg University, in 2011 and 2017.

Rasmus Waagepetersen Rasmus Waagepetersen received the MSc degree in mathematical statistics in 1995 and the PhD degree in mathematical statistics in 1997 both from Aarhus University, Aarhus, Denmark. Rasmus Waagepetersen was a Scientist at the Danish Institute of Agricultural Sciences, Tjele, from 1997 to 2000. From 2000 to 2007 he was Assistant and subsequently Associate Professor at Department of Mathematical Sciences, Aalborg University. In the period 2008-2009 he was a Credit Risk Analyst at Spar Nord Bank, Aalborg. Since 2010 he has been full Professor at Department of Mathematical Sciences, Aalborg University.