Bayesian Inference for Stochastic Multipath Radio Channel Models

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

Abstract—**Stochastic radio channel models based on underlying point processes of multipath components (MPCs) have been studied intensively since the seminal papers of Turin and Saleh–Valenzuela (SV). Despite this, inference regarding parameters of these models has remained a major challenge. Current methods typically have a somewhat ad hoc flavor involving a multitude of steps requiring user specification of tuning parameters. In this article, we propose to instead adopt the principled framework of Bayesian inference to conduct inference for the SV model. The posterior distribution is not analytically tractable and we therefore compute approximations of the posterior using Markov chain Monte Carlo (MCMC) methods specific to point processes. To demonstrate the flexibility of our approach, we additionally propose a new multipath model and apply our inference method to it. The resulting inference methodology is computationally demanding and our successful implementation relies critically on our novel MPC updates within the MCMC sampler. We demonstrate the usefulness of our approach on simulated and real radio channel data.**

Index Terms—**Bayesian inference, Markov chain Monte Carlo (MCMC) sampling, multipath propagation, parameter estimation, radio channel modeling.**

I. INTRODUCTION

STOCHASTIC multipath models, such as the one proposed by Turin et al. [1] and Saleh and Valenzuela [2], are widely used for simulating radio channels due to their simple formulation and low computational cost. In such a model, the collection of delays and gains of the multipath components (MPCs) is assumed to be a realization of a marked point process [3], [4], with the received signal being a superposition of the MPCs and additive random noise. The parameters of this underlying point process constitute the parameters of the stochastic multipath model. By adjusting these parameters, the multipath model can mimic different propagation scenarios. Usually, suitable values of the model parameters are estimated from channel measurements collected from a particular

Christian Hirsch is with the Department of Mathematics, Aarhus University, 8000 Aarhus, Denmark (e-mail: hirsch@math.au.dk).

Ayush Bharti is with the Department of Computer Science, Aalto University, 02150 Espoo, Finland (e-mail: ayush.bharti@aalto.fi).

Troels Pedersen is with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: troels@es.aau.dk).

Rasmus Waagepetersen is with the Department of Mathematical Sciences, Aalborg University, 9220 Aalborg, Denmark (e-mail: rw@math.aau.dk).

environment. Despite the simple formulation of stochastic multipath models, this is a nontrivial task since the likelihood function of the multipath model is analytically intractable. The computational problems related to the likelihood arise because the MPCs are unobservable due to the finite measurement bandwidth and additive noise.

Currently, parameters of stochastic multipath models are usually estimated using a multistep procedure as in [1], [2], [5], [6], [7], [8], and [9]. First, the MPCs are extracted from channel measurements using high-resolution algorithms such as multiple signal classification (MUSIC), RiMAX, and space alternating generalized expectation-maximization (SAGE), among others; see [10, Ch. 5] for an overview. In the case of a cluster-based model like the Saleh–Valenzuela (SV) model, the next step involves clustering the extracted MPCs. This is achieved either via manual clustering or through the use of automated clustering algorithms such as [11], [12], and [13]. As noted recently [14], [15], the multistep approach leads to a number of concerns. Despite being widespread, this approach is not trivial as it involves implementing sophisticated high-resolution and clustering algorithms. These algorithms involve certain heuristic choices and assumptions. The performance of these algorithms is sensitive to these choices, and their assumptions may conflict. A major drawback of this approach is that the performance of the overall estimator is difficult to assess. As a result, error bounds on the parameter estimates are hardly reported in the literature, and the uncertainty of the parameter settings used in, e.g., standardized channel models remains unknown.

To overcome these challenges, a host of estimation methods have been proposed recently that circumvent the need for extracting MPCs and clustering [14], [15], [16], [17], [18], [19], [20], [21]. In [16] and [18], a method of moments estimator is proposed for the SV model [2] and the Turin model [1], respectively. In [19], a multilayer perceptron is used to estimate parameters in the SV model. The methods in [14], [20], and [21] are based on approximate Bayesian computation (ABC) [22], which is a likelihood-free inference method that relies on the summary statistics of the channel measurements to approximate the likelihood function. It is possible to use automatically generated summary statistics, as shown in [21], although these did not perform as well as handcrafted summaries. In [15], a kernel-based distance metric is used in the ABC algorithm, which alleviates the need for summary statistics at the cost of choosing a kernel. Although some good choices for summaries and kernels have been proposed, these choices are nontrivial and affect the accuracy of the method.

Moreover, the accuracy of the ABC methods depends on two approximations: the Monte Carlo approximation pertaining to all the sampling methods and the posterior approximation depending on the choice of summaries or kernel functions. The magnitude of the latter effect can be assessed in simulation studies, but is in general unknown. The ABC methods are easily applicable to a wide range of stochastic models since they do not depend on the particular mathematical structure of the model, which is clearly an appealing feature to the practitioners. However, a common feature of the methods in [14], [15], [16], [17], [18], [19], [20], and [21] is that the parameters of the model are estimated without obtaining the MPCs, e.g., the delays and gains, for the particular set of measurement data. While multipath extraction is in some cases not interesting, it does provide insight into the particular set of measurement data, which is useful for some applications.

In this article, we propose to carry out approximate likelihood-based inference for the model parameters. As in [23], we apply Markov chain Monte Carlo (MCMC) techniques to handle the computational problems arising from the analytically intractable likelihood function. In the context of the Turin model, [1], [23] suggested to obtain estimates by maximizing a Monte Carlo estimate of the likelihood function, integrating out the unobserved MPCs.

In the current article, in the context of the more complex SV model [2], we instead adopt a Bayesian approach and use MCMC to approximate the analytically intractable posterior distribution. Unlike [14], [15], [16], [17], [18], [19], [20], [21], our method is general in the sense that it is not tied to a specific multipath channel model. Moreover, the proposed method is not dependent on the choice of summaries and kernels as the ABC method of [15], thus alleviating the posterior approximation arising due to them. Recasting the SV model in the framework of Bayesian hierarchical modeling further enables us to circumvent intractable likelihood computations by augmenting the posterior distribution with the MPCs that comprise an intermediate level of the model. This means that unlike the ABC method, we are also able to infer number and locations of MPCs.

In Bayesian analysis, inference is notoriously difficult when mixtures of distributions are involved and the clustered structure of the SV model imposes similar challenges. A crucial technical contribution is therefore that we introduce novel MCMC so-called split–merge updates that significantly improve the efficiency of the Monte Carlo computations relative to the MCMC scheme in [23].

Our new methodology enables us to assess and criticize the existing models. In this article, we demonstrate the shortcomings of the SV model and propose a new model that may be a promising candidate for providing a better fit to real indoor radio channel data.

The rest of the article is organized as follows. Section II presents the signal model of radio channel measurements, along with the description of the SV model. The proposed modified version of the SV model is presented in Section II-C. The proposed Bayesian inference method and the MCMC sampling procedure are described in Sections III and IV, respectively. In Section V, we conduct a simulation study to assess the performance of the proposed inference method. The method is applied to measured data in Section VI, and the concluding remarks are given in Section VII.

II. Model for Radio Channel Data

A. Signal Model

As in [23], we 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. In each measurement run, the transfer function is sampled at K equispaced frequencies $f_k = (k-1)\Delta f$, $k = 1, \ldots, K$, where $\Delta f = B/(K-1)$ is the frequency spacing between two measurement points, giving the period $\tau_{\max} = 1/\Delta f$ of the time-domain signal. The measurement data are 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 f_k, and N_k denotes the measurement noise. The noise variables $N_1, \ldots, N_K$ are assumed to be independent and identically distributed circular symmetric zero-mean Gaussian 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)

$$Y(t) = \frac{1}{K} \sum_{k=1}^{K} Y_k \exp(i2\pi f_k t) \tag{2}$$

where i denotes the imaginary unit.

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

$$H_k = H_k(Z_{\mathsf{m}}) = \sum_{\tau \in Z} \alpha_\tau \exp(-i2\pi f_k \tau), \quad k = 1, \ldots, K \tag{3}$$

where $Z_{\mathsf{m}} = \{(\tau, \alpha_\tau)\}_{\tau \in Z}$ constitutes a *marked* point process on $\mathbb{R}_+ \times \mathbb{C}$. The point process Z contains the *propagation time delays* τ. The mark associated with a delay τ is the complex-valued *gain* α_τ. The support of the point process Z is the interval $I = [\tau_0, \tau_{\max}]$, where τ_0 is the delay of the line-of-sight (LOS) path.

For the Turin model [1], Z is assumed to be an inhomogeneous Poisson process. That is, delays τ essentially appear independently of each other. This may not be realistic as clusters of delays may arise where each cluster is due to reflections from a particular environmental element. Saleh and Valenzuela [2] instead suggested a more elaborate model for Z, which according to [3] and [4] can be viewed as a clustered point process as detailed in the following.

B. SV Model

We let $C = \{t_0, t_1, \ldots\}$ denote the arrival times of clusters of MPCs. For the SV model, $t_0 = \tau_0$ and the remaining arrival times $\{t_1, t_2, \ldots\}$ form a homogeneous Poisson process of

intensity[1] $\lambda_0 = \exp(\kappa_0)$ on $[\tau_0, \infty)$. To each t_i, $i \geqslant 0$, a cluster $X_{t_i} = \{\tau_{i,1}, \tau_{i,2}, \ldots\}$ is associated, where conditional on C, X_{t_i}, $i \geqslant 0$, are independent and each X_{t_i} is a Poisson process on $[t_i, \infty)$ of intensity $\lambda_1 = \exp(\kappa_1) > 0$. Let $X = \cup_{i \geqslant 0} X_{t_i}$. Then $Z = C \cup X$ and $Z_{\mathsf{m}} = C_{\mathsf{m}} \cup X_{\mathsf{m}}$ where C_{m} and X_{m} are the marked point processes obtained by attaching gains α_{t_i} or $\alpha_{\tau_{i,k}}$ as marks to each point in C or X. Following terminology in spatial statistics, we will often refer to points in C and X as "parents" and "offspring," respectively.

Conditional on the delays Z, the gains are assumed to be independent zero-mean complex Gaussian random variables with a phase that is uniform on $[0, 2\pi)$. Thus, the magnitudes of the gains are Rayleigh distributed. Following Saleh and Valenzuela [2], we assume that their second moments satisfy:

$$\mathbb{E}[|\alpha_{t_i}|^2 \mid Z] = \exp(\gamma_0 + \gamma_1 t_i)$$
$$\mathbb{E}[|\alpha_{\tau_{i,k}}|^2 \mid Z] = \exp(\gamma_0 + \gamma_1 t_i + \gamma_2(\tau_{i,k} - t_i)) \quad (4)$$

for some parameters $\gamma_0, \gamma_1, \gamma_2 > 0$.

As noted in [3], the cluster process X is an example of a Cox point process [24]. This is because conditional on C, X is a Poisson process with intensity function

$$\Lambda(\tau) = \sum_{t \in C} \lambda_1 k_t(\tau) \quad (5)$$

where $k_t(\tau) = \mathbb{1}\{\tau \geqslant t\}$. Here, the indicator function $\mathbb{1}\{A\}$ is equal to 1 if event A holds. Hence, evaluation of the conditional distribution of X given C does not require information regarding the parent–offspring relationships. This simplifies greatly the construction in Section IV of an MCMC sampler for simulation of C and X given the radio channel data $\mathbf{Y}$. In particular, we can modify the MCMC sampler from [23] to sample X given $\mathbf{Y}$ and C.

However, with $\gamma_1 \neq \gamma_2$, the formulation of the gain distribution in (4) presents an obstacle for exploiting the aforementioned simplification. Evidently, unless $\gamma_1 = \gamma_2$, the evaluation of the right expression in (4) requires knowledge of the parent t_i for the offspring $\tau_{i,k}$. In the interest of computational feasibility, we have chosen to restrict attention to the case $\gamma_1 = \gamma_2$. In the data example in Section VI, we in fact obtain a very good fit assuming $\gamma_1 = \gamma_2$ indicating that this is not a severe restriction for the considered data.

C. Generalizations and Modifications of the SV Model

The physical interpretation of the SV model is not obvious and is still being discussed in the literature [25], [26]. In particular, the parent point process C is not necessarily Poisson and may be generalized in numerous ways. For example, C could be replaced by a regular point process to ensure a desired degree of separation between the parent times in C. In this article, we consider the case where C is the so-called Strauss process and coin the resulting model the Strauss–SV model.

For a specific parent configuration c, the Strauss point process density is of the form (see [24])

$$f(c) = \frac{1}{Z(\kappa_0, \psi, R)} \exp(\kappa_0 n(c)) \psi^{s_R(c)} \quad (6)$$

where $n(c)$ is the number of parents in c, $s_R(c)$ is the number of pairs of parents within distance $R > 0$ from each other, $0 \leqslant \psi \leqslant 1$ are the parameters, and $Z(\kappa_0, \psi, R)$ is a normalizing constant. Choosing a small ψ means that there is a low probability of seeing parents close to each other. With $\psi = 0$, a so-called hard core process is obtained where parents can never be closer than R. A Poisson process of intensity $\exp(\kappa_0)$ is obtained with $\psi = 1$.

A complicating feature of the Strauss model is that the normalizing constant $Z(\kappa_0, \psi, R)$ is an analytically intractable function of κ_0, ψ, and R. For any fixed (κ_0, ψ, R), it can be evaluated using the MCMC methods; see [24]. However, this is computationally demanding, and in our Bayesian setting we would need to recompute $Z(\kappa_0, \psi, R)$ each time an update is proposed for κ_0, ψ, or R. This presents a major computational problem; see also the discussion in Section VII. However, for any fixed (κ_0, ψ, R), the simulations of the Strauss process can easily be obtained by a simple birth–death MCMC sampler [24] that does not require knowledge of $Z(\kappa_0, \psi, R)$. Hence, simulation of Strauss–SV does not present a problem.

Furthermore, in (5), the kernel function $k_t(\cdot)$ can be replaced by an arbitrary positive function instead of the indicator function used for the SV model. For the SV model, X is an infinite point process. However, due to fast decay of the gains, gains are negligible when their delays are far from the LOS delay τ_0. This suggests replacing $k_t(\tau) = \mathbb{1}\{\tau \geqslant t\}$ in (5) by a rectangular kernel $k_t(\tau) = w^{-1}\mathbb{1}\{t \leqslant \tau \leqslant t + w\}$ of kernel width $w > 0$. By similar considerations, we also restrict C to the finite interval $[\tau_0, \tau_{\max}]$. Then X becomes a finite point process on $I_w = [\tau_0, \tau_{\max} + w]$ so that it can be stored on a computer and is amenable to MCMC computations. Moreover, for single-point MCMC updating of the parents C, using a k_t of bounded support simplifies computations because removing or inserting a point τ in C only affects a finite number of MPCs in X.

According to our simulation studies, once the kernel width w is chosen sufficiently large (relative to the decay of the gains), the results obtained are representative for the original SV model.

III. BAYESIAN INFERENCE FOR THE SV MODEL

To calibrate the SV model parameters $\boldsymbol{\theta} = (\gamma_0, \gamma_1, \kappa_0, \kappa_1, \sigma^2)$ to data, we use Bayesian inference assigning a prior density $p(\boldsymbol{\theta})$ to $\boldsymbol{\theta}$. Alternatively, one could use maximum likelihood estimation; see [23]. However, there are decisive benefits of Bayesian inference in the context of the SV model as follows.

1) *Prior Information:* The most obvious advantage of Bayesian inference is of course that it allows use of prior information e.g., obtained from physical considerations or previous experience with radio channel models in similar settings. Prior information is especially useful in the context of complex stochastic models such as the SV model.

2) *Parameter Identifiability:* Even for the simpler Turin model, several parameter configurations can lead to local maxima of very similar quality; see [23]. This

---

[1]The term "intensity" is widely used in the theory of spatial point processes. In propagation studies, the term "arrival rate" is often used instead.

issue is amplified for the more complex SV model. However, maximum likelihood estimation focuses only on the global maximum and ignores all the local modes. In contrast, in a Bayesian analysis, the different maxima are visible in a multimodal posterior distribution.

3) *Optimization Instability:* The aforementioned multi-modality is obviously a challenge for optimization of the likelihood function. Furthermore, for the Turin or SV model, the likelihood function or its derivatives must be approximated using Monte Carlo methods [23]. This can render the optimization process numerically unstable and the final result can be quite sensitive to Monte Carlo errors. In contrast, Monte Carlo approximation of posterior means and standard deviations is numerically more stable.

4) *Bayesian Hierarchical Modeling and Posterior Augmentation:* The stochastic multipath models can be viewed as hierarchical models with $\boldsymbol{\theta}$ as the upper level, the MPC point process Z_m at the intermediate level, and the data $\boldsymbol{y}$ at the lower and final level. Computing the likelihood $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ requires marginalization with respect to Z_m which is analytically intractable. Hence, direct evaluation of the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$ is not feasible. In contrast, the augmented likelihood $p(\boldsymbol{y}, c_m, x_m \mid \boldsymbol{\theta})$ including the parents c_m and offspring x_m in Z_m is analytically tractable; see Section III-A. We can therefore explore the joint posterior $p(\boldsymbol{\theta}, c_m, x_m \mid \boldsymbol{y}) \propto p(\boldsymbol{y}, c_m, x_m \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$ and hence also $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ using MCMC (see Section IV). More specifically, if $\{(\boldsymbol{\theta}_i, c_{mi}, x_{mi})\}_{i=1}^M$ represents a sample from the joint posterior $p(\boldsymbol{\theta}, c_m, x_m \mid \boldsymbol{y})$, then obviously $\{\boldsymbol{\theta}_i\}_{i=1}^M$ is a sample from the marginal posterior $p(\boldsymbol{\theta} \mid \boldsymbol{y})$. The joint posterior further informs about the unobserved MPCs which are of interest in their own right.

A. Augmented Likelihood

In the following, we derive $p(\boldsymbol{y}, c_m, x_m \mid \boldsymbol{\theta})$. First, $p(\boldsymbol{y}, c_m, x_m \mid \boldsymbol{\theta})$ factorizes as

$$p(\boldsymbol{y}, c_m, x_m \mid \boldsymbol{\theta}) = p(\boldsymbol{y} \mid z_m, \sigma^2)p(x_m \mid c_m, \kappa_0, \gamma)p(c_m \mid \kappa_0, \gamma). \tag{7}$$

The first factor on the right-hand side is the Gaussian probability density for the channel response data given the MPCs, the second factor is the conditional density of the offspring MPCs x_m given c_m and the parameters κ_0 and $\gamma = (\gamma_0, \gamma_1)$, and the third factor is the density of c_m given κ_0 and γ.

More precisely, the complex Gaussian density is given by

$$p(\boldsymbol{y} \mid z_m, \sigma^2) = (2\pi\sigma^2)^{-K} \exp\left(-\frac{1}{2}\sum_{k=1}^{K}\left(|y_k - H_k(z_m)|/\sigma\right)^2\right) \tag{8}$$

where we recall that $\sigma^2 > 0$ is the noise variance, and $H_k(z_m)$ is the complex-valued channel transfer function (3) evaluated at the frequency f_k.

Second, technically, the densities $p(x_m \mid c_m, \kappa)$ and $p(c_m \mid \gamma)$ are to be understood as densities of the offspring and

parent MPCs with respect to unit-intensity Poisson processes on I [24, Sec. 6.1]. Since the parents C_m form an independently marked homogeneous Poisson point process with intensity $\exp(\kappa_0)$, we deduce from [24, Sec. 3.3] that its density is given by

$$p(c_m \mid \kappa_0, \gamma) = \prod_{\tau \in c} f(\alpha_\tau \mid \gamma) \times \exp\left(\kappa_0 \# c + |I|(1 - \kappa_0)\right) \tag{9}$$

where $f(\alpha_\tau \mid \gamma)$ is the complex Gaussian density for the channel gains, $\#c$ is the cardinality of the set c, and $|I|$ is the length of the interval I.

Concerning the density of the offspring MPC, we use that conditional on C_m, X is a Poisson point process of intensity Λ given in (5). Hence, the density of X_m becomes

$$p(x_m \mid c_m, \kappa_1, \gamma)$$
$$= \prod_{\tau \in x} f(\alpha_\tau \mid \gamma) \times \exp\left(|I| - \int_I \Lambda(t)\mathrm{d}t\right) \times \prod_{\tau \in x} \Lambda(\tau) \tag{10}$$

with κ_1 entering through $\Lambda(\cdot)$; see (5).

B. Strauss–SV Model

For the Strauss–Valenzuela model, complications arise from the intractable normalizing constant of the Strauss point process density. The problem of conducting Bayesian inference in cases where the prior or likelihood involves an intractable normalizing constant has been studied in several papers, e.g., [27], [28], [29], [30]; see also the review [31]. Unfortunately, the methods proposed tend to add considerable additional computational complexity to an already computationally challenging problem. However, recent results [32] in the context of mixture models suggest that this approach may be worth exploring in future work on channel data inference.

For simplicity, we do not attempt efficient sampling of the Strauss process at this stage. Instead, leaving such refinements to future work, we assume that the user is able to specify fixed sensible values of κ_0, ψ, and R reflecting prior beliefs of the parent MPCs. For the remaining parameters, we proceed as for the SV model.

IV. MCMC SAMPLING OF THE POSTERIOR

The posterior $p(\boldsymbol{\theta}, c_m, x_m \mid \boldsymbol{y})$ is a high-dimensional and complex density and only known up to a constant of proportionality. We therefore apply the MCMC methods to sample from the posterior. Briefly, an MCMC algorithm generates an ergodic Markov chain whose equilibrium distribution is the target distribution of interest—in our case the posterior density $p(\boldsymbol{\theta}, c_m, x_m \mid \boldsymbol{y})$. After a suitable burn-in, the states of the Markov chain represent a correlated sample from the target distribution. We say that the Markov chain is mixing fast if the correlation between consecutive states is small. The Markov chain is constructed using the Metropolis–Hastings scheme where in each step of the Markov chain, updates are suggested for one or more components of the posterior. These updates are then accepted or rejected according to the Metropolis–Hastings ratio.

A complicating feature of our posterior is that the point process components c_m and x_m are not of fixed dimension.

Therefore, specialized Metropolis–Hastings updates are needed, as outlined in [24] and implemented for radio channel data in [23]. More specifically, [23] used a birth-death–move algorithm to sample the conditional distribution of MPCs given radio channel data. We refine this algorithm by adding split–merge steps and also extend the algorithm by adding updates for $\boldsymbol{\theta}$. Below we give a brief summary of the different types of updates used. We refer to [23] for a detailed description of MCMC for sampling MPCs, where several MCMC chains were run using a parallel tempering scheme to improve mixing. However, the introduction of split–merge steps improves mixing considerably so that the further computational task of using parallel tempering can be avoided.

A. Updates of MPCs

Given a current state $(\boldsymbol{\theta}, c_m, x_m)$, we propose a new state for the MPC components using a birth/death proposal, a move, or a split/merge proposal. These three proposals are selected each with probability 1/3. These probabilities were selected after a few pilot runs. We believe the performance of the MCMC algorithm is fairly robust to the choice of these probabilities as long as extreme values close to 0% or 100% are avoided. Henceforth, we describe in detail how to proceed for updating the MPC components x_m. For the parent MPCs, we proceed similarly, except that we do not perform a split/merge step.

1) Birth/Death: In case of a birth/death step, with probability a half, we propose a birth, i.e., to add a new MPC to x_m. The delay τ of the proposed new MPC is sampled on I from a density proportional to the intensity (5). Next the associated gain α_τ is sampled from the symmetric Gaussian density $f(\alpha_\tau \mid \gamma)$. Otherwise, we propose a death, where a current MPC is selected at random for removal.

2) Move: Relying exclusively on birth/death steps leads to a very slowly mixing MCMC sampler. To move from an MPC configuration to another which differs only by a minor modification of a single MPC, this MPC would have to be first removed and then reinserted slightly changed in a subsequent birth step. If the configuration between these two steps is highly unlikely a posteriori, the MCMC sampler will rarely explore this option. We therefore introduce a move step that simply proposes to update the delay or path gain of an MPC without changing the dimensionality.

3) Split/Merge: An MCMC sampler relying solely on birth/death and move steps may get stuck in configurations where two MPCs (τ, α_τ) and $(\tau', \alpha_{\tau'})$, with τ close to τ', represent the contribution of a single underlying MPC in the true unobserved MPC configuration. For any $k = 1, 2, \ldots$, $f \in \mathbb{R}$, if τ is close to τ'

$$\alpha_\tau \exp(-i2\pi \Delta f k\tau) + \alpha_{\tau'} \exp(-i2\pi \Delta f k\tau') \approx$$
$$(\alpha_\tau + \alpha_{\tau'}) \exp(-i2\pi \Delta f k\tau). \quad (11)$$

If the right-hand side represents well the contribution of a true MPC, then removal of any of (τ, α_τ) or $(\tau', \alpha_{\tau'})$ will be highly unlikely. Hence, the MCMC algorithm will not be able to explore the configuration where a single MPC replaces the two current MPCs.

To deal with this issue, we introduce a split/merge step. The split–merge steps have previously been considered in the context of Bayesian inference for mixture models or cluster point processes [33], [34], [35] where groups of observations are split or merged according to the underlying mixture component memberships. Our split–merge step is quite different since it pertains splitting one MPC into two MPCs or merging two close MPCs into one. More precisely, we allow an MPC (τ, α_τ) to merge with its nearest neighbor $(\tau', \alpha_{\tau'})$ so as to form a new MPC with magnitude given by $|\alpha_\tau| + |\alpha_{\tau'}|$ and located at $(\tau + \tau')/2$. We found it most effective to sample the phase of the new component at random. Note that from a conceptual point of view, it would have also been possible to define a Metropolis–Hasting step by updating the magnitude as $|\alpha_\tau + \alpha_{\tau'}|$. However, in the implementation the chosen approach made it easier to treat the magnitude and the phase separately.

To ensure reversibility of the MCMC sampler, the merge step is complemented by a split step, where a single MPC can split into two, where the magnitudes of the resulting path gains add up to the magnitude of the path gain of the original MPC and the delays of the two new MPCs are generated in a neighborhood of the original MPC. In case of a split–merge proposal, a split step or a merge step is chosen at random each with probability of half. The details of the split–merge Metropolis-Hastings ratios are given in the Supplementary, along with a block diagram of the structure of the MPC MCMC updating scheme.

B. Remaining Details of the MCMC Sampler

We use simple random walk Metropolis updates for the components of $\boldsymbol{\theta}$. With the model specification (4), the parameters γ_0 and γ_1 tend to be strongly negatively correlated which hampers the mixing of the MCMC sampler. To mitigate this, we found it helpful to reparameterize

$$\gamma_0 + \gamma_1 \tau = (\gamma_0 + \gamma_1 t_m) + \gamma_1(\tau - t_m).$$

Replacing the original parameter γ_0 by

$$\bar{\gamma}_0 = \gamma_0 + \gamma_1 t_m$$

where t_m is the center of the observation interval. This reparameterization is akin to centering the covariate of a linear regression.

Each step of the MCMC sampler comprises an update of the offspring MPC (using birth–death, move, or split–merge). The Metropolis–Hastings steps updating the parent MPCs are expensive since they entail recomputing the likelihood of both the process of parent MPCs and the process of offspring MPCs. Therefore, we update c_m only each 10th step in the sampler. The same goes for $\boldsymbol{\theta}$ since it does not seem useful to update $\boldsymbol{\theta}$ after each minor change in the point process state $x_m \cup c_m$ due to an MPC update. Hence, $\boldsymbol{\theta}$ is only updated each 25th step (updating all the components of $\boldsymbol{\theta}$).

The MCMC algorithm for the Strauss–SV model follows the specification above, except, following Section III-B, which we fix k_0, ψ, and R at prescribed values.
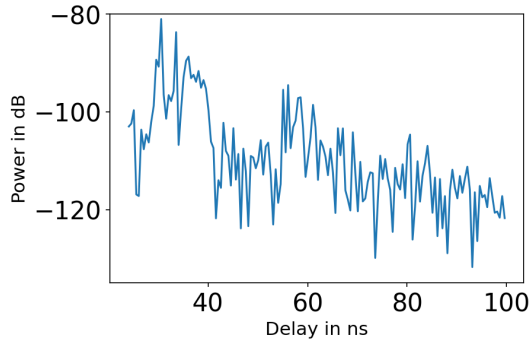
Fig. 1. Power-delay profile (magnitude of time-domain measurements (2) against time) of the indoor channel data.

C. Setting Prior Distributions

We now elaborate on the elicitation of priors. For each of the parameters $\kappa_0, \kappa_1, \bar{\gamma}_0, \gamma_1$, and σ^2, we use a uniform prior on a bounded interval with endpoints informed by the measurement settings in [36]. The time window for the data is of length 75. We expect to see on average at least 0.5 and at most 80 parents for the given time window. Similarly, the range for arrival rate of offspring is set based on the time resolution in the data, i.e., we can only observe as many offspring points as the number of time points in the window. The ranges of $\bar{\gamma}_0$ and γ_1 are set to reflect reasonable behaviors of the decay of the gains. For $\bar{\gamma}_0$, the prior range corresponds to a dynamic range of $[-108, -32$ dB] for the intercept of the power-delay spectrum, encompassing that of the measurement data shown in Fig. 1. The range of γ_1 corresponds to a drop of around 65 dB (data in Fig. 1 has a drop of around 40 dB) to a drop of 0 dB over the measurement interval. Finally, the range for σ is set by inspecting the noise level in the power-delay profile of the data.

V. SIMULATION STUDY

From a strictly Bayesian point of view, the posterior distribution obtained given the data at hand and an appropriately elicited prior is the (self-contained) solution to the statistical inference problem. Following this line of thought, frequentist properties of the inference procedure are irrelevant. Nevertheless, from a more pragmatic point of view, users may not be so sure about the priors chosen and often posterior means are used as computationally convenient point estimates of unknown parameters. We have therefore conducted a simulation study to explore in our setting the performance of MCMC approximate posterior means as point estimates.

To that end, we consider the SV model on the time interval $I = [25, 100$ ns]. We generate 100 mock datasets with on average ten cluster parents each giving rise to on average four offspring MPCs distributed according to a rectangular kernel with bandwidth 15. Furthermore, we choose the Rayleigh parameters $\bar{\gamma}_0 = -22$, $\gamma_1 = -0.1$, and a thermal noise with log-variance $\ln(\sigma^2) = -20$. We set $M = 1$ and the number of frequency points to $K = 750$ as per [36], and then perform Bayesian inference with 40 000 000 iterations of the MCMC sampler from Section III. To avoid the burn-in of the

Parameter	Prior range
κ_0	[-5.01, 0.064]
κ_1	[-0.693, 4.38]
$\bar{\gamma}_0$	[-25, -20]
γ_1	[-0.2, 0]
$\ln(\sigma)$	[-11, -9]

chain, we discard the first 20 000 000 iterations. Uniform prior distributions are set for each parameter as per Section IV-C, and the resulting prior ranges are shown in Table I.

The left panel of Fig. 2 shows the histograms for the distributions of the 100 approximate posterior means for each of the parameters $\kappa_0, \kappa_1, \bar{\gamma}_0, \gamma_1$. For the parameters κ_1 and γ_1, the true parameter value is within two standard errors from the overall estimated posterior mean. Although the true values of κ_0 and γ_0 fall outside the intervals, they are still fairly close to the corresponding overall means. We hypothesize that the deviations for κ_0 and $\bar{\gamma}_0$ are caused by the simulated datasets where there is little difference between labeling an MPC as parent or offspring, thereby making it harder to statistically identify the parameters. The general impression is that the bias of the posterior means is relatively small compared with the sampling variability of the posterior means. Moreover, the sampling distributions of the posterior means are unimodal with moderate deviations from normality.

As mentioned in Section III, it is natural to expect cross-correlations for different pairs of parameter estimates. To quantify these effects, Fig. 2 (right) shows pair plots of the posterior means for the 100 mock datasets. For instance, we can observe a substantial negative correlation between the intensity parameters κ_0 and κ_1. This indicates that it is often difficult to discriminate between parents and offspring. On the other hand, upon reparametrization the posterior means of $\bar{\gamma}_0$ and γ_1 are essentially uncorrelated. Since substantial correlations often translate to higher rejection rates in the MCMC sampler, this decorrelation helps accelerate the procedure. The Bayesian inference for one dataset takes roughly 10 h on an Intel Xeon 2.5 GHz with 1 GB allocated RAM running on a CentOS Linux distribution.

Studying further frequentist properties of our procedure, we considered coverage of 90% credibility intervals by considering the percentages of simulations where the true parameters fall outside the credibility intervals. For the parameters $\kappa_0, \kappa_1, \bar{\gamma}_0, \gamma_1$, and $\ln(\sigma^2)$, this happened in 17%, 36%, 25%, 11%, and 14% of the simulations, respectively, showing that the intervals in general have less coverage than the nominal 90%. Bayesian credibility intervals are in general not guaranteed to attain nominal frequentist coverage so these results are not indicative of a failure of our method. According to the Bernstein–von Mises theorem, nominal frequentist coverage may be obtained asymptotically by including more data in the inference. However, it is not sufficient to increase the number of observations by just decreasing the sampling frequency Δf since we still observe just one single signal realization. Instead, one should aim for increasing the
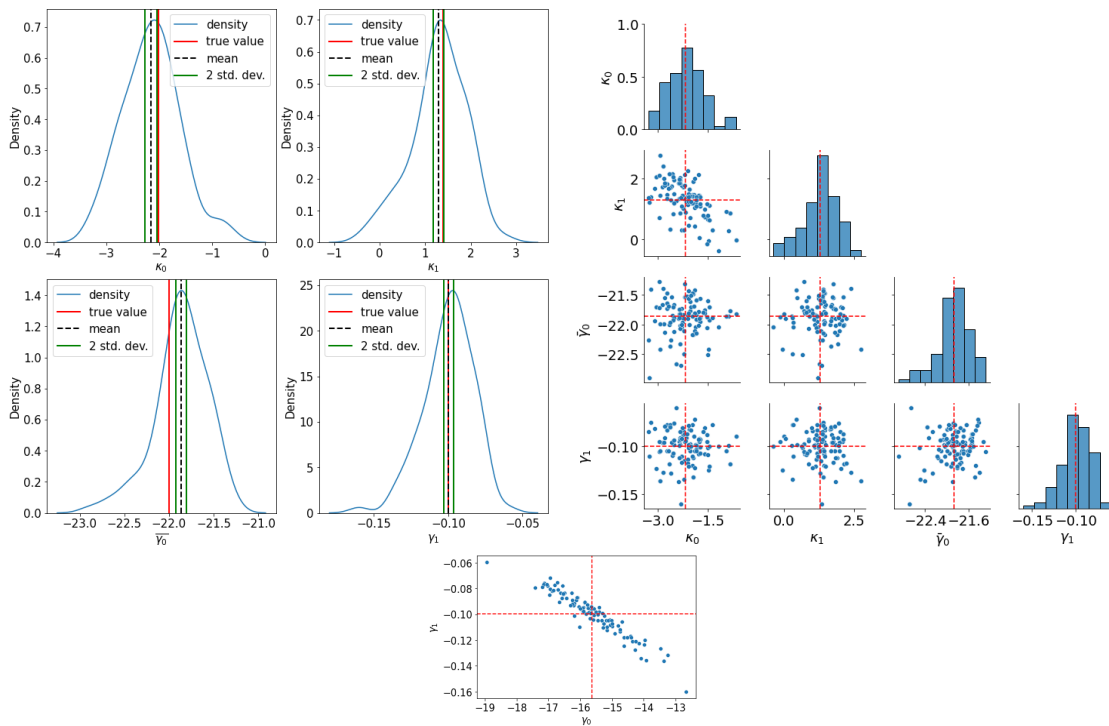
Fig. 2. Histograms for the distributions of the MCMC approximate posterior means of the parameters $\kappa_0, \kappa_1, \bar{\gamma}_0, \gamma_1$ in the 100 simulation runs (left); pair plots of the posterior means in the 100 simulation runs (right). The red vertical lines indicate the true parameter values and the black vertical dashed lines show the overall mean of the 100 approximate posterior means. The green lines show the boundaries of 95% (frequentist) confidence intervals for the true posterior mean obtained from the sample of the 100 approximate posterior means for each parameter considered. The confidence intervals are obtained in a standard way as overall mean $\pm 2 \times$ the standard error of the overall mean. The lower row shows the correlation between posterior means of γ_0 and γ_1 without reparametrization.

number M of independent signal realizations. Alternatively, to obtain better finite sample frequentist coverage, one may explore more sophisticated (although challenging to implement) noninformative priors as discussed in [37].

We end this section by illustrating the effectiveness of the split/merge step. Considering first the Turin model, we generate 150 randomly scattered MPCs on the time interval $I = [25, 100 \text{ ns}]$. We again take $K = 750$ together with Gaussian noise of variance $\sigma^2 = \exp(-25)$. Then, we run the MCMC sampler for 500 000 steps with a random initial MPC structure to sample the conditional distribution of the MPCs given the simulated radio channel data. Fig. 3 (top) shows MCMC samples for exploring the posterior distribution of the number of MPCs under the Turin model where the samples are generated with and without split–merge steps. It appears that the posterior mean of the number of MPCs is higher than the true number of MPCs for the particular simulated dataset considered. The important point, however, is that the convergence to the equilibrium distribution around the posterior mean is faster when split–merge steps are used.

For the SV model, the conclusion is similar. We keep the basic simulation setting from the Turin model, but now with MPCs organized in ten clusters each consisting of 15 offspring. MCMC samples of the number of MPCs under the SV model are shown in Fig. 3 (bottom). The improvement in terms of convergence speed when using split–merge steps is even more pronounced than for the Turin model.
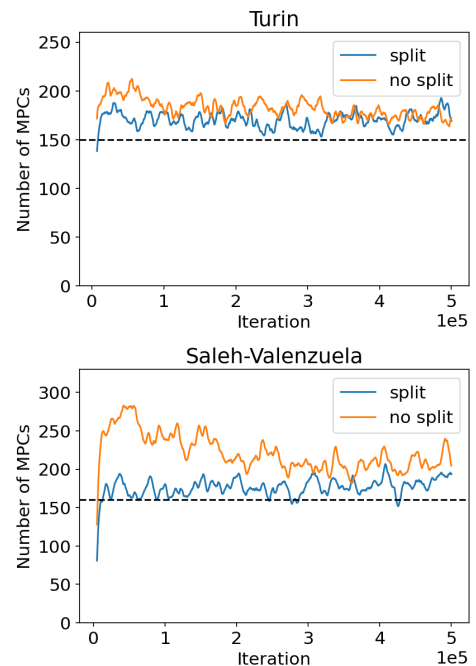

Fig. 3. Total number of MPCs in an MCMC sampler with and without the split/merge step in the Turin model (top) and the SV model (bottom). The dashed line shows the true number of MPCs.

VI. APPLICATION TO MEASUREMENT DATA

We now analyze real indoor channel data from [36]. The data consist of the channel response for $K = 750$ equally

TABLE II
POSTERIOR MEANS (STANDARD DEVIATIONS) FOR FIT MODEL PARAMETERS

Model	κ_0	κ_1	$\bar{\gamma}_0$	γ_1	$\ln(\sigma)$
Turin	0.83 (0.15)	—	-24.7 (0.20)	-0.107 (0.0074)	-10.59 (0.02)
SV	-1.23 (0.36)	2.29 (0.27)	-23.6 (0.233)	-0.109 (0.010)	-10.59 (0.02)
Strauss-SV	—	2.49 (0.20)	-23.58 (0.18)	-0.11 (0.0080)	-10.59 (0.02)
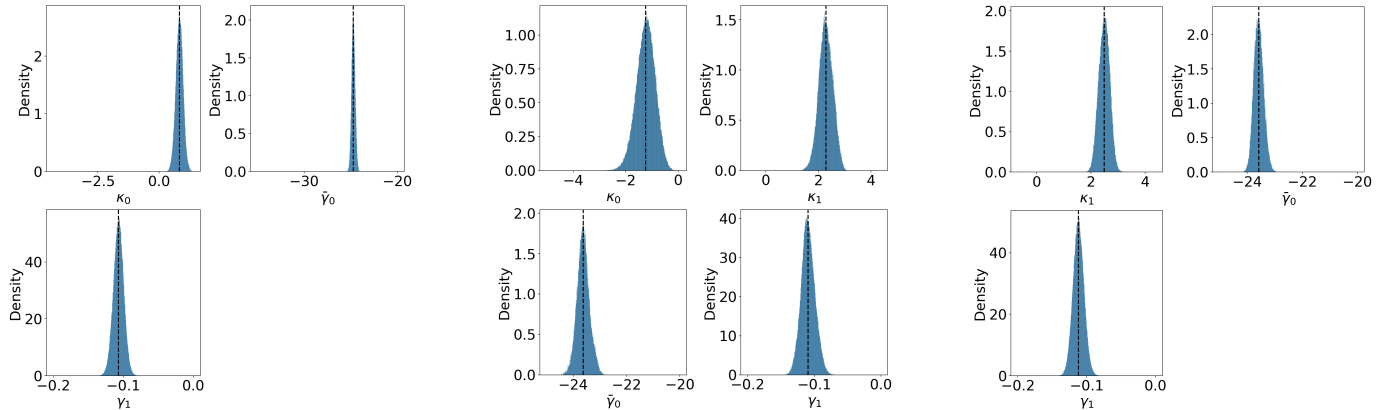
Fig. 4. Posterior distributions of the parameters for the Turin model (left), SV model (middle), and the Strauss–SV model (right). The dashed lines indicate the posterior means. The x-axis are the prior ranges.
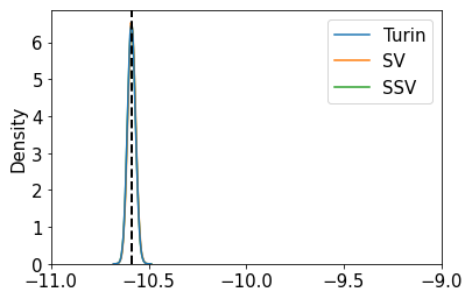
Fig. 5. Posterior distribution of $\ln(\sigma)$ in the three models. The dashed line indicates the posterior mean. The x-axis is the prior range.

spaced frequencies in the range from 2 to 4 GHz. To fit the model to data, we rely on Bayesian inference as elucidated in Section V. After transforming the data into the time domain, Fig. 1 illustrates that we may restrict the attention to MPCs located in the time interval $I = [25, 100$ ns]. Beyond that, the channel impulse response becomes negligible.

We fit three different models to the data, namely, the SV model, the Strauss–SV model, and the Turin model (i.e., the SV model with $\lambda_1 = 0$). In each of the cases, the fits are based on 8×10^7 iterations of the MCMC sampler with the priors from Section IV. For the SV and Strauss–SV model, we choose a kernel width of 15 ns. For the Strauss–SV model, we fix $\psi := 0.5$ as the interaction parameter and $R := 15$ ns as the interaction radius. That is, the interaction radius is of the same magnitude as the kernel width. We also fix the intensity parameter as $\exp(\kappa_0) := 2.5$ whereby the Strauss process has on average 25 points in the sampling window I.

A. Posterior Results

The posterior distributions of the parameters obtained from the MCMC samples are illustrated in Figs. 4, 5, and Table II.

The model parameters seem well-determined by the data since the posterior distributions are close to normal and concentrated relative to the prior ranges. It is remarkable that the posterior results for the decay parameters $\bar{\gamma}_0$ and γ_1 as well as the log noise variance $\ln(\sigma^2)$ are very similar for each of the three models, indicating that these parameters have a physical interpretation not depending on the specific model choice. Moreover, the posterior distribution for κ_1, the intensity of offspring, is clearly bounded away from the lower end of the prior support. This indicates that the SV model fits the data better than the Turin model which is the limiting case of SV when κ_1 tends to $-\infty$.

Although considering the log intensities κ_i, $i = 0, 1$ are convenient for computation, they are less convenient for model interpretation. We therefore also report the posterior means of the intensities λ_i. For the Turin model, the posterior mean yields an MPC intensity of $\lambda_0 = 2.31$. For the SV model, we obtain a cluster center intensity of $\lambda_0 = 0.201$ and an intensity of $\lambda_1 = 16.57$ offspring per center. In particular, not distinguishing between centers and offspring, the posterior mean of the combined intensity $\lambda_0(1 + \lambda_1)$ is 3.68 which exceeds substantially the value in the Turin model. Finally, for the Strauss–SV model we arrive at $\lambda_1 = 10.77$.

Fig. 6 shows the posterior intensities of the locations of cluster centers (blue) and total intensity (green), i.e., the centers and offspring combined. To highlight the connection to the data, we overlay the intensities with the power-delay profile. The intensity of all the MPCs is quite similar for the three models. The blue parts of the middle and right panels illustrate how including Strauss interactions in the model leads to a clear separation of the cluster centers in the posterior distribution. The pair plots in Fig. 7 illustrate
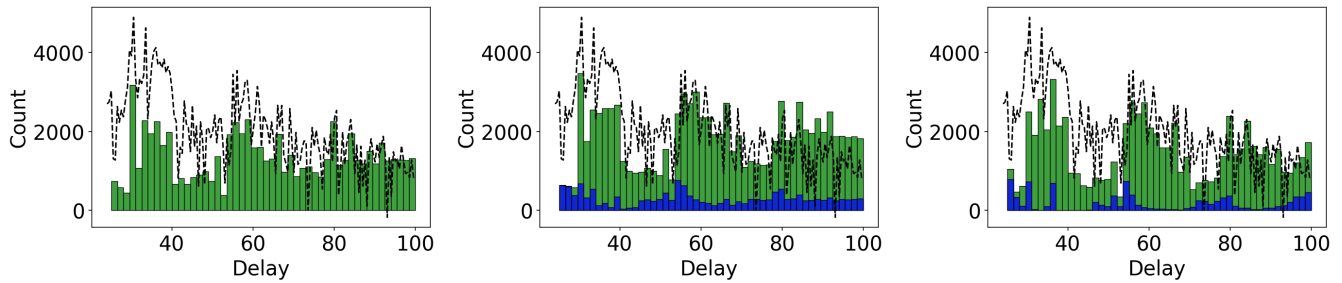
Fig. 6.   Histograms corresponding to the posterior intensity of cluster centers (blue) and all MPCs (green) for the Turin model (left), SV model (middle), and the Strauss–SV model (right).
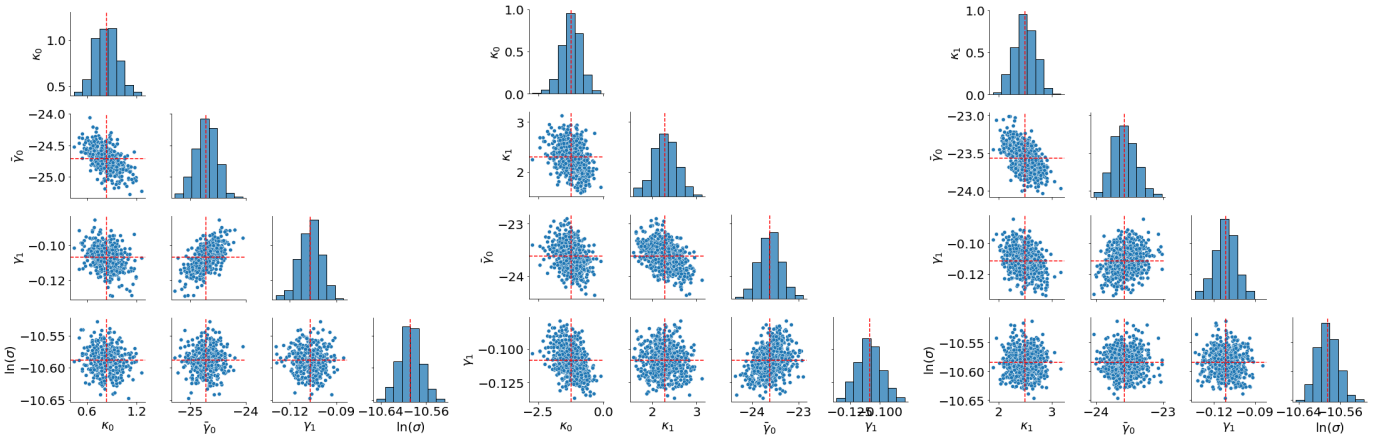

Fig. 7.   Pair plots of samples taken from the posterior in the Turin model (left), the SV model (center), and the Strauss–SV model (right).

parameter dependencies when posterior samples are extracted from the MCMC sampler.

B. Model Assessment

To compare the fits of the three models, we use posterior predictive model checks [38], [39]. The checking method proceeds by first drawing a parameter sample from the posterior distribution, which is then used in the model to simulate synthetic data. Repeating this process a large number of times provides a set of data which can be compared with the measurement data in terms of summary statistics. Here, we compare in terms of the posterior 95% envelopes of the simulated summaries, given as the interval between a lower quantile (2.5%) and an upper (97.5%) quantile. If the data are well-represented by the model, the data should lie within the posterior envelope. Further details of the method of posterior predictive model checking can be found in [38] and [39].

Fig. 8 shows the squared magnitude of the transfer function of the data together with 95% posterior predictive envelopes under the Turin model. The envelopes were computed pointwise in frequency for 200 posterior predictive simulations of channel data. Each simulation was obtained by first sampling the model parameters and MPCs from their posterior distribution under the Turin model, and next, given the resulting channel transfer function, sampling the channel data from the model (1). The observed data fall nicely within the envelopes and it can be concluded that the model is able to explain

the data, i.e., the plot does not lead to questioning the fit of the Turin model. Very similar envelopes (not shown) were obtained for the SV model. For comparison, we show in Fig. 8 also the means of the posterior predictive distributions which are very similar for the Turin model and the SV model.

The similarity between the Turin and SV models regarding the posterior predictive distributions for the channel data magnitudes suggests that the posterior distributions of MPCs are more determined by the observed data than the underlying models. However, this does not mean that any configuration of MPCs is equally likely under the two models. The fact that the posterior distribution of κ_1 under the SV model is clearly bounded away from small values is indicative that the posterior distribution of delays is concentrated on clustered configurations which have a higher likelihood for large κ_1 for the SV model than for the Turin model ($\kappa_1 = -\infty$).

To further show that there is a substantial difference in terms of model fit between Turin and SV, we consider the power autocorrelation, i.e., the autocorrelation function of the squared magnitudes of the transfer function. The power autocorrelation function for the SV model was previously studied in [40]. As discussed in relation to Fig. 8, the posterior distribution of the MPCs is to a high degree controlled by the observed data. Hence, if both the model parameters and MPCs are drawn from their posterior distribution, it becomes hard to assess the implications of the different distributional (clustered or not) assumptions for the MPCs. We therefore use
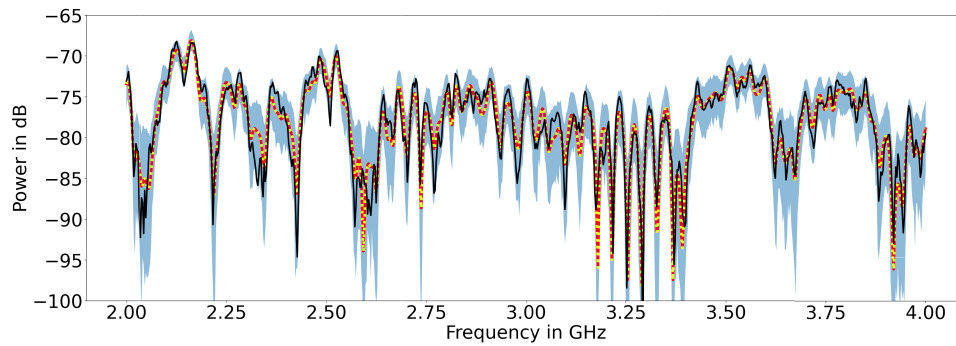
Fig. 8. Squared magnitude of the transfer function of the channel data (black) together with means in the Turin model (dotted red) and SV model (dotted yellow), and 95% envelopes for the Turin model (shaded blue).
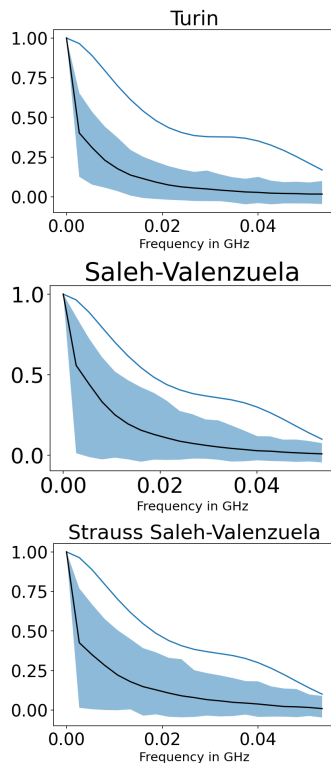

Fig. 9. Autocorrelation function for the squared magnitude of the transfer function of the channel data (blue) together with the average autocorrelation (black) for the Turin model (top), the SV model (middle), and the Strauss–SV model (bottom). The shaded blue areas are 95% envelopes.
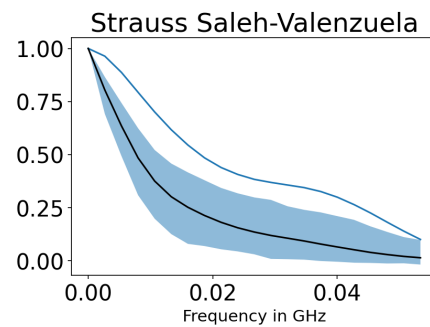

Fig. 10. Autocorrelation function for the squared magnitude of the transfer function of the channel data (blue) together with the average autocorrelation (black) for a Strauss–SV model with parameters $\beta = 20$ and $\psi = 0.75$. The shaded blue areas are 95% envelopes.
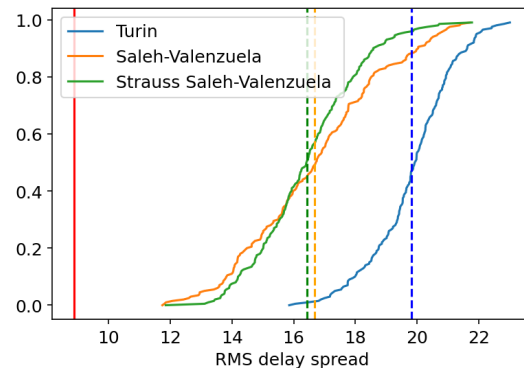

Fig. 11. Posterior predictive ecdfs for the rms-delay spread in the different models; the dashed vertical lines indicate the posterior predictive means of the rms-delay spread. The solid black horizontal lines indicate the 2.5% and 97.5% levels so that the 95% posterior predictive intervals are obtained by forming the intersections with the corresponding cdfs. The solid red vertical line shows the rms-delay spread of the real channel data.

the following variant of posterior predictive model checking. As before, we first draw the model parameters from their posterior distributions. Next, given the model parameters, both MPCs and the final channel data are generated from (7). For example, in case of the Turin model, the delays are generated from a homogeneous Poisson process. In this way, for each model, we generate 200 posterior predictive simulations of channel data and compute the autocorrelation function for each simulation.

Fig. 9 illustrates the posterior predictive means and 95% envelopes for the power autocorrelation function. In particular, we observe that both the SV model and the Strauss–SV model lead to autocorrelations that are substantially higher than those in the Turin model. However, neither the SV model nor the

Strauss–SV model fits the data perfectly. Hence, using our new methodology we are able to disclose shortcomings of the widely used existing models. Such shortcomings seem to have been overlooked in the existing literature using the multistep approach. It is quite possible that the chosen parameters for the Strauss parameters are not optimal. Based on the Strauss parameters $(\psi, \beta, R) = (0.75, 20, 0.15)$ obtained with trial and error, Fig. 10 illustrates that there is scope for improved fit of the Strauss–SV model by changing the values of the
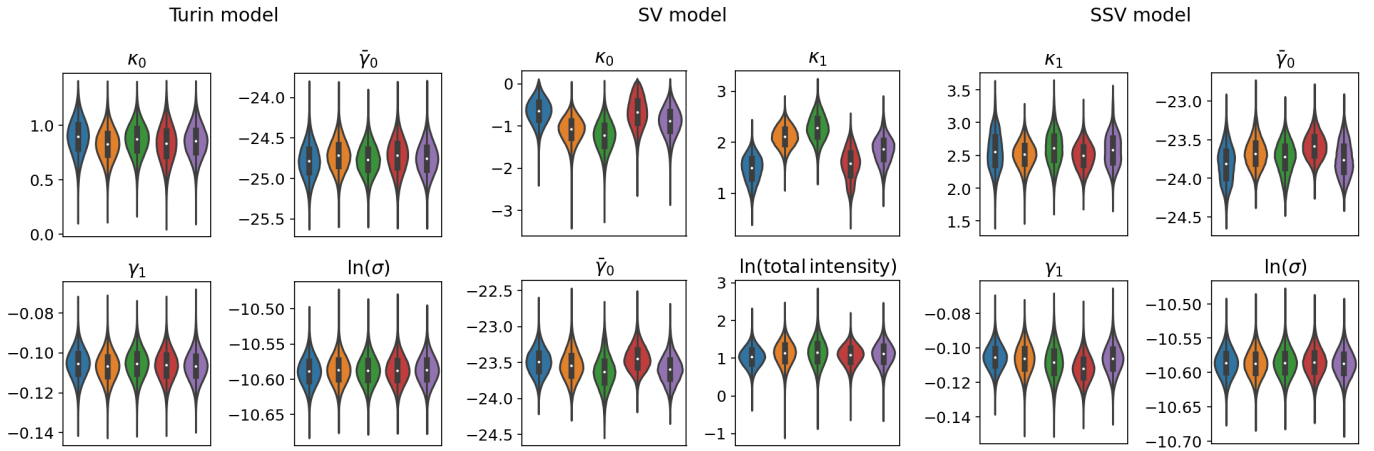
Fig. 12.   Posterior distribution of parameters for five runs of the inference for the Turin model (left), SV model (middle), and the Strauss–SV model (right). In case of the SV model, the total intensity is $\exp(\kappa_0)(1 + \exp(\kappa_1))$.

Strauss parameters, and we think it could be a topic for future work.

Finally, we analyze the rms-delay spread defined as

$$\tau_{\text{rms}} = \sqrt{\frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2} \tag{12}$$

where m_j is the jth temporal moment of a channel impulse response measurement, $y(t)$, and is computed as

$$m_j = \int_0^{\tau_{\text{max}}} t^j |y(t)|^2 \text{dt}, \quad j = 0, 1, 2, \dots \tag{13}$$

The temporal moments and thus the rms-delay spread are calculated per measurement realization and are considered as random variables. As shown in [41], the variance is related to the power autocorrelation. From the posterior predictive empirical cdfs (ecdfs) of the rms-delay spread in Fig. 11, it appears that the rms-delay spread for the SV model and the Strauss–SV model is much closer to the value found in the data than that of the Turin model. This confirms the conclusion from the power autocorrelation analysis. It should be noted, however, that since the Turin model is more parsimonious in terms of parameters, we expect it to be less flexible than the two other models. Nevertheless, all the three models overestimate the rms-delay spread found in the data. This suggests that there is still scope for achieving a better fit of the data by resorting to a more refined model.

C. Diagnostics for the MCMC Sampler and Posterior Multimodality

We now analyze the performance of the MCMC sampler in the context of the data example. In addition to considering estimated autocorrelations and trace plots for the MCMC samples, we also assessed the chosen burn-in using Geweke's diagnostic. This did not reveal problems in case of the Turin and Strauss–SV models, but a critical value of Geweke's diagnostic was obtained for κ_0 in case of the SV model; see the Supplementary for results.

Fig. 12 shows comparison of the posterior results obtained for replicated MCMC runs. In case of the SV model, the posterior results for κ_0 and κ_1 differ over MCMC runs while the posterior results for the total MPC intensity $\exp(\kappa_0)(1 + \exp(\kappa_1))$ are very similar. This shows a problem of identifiability where the pattern of MPCs is explained well by several modes corresponding to either few large clusters (small κ_0, large κ_1) or many small clusters (large κ_0 and small κ_1). The posterior results reported in this section cover the few large cluster mode.

The problem of identifiability is not uncommon when fitting a complex model to a dataset with insufficient information for resolving the nature of the underlying latent components. In our simulation study, we did not experience this problem. However, in contrast to the simulation study, our posterior predictive model checking indicates that the SV model is misspecified for the data considered. This may further amplify the problem of identifiability.

Discovering both the modes within one MCMC run with the current MCMC sampler requires excessively long MCMC runs and designing an MCMC sampler that could move quickly between the modes is a very difficult task. From a practical point of view, the best remedy of the identifiability problem might be to include stronger prior information on either κ_0 or κ_1. Moreover, more research is needed to obtain improved models for channel data as considered in this section.

VII. Conclusion

We proposed a Bayesian inference method for the Turin and SV models implemented using the MCMC methods. The proposed method is a statistically sound inference technique as opposed to the currently used multistep approaches involving multipath extraction and clustering algorithms. Furthermore, the method does not rely on side information of hard-to-estimate quantities such as the number of MPCs or the number of clusters. Being a Bayesian approach, not only point estimates but also the whole posterior distribution, and thus information on estimation accuracy, are provided as the output.

In addition to the Turin and SV models, we considered a new model, coined the Strauss–SV model, where it is possible to model regularity of cluster parent configurations. This can

lead to more parsimonious configurations of cluster parents in cases where clusters of MPCs are believed to be well-separated in time.

Our Bayesian methodology is able to fit the considered models to both the simulated and measured data. To validate the fit models, we performed posterior predictive model checking. This check revealed that all the three considered models are able to fit the transfer function data, but have shortcomings in fitting the power autocorrelation in the frequency domain and the rms-delay spread.

For the Strauss–SV model, we fixed the Strauss parameters in an ad hoc manner. Preliminary experiments indicated that a better fit might be obtained if the additional flexibility of the Strauss–SV model is used by including the Strauss parameters in the Bayesian inference. A thorough investigation of the new Strauss SV model could be done in future work by adapting advanced inference methods for doubly intractable posterior distributions available in the statistics literature.

Our main methodological focus regarding the MCMC sampler was the updating of the MPCs where we in particular introduced a novel split–merge step. However, there is certainly also scope for replacing our rather naive random walk updates with more efficient updates of the model parameters using, e.g., normal approximations of the posterior [42], [43] or adaptive MCMC [44].
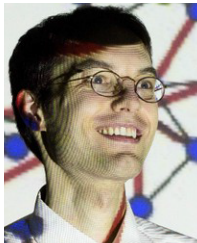
ACKNOWLEDGMENT

REFERENCES

[1] G. L. Turin, F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry, "A statistical model of urban multipath propagation," *IEEE Trans. Veh. Technol.*, vol. VT-21, no. 1, pp. 1–9, Feb. 1972.

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

[3] M. L. Jakobsen, B. H. Fleury, and T. Pedersen, "Analysis of the stochastic channel model by Saleh & Valenzuela via the theory of point processes," in *Proc. Int. Zurich Seminar on Communications (IZS)*, in Zürich, Eidgenössische Technische Hochschule Zürich, Mar. 2012, pp. 1–5.

[4] M. L. Jakobsen, T. Pedersen, and B. H. Fleury, "Analysis of stochastic radio channels with temporal birth-death dynamics: A marked spatial point process perspective," *IEEE Trans. Antennas Propag.*, vol. 62, no. 7, pp. 3761–3775, Jul. 2014.

[5] 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, Jun. 2015.

[6] 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, Mar. 2014.

[7] 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 *Proc. Eur. Conf. Antennas Propag.*, 2011, pp. 3606–3610.

[8] J. Li, B. Ai, R. He, M. Yang, Z. Zhong, and Y. Hao, "A cluster-based channel model for massive MIMO communications in indoor hotspot scenarios," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 3856–3870, Aug. 2019.

[9] M. Yang et al., "Measurements and cluster-based modeling of Vehicle-to-Vehicle channels with large vehicle obstructions," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5860–5874, Sep. 2020.

[10] X. Yin and X. Cheng, *Propagation Channel Characterization, Parameter Estimation, and Modeling for Wireless Communications*. Singapore: Wiley, 2016.

[11] 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 Veh. Technol. Conf.*, Sep. 2006, pp. 1–5.

[12] 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, Jun. 2013.

[13] R. He et al., "On the clustering of radio channel impulse responses using sparsity-based methods," *IEEE Trans. Antennas Propag.*, vol. 64, no. 6, pp. 2465–2474, Jun. 2016.

[14] A. Bharti and T. Pedersen, "Calibration of stochastic channel models using approximate Bayesian computation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.

[15] A. Bharti, F.-X. Briol, and T. Pedersen, "A general method for calibrating stochastic radio channel models with kernels," *IEEE Trans. Antennas Propag.*, vol. 70, no. 6, pp. 3986–4001, Jun. 2022.

[16] W.-D. Wu, C.-H. Wang, C.-C. Chao, and K. Witrisal, "On parameter estimation for ultra-wideband channels with clustering phenomenon," in *Proc. IEEE 68th Veh. Technol. Conf.*, Sep. 2008, pp. 1–5.

[17] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter estimation for stochastic channel models using temporal moments," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting*, Jul. 2019, pp. 1267–1268.

[18] A. Bharti, R. Adeogun, and T. Pedersen, "Estimator for stochastic channel model without multipath extraction using temporal moments," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[19] R. Adeogun, "Calibration of stochastic radio propagation models using machine learning," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 12, pp. 2538–2542, Sep. 2019.

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

[21] A. Bharti, R. Adeogun, and T. Pedersen, "Auto-generated summaries for stochastic radio channel models," in *Proc. 15th Eur. Conf. Antennas Propag. (EuCAP)*, Mar. 2021, pp. 1–5.

[22] S. A. Sisson, *Handbook of Approximate Bayesian Computation*. Boca Raton, FL, USA: Chapman & Hall/CRC, Sep. 2018.

[23] C. Hirsch, A. Bharti, T. Pedersen, and R. Waagepetersen, "Maximum likelihood calibration of stochastic multipath radio channel models," *IEEE Trans. Antennas Propag.*, vol. 69, no. 7, pp. 4058–4069, Jul. 2021.

[24] J. Møller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, FL, USA: CRC Press, 2004.

[25] A. Meijerink and A. F. Molisch, "On the physical interpretation of the Saleh–Valenzuela model and the definition of its power delay profiles," *IEEE Trans. Antennas Propag.*, vol. 62, no. 9, pp. 4780–4793, Sep. 2014.

[26] T. Pedersen, "Stochastic multipath model for the in-room radio channel based on room electromagnetics," *IEEE Trans. Antennas Propag.*, vol. 67, no. 4, pp. 2591–2603, Apr. 2019.

[27] J. Møller, A. N. Pettitt, R. Reeves, and K. K. Berthelsen, "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants," *Biometrika*, vol. 93, no. 2, pp. 451–458, Jun. 2006.

[28] I. Murray, Z. Ghahramani, and D. J. C. MacKay, "MCMC for doubly-intractable distributions," in *Proc. 22nd Annu. Conf. Uncertainty Artif. Intell. (UAI)*. Arlington, VA, USA: AUAI Press, 2006, pp. 359–366.

[29] Y. F. Atchadé, N. Lartillot, and C. Robert, "Bayesian computation for statistical models with intractable normalizing constants," *Brazilian J. Probab. Statist.*, vol. 27, no. 4, pp. 416–436, Nov. 2008.

[30] F. Liang and I.-H. Jin, "A Monte Carlo metropolis-hastings algorithm for sampling from distributions with intractable normalizing constants," *Neural Comput.*, vol. 25, no. 8, pp. 2199–2234, Aug. 2013.

[31] J. Park and M. Haran, "Bayesian inference in the presence of intractable normalizing functions," *J. Amer. Stat. Assoc.*, vol. 113, no. 523, pp. 1372–1390, Jul. 2018.

[32] M. Beraha, R. Argiento, J. Møller, and A. Guglielmi, "MCMC computations for Bayesian mixture models using repulsive point processes," 2020, *arXiv:2011.06444*.

[33] S. Jain and R. M. Neal, "A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model," *J. Comput. Graph. Statist.*, vol. 13, no. 1, pp. 158–182, Mar. 2004.

[34] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *J. Roy. Statist. Soc. B*, vol. 59, no. 4, pp. 731–792, 1997.

[35] P. Guttorp and T. L. Thorarinsdottir, "Bayesian inference for non-Markovian point processes," in *Advances and Challenges in Space-time Modelling of Natural Events*, E. Porcu, J.-M. Montero, and M. Schlather, Eds. Berlin, Germany: Springer, 2012, pp. 79–102.

[36] W. Fan, I. Carton, J. O. Nielsen, K. Olesen, and G. F. Pedersen, "Measured wideband characteristics of indoor channels at centimetric and millimetric bands," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, p. 58, 2016.

[37] J. O. Berger and J. M. Bernardo, "On the development of reference priors (with discussion)," in *Bayesian Statistics*, vol. 4, London, U.K.: Oxford Univ. Press, 1992, pp. 35–60.

[38] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2013.

[39] A. Gelman, X.-L. Meng, and H. Stern, "Posterior predictive model assessment via realized discrepancies," *Statistica Sinica*, vol. 6, pp. 733–807, May 1996.

[40] M. S. Derpich and R. Feick, "Second-order spectral statistics for the power gain of wideband wireless channels," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1013–1031, Mar. 2014.

[41] T. Pedersen, "First- and second order characterization of temporal moments of stochastic multipath channels," in *Proc. 33rd Gen. Assem. Scientific Symp. Int. Union Radio Sci.*, Aug. 2020, pp. 1–4.

[42] S. Chib, "Analysis of multivariate probit models," *Biometrika*, vol. 85, no. 2, pp. 347–361, Jun. 1998.

[43] R. Waagepetersen, N. Ibanez-Eschriche, and D. A. Sorensen, "Strategies for MCMC computation in quantitative genetics," *Genet., Selection, Evol.*, vol. 40, no. 2, pp. 161–176, Feb. 2008.

[44] C. Andrieu and J. Thoms, "A tutorial on adaptive MCMC," *Statist. Comput.*, vol. 18, no. 4, pp. 343–373, Dec. 2008.

Ayush Bharti received the B.E. degree in electrical and electronics engineering from the Birla Institute of Technology and Sciences, Pilani, India, in 2015, and the M.Sc. degree in signal processing and computing and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2017 and 2021, respectively.

He is currently working as a Post-Doctoral Researcher with the Department of Computer Science, Aalto University, Espoo, Finland, and also affiliated with the Finnish Centre for Artificial Intelligence. His research interests include simulation-based inference, approximate Bayesian computation, and kernel methods.

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 d'Électronique et de Télécommunications 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.

Christian Hirsch received the Diploma degree in mathematics from the Ludwig Maximilian University of Munich (LMU Munich), Munich, Germany, in 2010, and the Ph.D. degree in mathematics from Ulm University, Ulm, Germany, in 2014.

He was a Junior Professor with the University of Mannheim, Mannheim, Germany, and an Assistant Professor with the University of Groningen, Groningen, The Netherlands. He is currently an Associate Professor of data science and statistics with Aarhus University, Aarhus, Denmark. Before that, he held post-doctoral positions at the Weierstrass Institute for Applied Analysis and Stochastic (WIAS), Berlin, Germany, LMU Munich, and Aalborg University, Aalborg, Denmark. His research interests include spatial statistics, topological data analysis, statistical signal processing, and applied probability.

Rasmus Waagepetersen received the M.Sc. and Ph.D. degrees in mathematical statistics from Aarhus University, Aarhus, Denmark, in 1995 and 1997, respectively.

He was the Scientist with the Danish Institute of Agricultural Sciences, Tjele, Denmark, from 1997 to 2000. From 2000 to 2007, he was an Assistant Professor and subsequently an Associate Professor with the Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark. From 2008 to 2009, he was a Credit Risk Analyst with Spar Nord Bank, Aalborg. Since 2010, he has been a Full Professor with the Department of Mathematical Sciences, Aalborg University.