Calibrating Stochastic Radio Channel Models

An Approximate Bayesian Computation Approach

Bharti, Ayush

DOI (link to publication from Publisher):
10.54337/aau429765530

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

CALIBRATING STOCHASTIC RADIO CHANNEL MODELS: AN APPROXIMATE BAYESIAN COMPUTATION APPROACH

BY
AYUSH BHARTI

DISSERTATION SUBMITTED 2021

AALBORG UNIVERSITY
DENMARK

Calibrating Stochastic Radio Channel Models: An Approximate Bayesian Computation Approach

Ph.D. Dissertation

Ayush Bharti

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7A
DK-9220 Aalborg

Dissertation submitted: April 2021

PhD supervisor: Associate Professor Troels Pedersen
Aalborg University

PhD committee: Associate Professor Beatriz Soret (chairman)
Aalborg University

Professor Xuefeng Yin
Tongji University

Associate Professor Klaus Witrisal
Graz University of Technology (TU Graz)

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-932-9

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2021

To my parents,
Arvind and Rupam.

Abstract

Stochastic models of the radio channel are widely used simulation tools due to their simplicity and low computational cost compared to the deterministic approaches. Another appealing feature of stochastic channel models is that they are versatile, i.e., they can be applied to different radio propagation scenarios by simply calibrating their parameters. This is easier said than done as most stochastic channel models have intractable likelihood functions, leaving standard parameter estimation methods inapplicable. The models are typically constructed with a focus on the underlying phenomenon and less attention to any calibration concerns. This leads to models which, despite representing the channel well, can be difficult to calibrate using standard estimation approaches. As a way to remedy this, engineers have mainly relied on ad-hoc methods.

Since the early days of the Turin model in the 70s, it has been customary in the field to split the calibration problem into multiple steps. The first step involves estimating the multipath components by implementing high-resolution algorithms like CLEAN, SAGE, and RiMAX. The extracted multipaths are then used to estimate the parameters of the stochastic channel models. With the advent of the cluster-based Saleh-Valenzuela (S-V) model, an additional step of clustering the multipath components has been added to the calibration procedure. This multi-step approach has fueled the independent development of new multipath extraction and clustering algorithms, each having their own particular settings. However, evaluating performance of the end-to-end calibration method becomes infeasible, not to mention the difficulty in implementing these sophisticated algorithms.

In search of a calibration method applicable to stochastic channel models irrespective of their underlying mathematical structure, we look towards other scientific fields where calibrating models with intractable likelihoods is commonplace. One such field is population genetics, where a likelihood-free calibration method called *approximate Bayesian com-*

putation (ABC) has been developed in the past couple of decades. Suited for models that are easy to simulate from, ABC yields samples from the approximate posterior of the parameters by comparing summary statistics of the measured and simulated data in some distance metric.

In this work, we develop likelihood-free methods, primarily based on ABC, to calibrate stochastic radio channel models without multipath extraction and clustering. Our methods utilize the *temporal moments*, which are frequently used statistics for characterizing the radio channel. We show that the temporal moments are informative about the parameters of stochastic channel models, and hence, can be used to calibrate them. Taking the Turin, the S-V, and the *propagation graph* (PG) model as examples, we propose calibration methods that accurately estimate their parameters in simulations. We also validate the proposed methods by applying them to real channel measurements.

Considering the usefulness of the temporal moments, we empirically investigate their distribution using a wide range of measurements. We find them to be well-modeled by a multivariate log-normal distribution in different propagation scenarios and frequencies. The proposed model is simple to use, and can simulate channel characteristics such as *mean delay* and *rms delay spread* jointly.

Our main contribution is the development of a general ABC method that can calibrate stochastic channel models with different mathematical structures using the exact same procedure and settings. The end-to-end calibration method is unlike any available in the literature, and it enables model comparison studies that was not possible before. Exemplified by the S-V and the PG model, the method circumvents the need for specialization of summaries, and instead, relies on comparing distributions of the temporal moments.

Resumé

Stokastiske modeller af radiokanalen er meget anvendte simuleringsværktøjer på grund af deres enkelhed og lave beregningsomkostninger sammenlignet med de såkaldte deterministiske metoder. En styrke ved stokastiske modeller er, at de er alsidige, dvs. de kan anvendes til forskellige radiomiljøer ved blot at kalibrere deres parametre. Dette er dog lettere sagt end gjort, da de fleste stokastiske modeller for radio miljøer med multiple udebredelsesveje har uhåndterlige eller utilgængelige likelihoodfunktioner, hvorfor estimeringsmetoder som maximum likelihood eller de fleste Bayesianske metoder ikke kan anvendes. Modellerne er typisk konstrueret med fokus på det underliggende radioudbredelsesfænomen (fx flervejsudbredelsen) og ringe opmærksomhed på muligheden for kalibrering. Dette fører til modeller, der, selv om de repræsenterer radiokanalen godt, kan være vanskelige at kalibrere ved hjælp af standard estimatorer. I stedet har ingeniører hovedsagelig benyttet sig af metoder konstrueret ad-hoc til den enkelte model.

Siden Turins arbejde med stokastiske flervejsmodeller i 1970'erne har det været almindelig praksis at dele kalibreringsproblemet op i flere mindre trin. Det første trin involverer ekstrahering af flervejskomponenter (multipaths) ved hjælp af algoritmer som CLEAN, SAGE og RiMAX. De ekstraherede mulitpaths bruges derefter til at estimere parametrene for de stokastiske kanalmodeller. Med tilføjelsen af den clusterbaserede Saleh-Valenzuela (S-V) -model er der tilføjet et yderligere trin til kalibreringsprocessing, nemlig gruppering (clustering) af multipath komponenter. Opdeling af estimeringsproblemet i flere trin har ansporet en del arbejder med at udvikle nye algoritmer til multipath ekstraktion og clustering. Hver især har disse algoritmer har særlige parametervalg og antallet af kombinationsmuligheder er mange. Dette gør det til en vanskelig opgave at evaluere og lave end-to-end sammenligninger mellem for skellige kalibreringsmetoder.

I jagten efter en kalibreringsmetode, der kan anvendes på stokastiske

kanalmodeller uanset deres underliggende matematiske struktur, søges fra andre fagområder, hvor kalibrering af modeller med utilgængelige likelihood funktioner er almindelig. Et sådant område er populationsgenetik, hvor en likelihood-fri kalibreringsmetode kaldet Approximate Bayesian Computation (ABC) er blevet udviklet i løbet af de sidste par årtier. ABC er velegnet til modeller, der let kan simuleres fra, og tillader sampling fra den approksimerede a posteriori fordeling ved at sammenligne statistikker for målte og simulerede data i en vis afstandsmetrik.

I dette arbejde udvikles likelihood-fri metoder, primært baseret på ABC, til at kalibrere stokastiske multipath radiokanalmodeller. Derved undgås trinene med multipath ekstraktion og clustering. Vores metoder anvender de letberegnelige tids-momenter, som ofte bruges som statistikker til karakterisering af radiokanaler. Vi viser, at tidsmomenterne er informative om parametrene for stokastiske kanalmodeller og derfor kan bruges til kalibrering. De udviklede ABC metoder er generelle i den forstand at nøjagtig den samme algoritme kan anvendes til kalibrering forskellige modeller, fx Turin's model, S-V modellen og modellering med udbredelsesgrafer (PG-modellen). De udviklede metoder afprøves i simuleringer og på måledata.

Da tidsmomenterne findes nyttige som statistikker undersøges deres fordeling empirisk for en række forskellige datasæt fra forskellige radiomiljøer. Vi finder at fordelingen af tidsmomenter er godt modeleret ved en multivariat log-normalfordeling. Log-normal modellen er enkel at bruge i simuleringer.

Afhandlingens hovedbidrag er en generel ABC-metode, til end-to-end kalibrering af stokastiske kanalmodeller med forskellige matematiske strukturer ved hjælp af nøjagtig samme procedure og indstillinger. Dette gør det praktisk muligt at sammenligne at sammenligne forskellige modeller på de samme data.

Contents

Thesis Details

Thesis Title:	Calibrating Stochastic Radio Channel Models — An Approximate Bayesian Computation Approach
Ph.D. Student:	Ayush Bharti
Supervisor:	Assoc. Prof. Troels Pedersen, Aalborg University

The main body of this PhD thesis consists of the following nine (5 conference and 4 journal) papers:

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

Paper B A. Bharti, R. Adeogun, and T. Pedersen, "Estimator for stochastic channel model without multipath extraction using temporal moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, pp. 1-5, 2019.

Paper C C. Hirsch, A. Bharti, T. Pedersen, and R. Waagepetersen, "Maximum likelihood calibration of stochastic multipath radio channel models," *IEEE Trans. on Antennas and Propag.*, 2020.

Paper D A. Bharti and T. Pedersen, "Calibration of stochastic channel models using approximate Bayesian computation," in *Proc. IEEE Global Commun. Conf. Workshops (Globecom)*, pp. 1-6, 2019.

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

Paper F A. Bharti, R. Adeogun, and T. Pedersen, "Auto-generated summaries for stochastic radio channel models," in *15th Eur. Conf. on Antennas and Propag.*, pp. 1–5, 2021.

Paper G A. Bharti, L. Clavier, and T. Pedersen, "Joint statistical modeling of received power, mean delay, and delay spread for indoor wideband radio channels," in *14th Eur. Conf. on Antennas and Propag.*, pp. 1–5, 2020.

Paper H A. Bharti, R. Adeogun, X. Cai, W. Fan, F.-X. Briol, L. Clavier, and T. Pedersen, "Joint modeling of received power, mean delay, and delay spread for wideband radio channels," *IEEE Trans. on Antennas and Propag.*, 2021.

Paper I A. Bharti, F.-X. Briol, and T. Pedersen, "A general method for calibrating stochastic radio channel models with kernels," (under review) IEEE Trans. on Antennas and Propag., 2021.

Additionally, the following two papers were co-authored by Ayush Bharti during his PhD studies:

[1] R. Adeogun, T. Pedersen, and A. Bharti, "Transfer function computation for complex indoor channels using propagation graphs," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, pp. 566–567, Sep. 2018.

[2] R. O. Adeogun, A. Bharti, and T. Pedersen, "An iterative transfer matrix computation method for propagation graphs in multiroom environments," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 4, pp. 616–620, Apr. 2019.

This thesis has been submitted to the Doctoral School of IT and Design at Aalborg University, Denmark, for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis.

Preface

This thesis is based on the work that I did at the Wireless Communication Networks section at Aalborg University from November 2017 to March 2021. Chapters 1 – 4 provide an introduction and brief descriptions of the contributions of the thesis. The main body consists of nine research papers, referred to as Paper A – Paper I, that are published in peer-reviewed conferences and journals. This work has been funded by the Danish Council for Independent Research, Grant DFF 7017-00265.

As I sit down and reminisce about this three-and-a-half-year long journey that I undertook, I am reminded of a number of people that have helped and supported me along the way, none more so than my supervisor Assoc. Prof. Troels Pedersen. Thank you Troels for not just being a project supervisor, but more importantly a guide, a mentor, and a friend. You have guided me not only on what to think, but more importantly on how to think. The skills and practices I have acquired working with you make me confident of excelling in my future research endeavors.

I also wish to express my sincerest gratitude to Dr. Ramoni Adeogun and Dr. François-Xavier Briol for the numerous technical discussions that helped me grow as a researcher. I would like to thank Assoc. Prof. Carles Navarro Manchon for always finding time for my frequent unannounced visits, and to Sajad Rezaie for being the best office-mate. Thank you Dorthe Sparre and Linda Villadsen for taking care of all the administrative affairs.

I thank my parents for their endless support and dedicate this thesis to them. If not for the privilege of being their son, I wouldn't have dreamed of relocating to another continent and pursuing a career that they are still unsure whether qualifies as job or studies. Finally, I thank my constant source of motivation, Modesta, for her love and support.

Ayush Bharti
Aalborg University, April 16, 2021

Part I

Introduction

Introduction

Wireless or radio communications is ubiquitous in the modern world and their applications are increasing day-by-day. Wireless systems enable communication between devices over a wide range of distances, from a few metres between the WiFi router and our phones to millions of kilometres between Earth and the Perseverance rover on Mars. With the advent of self-driving cars, internet-of-things, and automation of industries, the reliance on radio communications is going to increase even more in the near future.

Radio communication pertains to the use of electromagnetic waves for sending and receiving information between two devices. A communication system primarily consists of three components — a transmitter, a receiver, and the environment in which the communication is taking place, termed the *radio channel*. The transmitter, equipped with one or more antennas, transmits an electromagnetic wave which has some information embedded in it. The wave, or the signal, propagates through the environment, interacting with surrounding objects and surfaces, and thus gets distorted and modified according to the radio channel. This modified signal is then sensed at one or more antennas of the receiver. The task of the receiver is now to extract the initial embedded information from the received signal, which is now contaminated with noise and possibly interference. To do that, the communication system needs to account for the effect of the radio channel on the transmitted signal. The radio systems of the modern era rely on understanding the behavior of the radio channel for better and more efficient recovery of the information. Hence, knowledge about the channel is imperative for the design, analysis, and testing of communication systems.

1 Radio Channel Modeling

Radio channel modeling is the craft of creating mathematical descriptions of the radio channel [1]. Typically, the radio channel is modeled as a linear system with the transmitted signal, $x(t)$, being the input and the received signal, $y(t)$, being its output. In many cases the channel is time-variant; however, we will assume throughout this thesis that the channel is time-invariant. A time-invariant system is agnostic to time-shifts in the input signal, and simply shifts the output signal correspondingly. This would correspond to a radio channel which does not change over time, i.e. when the transmitter, the receiver, and the environment is static, e.g. in an indoor scenario. The output of a linear, time-invariant system can be written in complex-baseband notation [2] using the convolution operator as (ignoring noise)

$$y(t) = \{h * x\}(t) \tag{1}$$

$$= \int h(\tau)x(t-\tau)\mathrm{d}\tau \tag{2}$$

where h is referred to as the *impulse response* of the radio channel [3]. The impulse response, as the name suggests, is the response of the system when the Dirac delta, or an impulse, is applied as input. The time-domain functions can be transformed into the frequency-domain using the Fourier transform, $\mathcal{F}$[1], and (1) can be alternatively expressed as

$$Y(f) = \mathcal{F}\{h(t)\}\mathcal{F}\{x(t)\} \tag{3}$$

$$= H(f)X(f). \tag{4}$$

Here, $Y(f) = \mathcal{F}\{y(t)\}$, and the quantity $H(f)$ is called the *transfer function*. The radio channel is fully characterized by its impulse response or transfer function, and therefore $h(t)$ or $H(f)$ are the quantities of interest for radio channel modeling experts.

Radio channel modeling can be roughly classified into *deterministic* modeling, *stochastic* modeling, a combination of the two termed as *geometric-stochastic* modeling [4], and *statistical* modeling, each having their own purpose. Deterministic channel models aim to mimic the exact propagation conditions and phenomenon for a given propagation environment. They therefore rely on the geometry, wave theory, geometric

[1] The Fourier transform of a function $g(t)$ is defined as $\mathcal{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)\exp(-j2\pi ft)dt$.

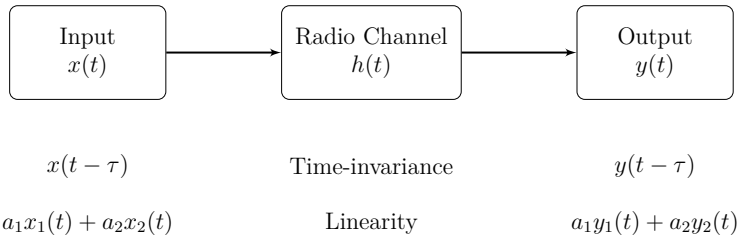


Fig. 1: A communication system with one transmitter and one receiver. The transmitted signal, $x(t)$, is transformed by a linear, time-invariant radio channel, $h(t)$, to produce the received signal, $y(t)$. Here $\tau \in \mathbb{R}_+$ is a time shift in the signal and $a_1, a_2 \in \mathbb{R}$ are constants.

optics and the uniform theory of diffraction [5] to simulate the interactions of the electromagnetic wave with the objects in the surroundings. They require detailed descriptions of the surrounding objects and their material properties, along with their precise locations. Deterministic models such as the ray tracer are therefore computationally expensive, but are considered to be quite accurate in settings where the environment and electromagnetic material properties are known with great accuracy.

Stochastic channel models, on the other hand, are agnostic to the minute details of the propagation environment and positions of the objects. They assume that the received signal is a realization from an unknown distribution which is the true state of nature. They then aim to construct a probabilistic model of the channel that approximates the description of the received signal based on the propagation phenomenon. The stochastic channel models are governed by a set of parameters, primarily the parameters of the proposed probability distributions. Since they are designed with no specific propagation environment in mind, they can be applied to different scenarios by adjusting their parameters. Their computational cost is low, and hence, they can be used for Monte Carlo methods where a large number of channel realizations are needed. They are also simpler to use for analysis and design of radio systems.

The geometric-stochastic models basically combine some aspects of both the deterministic and the stochastic models. For example, they may take into account the geometry of the scenario and the position of the objects in the environment, while the interactions of the signal with the environment are considered random.

Finally, the statistical models describe the output or the data, disregarding any internal workings of the system. Statistical models are

often used for modeling the path loss of the signal, i.e. the reduction in the power of an electromagnetic signal as it propagates through space. In this work, we focus on the radio channel models of stochastic nature.

2 Calibrating Stochastic Models

In order to use a model for simulation purposes, its parameters need to be either set based on physical considerations, e.g. knowledge of environment and material constants, or estimated according to some data. This procedure of setting parameters of a model is termed as *calibration*. A typical example in the field of signal processing would be the modeling of thermal noise in various measurement equipment with a zero-mean Gaussian distribution. Calibration from data then involves fitting noisy measurements from the equipment to the zero-mean Gaussian distribution and estimating its variance value. Using the estimated variance value, noise can be simulated from the model. Alternatively, this variance might be calibrated from a thermal noise model on knowledge of the temperature. In this thesis, we limit our discussion to calibration of models using data, also known as parameter estimation.

The aim of calibration is to find the parameter value that yields simulated data from the model that best fits the data. Thus, calibrating a model from data is essentially an inference or a model fitting problem. The data used for calibration, which we will refer to as *measured* or *observed* data throughout this thesis, is usually collected through measurement campaigns.

In the case of radio propagation, measurement campaigns are expensive to carry out and require significant time and effort. Hence, carrying them out for every propagation scenario or setting is infeasible. However, similar impulse response data can be simulated from a stochastic channel model if its parameters are calibrated using the measured data. Of course, the accuracy of a model's predictions depends on both the model and the calibration process. A calibration method can only provide an optimal fit of the model to the data in some metric, but cannot guarantee accurate model predictions or remedy the inadequacies of a model. The calibration problem can be framed as:

Calibration problem: *Consider a stochastic model whose output is treated as random realizations out of a family of distributions, $\mathbb{P}_\theta$, parameterized by the vector θ. The observed data, y, is a collection of possibly multiple realizations out of some unknown distribution, say $\mathbb{Q}$. The calibration problem can then be framed as: estimate θ given y such that $\mathbb{P}_\theta$ is "close to" $\mathbb{Q}$.*

An important component for calibrating stochastic models is the *likelihood function*, denoted $f_\theta(y)$ or $f(y|\theta)$. The likelihood function is defined as the probability density of the data y viewed as a function of the parameter θ. It measures the support provided by the data for each possible value of the parameter. For a given data y, if the value of the likelihood function at two parameter points, θ_1 and θ_2 is L_1 and L_2, respectively, such that $L_1 > L_2$, then it is more likely that y is observed for $\theta = \theta_1$ than $\theta = \theta_2$.

There are two approaches for tackling such a calibration problem, namely the *frequentist* and the *Bayesian* approach. We now describe how the likelihood function enters in both the frequentist and Bayesian approach for calibration.

Frequentist Inference In this calibration framework, the parameter vector θ is considered to be a deterministic quantity belonging to some parameter space Θ. We denote the likelihood function as $f_\theta(y)$. Maximizing the likelihood function with respect to the parameters θ results in the maximum likelihood estimate, $\hat{\theta}_{\mathrm{ML}}$, that is

$$\hat{\theta}_{\mathrm{ML}} = \underset{\theta \in \Theta}{\operatorname{argmin}} \; f_\theta(y). \tag{5}$$

The above minimization problem can be solved either analytically, or using standard optimization toolboxes that search the parameter space to obtain a point estimate of θ.

Bayesian Inference In the Bayesian approach, the parameter θ is considered as a random vector out of some distribution $p(\theta)$, called the *prior distribution*. Hence, the likelihood function is denoted as a conditional density of the data given the parameter, $f(y|\theta)$. The aim is then to characterize the distribution of the parameters given the observed data, $p(\theta|y)$, called the *posterior distribution*. The prior, the likelihood, and the posterior are related to each

other per the Bayes' theorem that reads,

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\theta)p(\theta)d\theta}. \tag{6}$$

Point estimates of the parameter can then be obtained either by maximizing the posterior, leading to the maximum *a priori* (MAP) estimate,

$$\hat{\theta}_{\text{MAP}} = \underset{\theta}{\text{argmax}} \;\; p(\theta|y), \tag{7}$$

or by taking its mean, leading to the minimum mean squared error estimate (MMSE),

$$\hat{\theta}_{\text{MMSE}} = \mathbb{E}[p(\theta|y)]. \tag{8}$$

The denominator in Eq. 6 contains the marginal density of the observations and is termed as the *evidence*. This constant term solely depends on the data, and is usually not known due to the unavailability of the distribution of y for most practical cases. In those cases, the posterior distribution is known up to a normalization constant and it cannot be characterized fully. However, it can approximated if we get enough samples out of the posterior. Sampling from the unnormalized posterior can be achieved via Markov chain Monte Carlo [6] or sequential Monte Carlo methods [7].

An important aspect of Bayesian inference is the choice of the prior distribution. The prior signifies our initial beliefs about the parameter, which gets updated based on the observed data. Assigning equal probabilities over the entire parameter space leads to frequentist inference, where $\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{ML}}$.

Note that for all practical purposes, $f_\theta(y)$ and $f(y|\theta)$ are the same, with most engineering textbooks using the latter notation. Here, we make the distinction in notation to highlight the difference between the two inference approaches, see [8] for more details.

2.1 Intractability of Likelihood

Irrespective of the approach, it is evident that the likelihood function plays an important role in model calibration. However, for some complex stochastic models the likelihood becomes intractable. Intractablity of likelihood refers to the inability to numerically evaluate the likelihood function for an arbitrary θ, and can occur in the following cases:

I1 The model is simply too complex, in which case it is either impossible or impractical to write down a closed-form expression for the likelihood function.

I2 Variables that are important for model description are unobserved, in which case the likelihood manifests in the form of a computationally intractable integral.

I3 The likelihood function has not been derived yet for a newly constructed model.

With the likelihood unavailable, maximizing it with respect to the parameters is infeasible, as is Bayesian inference since the posterior distribution is proportional to the likelihood. Thus, calibration of such models using the aforementioned standard techniques is not possible. As it turns out, stochastic radio channel models suffer from both I1 and I2, and often also from I3. They are therefore calibrated using ad-hoc methods, as we will see in the next chapter.

3 Measuring the Radio Channel

The data used for calibrating a stochastic radio channel model is usually obtained by conducting channel measurement campaigns, called channel sounding [9]. The campaign involves placing the the transmit and the receive antennas at a certain distance from each other in some indoor or outdoor environment. A known signal is then transmitted from the transmit antennas and recorded at the receive antennas using, e.g., a vector network analyzer (VNA) which is a measurement equipment used to measure two ports. By viewing the radio channel as a two-port system, we can use a VNA to measure it, as depicted in Fig. 2. Virtual multi-input, multi-output measurements are formed by repeating measurements for multiple positions of transmit and receive antennas. The received signal at the receive antennas depends upon the signal transmitted and the radio channel, and is contaminated with thermal noise from the measuring equipment. The measurement data is then obtained by accounting for the effect of the known transmitted signal.

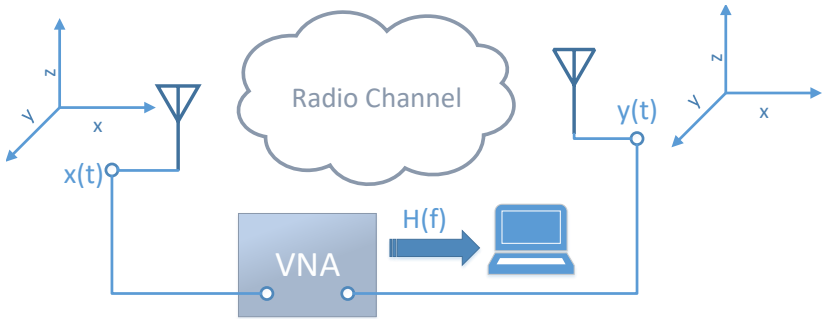


Fig. 2: Principle diagram for measurement of a radio channel using a vector network analyzer (VNA).

4 Thesis Structure

The rest of the thesis is organized as follows:

Chapter 2 introduces the problem being considered in this thesis. We present the calibration methods adopted for some of the common stochastic radio channel models and their drawbacks.

Chapter 3 presents the background on the calibration methods adopted in this thesis. We provide an overview of the likelihood-free calibration methods available in the literature.

Chapter 4 lists all the research questions to be answered and outlines the research methodology adopted.

Chapter 5 provides the main contributions of the thesis in terms of an overview and summary of Papers A–I.

Chapter 6 provides the conclusions of the thesis and the future outlook.

Calibration of Stochastic Radio Channel Models

In this chapter, we give three examples of stochastic radio channel models, namely the Turin model, the Saleh-Valenzuela model, and the propagation graph model, for the purpose of elevating the discussion on their calibration procedures. These models are used as examples throughout the thesis.

5 Stochastic Multipath Models

Consider the scenario where a radio signal $x(t)$ is transmitted in an environment as sketched in Fig. 3. The signal, being transmitted in multiple directions by the transmit antenna, bounces off of the objects in the environment before being sensed at the receiver. The phenomenon of multiple copies of $x(t)$ arriving at the receiver is termed as *multipath propagation*.

The time delay and the complex gain of each of those copies depends upon their exact propagation trajectory. For example, the line-of-sight path, which is the direct unobstructed path between the transmit and the receive antenna, has the minimum delay and the maximum gain out of all the possible paths. The l^{th} path has the delay τ_l and the gain α_l associated with it, called the *multipath components*. The received signal, $y(t)$, is hence a superposition of such multipath components, each having their own delay and gain, and can be written as

$$y(t) = (h * x)(t) + w(t) \tag{9}$$

$$= \sum_l \alpha_l x(t - \tau_l) + w(t) \tag{10}$$

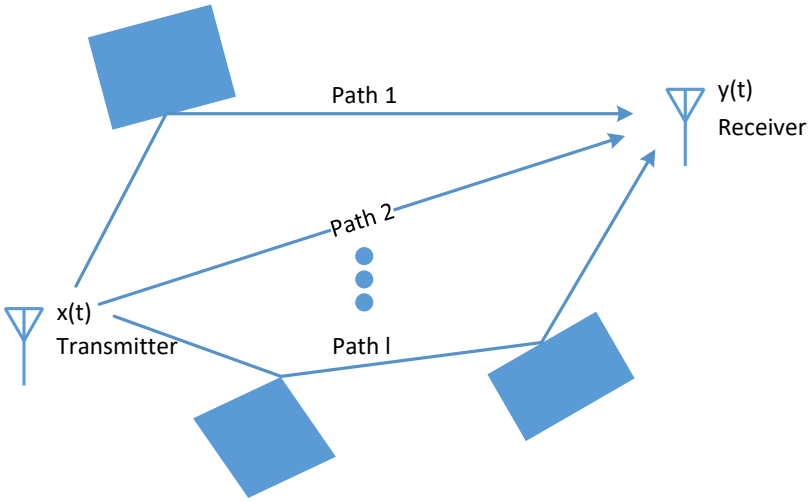


Fig. 3: Multipath propagation phenomenon. The objects reflecting the transmitted signal could be e.g. walls or ceiling in an in-room scenario or nearby building or trees in case of outdoor propagation.

where $h(t)$ is the channel impulse response and $w(t)$ denotes the complex additive white Gaussian noise.

A stochastic multipath radio channel model considers the delays and the gains as a marked point process, $Z = \{(\tau_l, \alpha_l)\}$, where $\{\tau_l\}$ are the points and $\{\alpha_l\}$ are their associated marks. Equivalently, Z may be viewed as a higher dimensional point process as done in [10]. In contrast, Z in the case of deterministic multipath models is computed from the given environment. Defining a stochastic multipath model then amounts to defining the marked point process Z, having parameters θ. The model is therefore a stochastic mechanism that outputs such a marked point process. In the following, we discuss two seminal stochastic multipath models for the radio channel, one proposed by Turin et al. in 1972 and the other by Saleh & Valenzuela in 1987.

5.1 Calibrating the Turin model

The model for multipath channel proposed by Turin *et al.* in 1972 [11] was a pioneering work in the field.

Example 5.1 (Turin model)
Turin defined the impulse response, $h(t)$, as

$$h(t) = \sum_l \alpha_l \delta(t - \tau_l), \tag{11}$$

where $\delta(\cdot)$ is the Dirac delta function. The delays, $\{\tau_l\}$, are modeled as a Poisson point process with arrival rate $\lambda(t)$. Assuming a constant arrival rate, say $\lambda(t) = \lambda_0$, yields the homogeneous Turin model. The gains conditioned on the delays, i,e, $\alpha_l|\tau_l$, are assumed as independent and identically distributed (iid), with their magnitude modeled using a log-normal distribution, and their phase uniformly distributed over $(0, 2\pi]$. The parameters of $\lambda(t)$ and that of the $\alpha_l|\tau_l$ distribution constitute the parameter vector of this model.

The Turin model suffers from intractability issue I2, with the underlying point process realization being the unobserved or hidden variables. Let y be a realization of the channel measurement, and z be the corresponding point process realization. Assuming that z is known, the joint likelihood of y and z can be written as

$$f_\theta(y, z) = p(y|z; \theta)p(z; \theta). \tag{12}$$

The desired likelihood function $f_\theta(y)$ can be obtained by marginalizing $f_\theta(y, z)$ with respect to z as

$$f_\theta(y) = \sum_k \mathbb{P}(|z| = k) \int p(y|z; \theta)p(z|\ |z| = k; \theta)dz, \tag{13}$$

where $|z|$ denotes the number of points in z or its cardinality, and $\mathbb{P}(|z| = k)$ is the probability that the cardinality of z is k. The likelihood is therefore computationally intractable as it is a sum of possibly infinite integrals since z is unobserved.

To calibrate their model to a set of measurement data collected in outdoor scenarios, Turin *et al.* reduced the dimension of the measured data by recording the peaks in the data along with their magnitudes. These served as estimates of the delays and the gains of the multipaths. The parameters of the model, such as the arrival rate of the Poisson process, were then estimated in a non-parametric manner based on the recorded multipath components. This calibration process is outlined in Fig. 4.

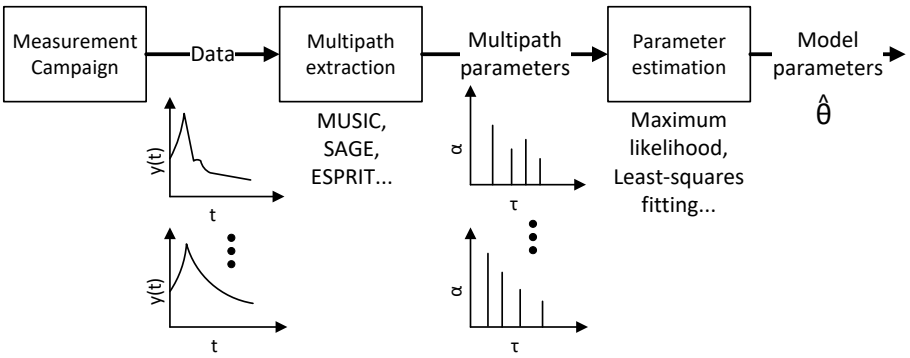


Fig. 4: Methodology for calibration of Turin model. First, the multipath components are extracted from the channel measurements, which are then used to estimate the model parameters.

Although the data-reduction procedure employed by Turin *et al.* was due to the limitations of the measurement equipment and data processing available at the time, their calibration methodology has been adopted by researchers till date. As a result, advanced high-resolution algorithms have been developed in the past couple of decades for extracting the multipath components from the channel response measurements. Examples of such high-resolution algorithms include CLEAN [12], SAGE [13], RiMAX [14], MUSIC, and ESPRIT, among others, see [15, Ch. 5] for an overview.

5.2 Calibrating the Saleh-Valenzuela model

In the seminal model proposed by Saleh and Valenzuela (S-V) in 1987 [16], which can be considered as a further advancement of Turin's model, the multipath components are modeled to arrive in clusters.

Example 5.2 (Saleh-Valenzuela model)
The impulse response can be stated as

$$h(t) = \sum_{l} \sum_{k} \alpha_{kl} \delta(t - T_l - \tau_{kl}), \tag{14}$$

where T_l, $l = 0, 1, \ldots$, is the delay of the l^{th} cluster, while τ_{kl} and α_{kl}, $k = 0, 1, \ldots$, are the delay and complex gain of the k^{th} ray within the l^{th} cluster, respectively. The delay of the first cluster and the first ray within the l^{th} cluster is set to zero, i.e. $T_0 = 0$ and $\tau_{0l} = 0$. The

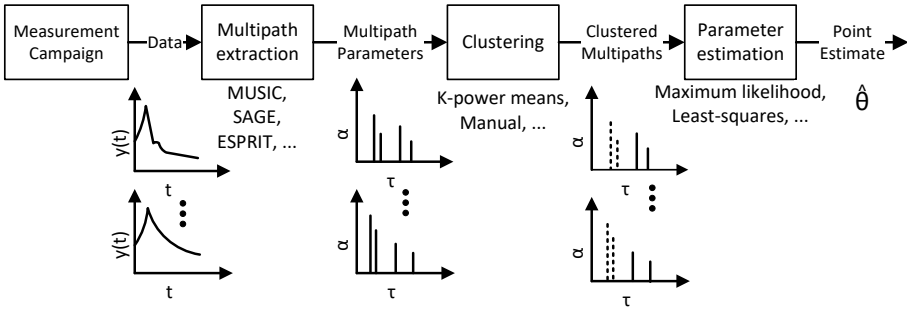


Fig. 5: Methodology for calibration of cluster-based models such as the one by Saleh-Valenzuela. The channel measurements are subjected to multipath extraction, followed by clustering. The parameters of the model are then estimated based on the extracted and clustered multipath components.

cluster delays, $\{T_l\}$, and the ray delays within each cluster, $\{\tau_{kl}\}$, are modeled as homogeneous Poisson point processes with arrival rates Λ and λ, respectively. Conditioned on $\{T_l\}$ and $\{\tau_{kl}\}$, the gains, $\{\alpha_{kl}\}$, are assumed iid zero-mean complex Gaussian random variables with conditional variance

$$\mathbb{E}\left[|\alpha_k l|^2 | T_l, \tau_{kl}\right] = Q \exp\left(-T_l/\Gamma\right) \exp\left(-\tau_{kl}/\gamma\right). \tag{15}$$

Here, Q is the average power of the first arriving multipath component, and $\Gamma, \gamma > 0$ are the cluster and ray decay constants. Thus, the parameter vector that needs to be calibrated for the S-V model is $\theta = [Q, \Lambda, \lambda, \Gamma, \gamma]^\top$.

The S-V model has intractable likelihood due to I2, and possibly I1. The calibration procedure of the S-V model is similar to that of the Turin model with the added step involving clustering of the multipath components, see Fig. 5. First, the multipath components such as the delays and the gains are extracted from channel responses using high-resolution algorithms. Then, the multipath components that are "close" to each other are assigned the same cluster. Finally, the S-V model parameters are estimated from the clustered multipath components via, e.g., least-squares fitting or maximum likelihood estimation.

The S-V model has formed the basis for many models such as the IEEE 802.15.3a [17], IEEE 802.15.4a [18], METIS [19], WINNER [20], and COST 2100 [21] models, among others. It has also been extended to

include the spatial properties of the radio channel in [22] and [23]. The multi-step calibration methodology of Fig. 5 has been followed to calibrate these models [17–27]. The clustering of the multipath components in [16], along with other early works [22, 26, 27], was done manually through visual inspection. However, the popularity of the clustering idea, coupled with the increasing size of data-sets, necessitated the need for automated clustering algorithms [28–30]. Development of clustering algorithms, e.g. for time-variant channel models [31–33], is still an active area of research.

6 Propagation graph model

Our third example model is the propagation graph (PG) model where the radio propagation mechanism has been modeled using graphs [34]. The PG model can account for propagation channels in multi-input, multi-output (MIMO) systems with N_t and N_r antenna ports at the transmitter and the receiver, respectively. In the PG framework, the propagation channel is represented as a directed graph where the vertices represent the transmitter, the receivers, and the objects in the environment called scatterers. The edges in the graph model the propagation conditions between the vertices.

Example 6.1 (Polarized propagation graph model)
Consider the directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the vertex set $\mathcal{V} = \mathcal{V}_t \cup \mathcal{V}_r \cup \mathcal{V}_s$ is a union of a set $\mathcal{V}_t$ of N_t transmitters, a set $\mathcal{V}_r$ of N_r receivers, and a set $\mathcal{V}_s$ of N_s scatterers in the environment, as shown in Fig. 6. The edge set $\mathcal{E} = \mathcal{E}_\mathrm{d} \cup \mathcal{E}_\mathrm{t} \cup \mathcal{E}_\mathrm{s} \cup \mathcal{E}_\mathrm{r}$ is a union of a set of direct edges, $\mathcal{E}_\mathrm{d}$, a set of transmitter to scatterer edges, $\mathcal{E}_\mathrm{t}$, a set of scatterer to scatterer edges, $\mathcal{E}_\mathrm{s}$, and a set of scatterer to receiver edges, $\mathcal{E}_\mathrm{r}$. A position vector $\mathbf{r}_v \in \mathbb{R}^3$ is associated with each vertex v. Thus, the length of an edge (v, w) is $||\mathbf{r}_v - \mathbf{r}_w||$, where c is the speed of light in vacuum and $\|\cdot\|$ is the Euclidean norm. The propagation delay from v to w is therefore $\tau_e = \|(\mathbf{r}_w - \mathbf{r}_v)\|/c$. The direction of propagation is specified by a unit vector Ω_e associated with edge e. Note that the transmit vertices and receive vertices can only have outgoing and incoming edges, respectively. In the following, we describe the extension of the propagation graph model that accounts for dual-polarized channels [35–37].

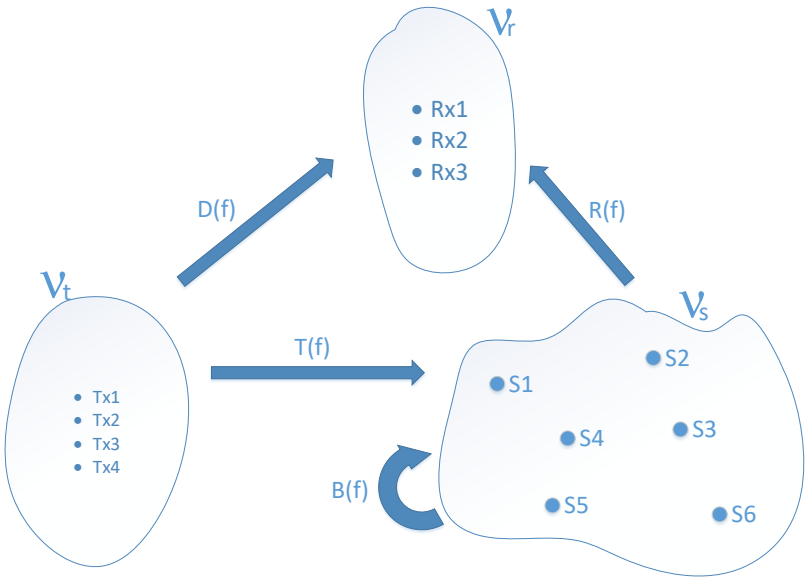


Fig. 6: A propagation graph with $N_t = 4$ transmit, $N_r = 3$ receive, and $N_s = 6$ scatterer vertices.

The transfer function matrix at a particular frequency, $\mathbf{H}(f) \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$, of the polarimetric PG model is given as

$$\mathbf{H}(f) = \mathbf{D}(f) + \mathbf{R}(f)[\mathbf{I} - \mathbf{B}(f)]^{-1}\mathbf{T}(f), \tag{16}$$

where $\mathbf{D}(f) \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ is the transmitter to receiver, $\mathbf{T}(f) \in \mathbb{C}^{2N_\mathrm{s} \times N_\mathrm{t}}$ is the transmitter to scatterer, $\mathbf{R}(f) \in \mathbb{C}^{N_\mathrm{r} \times 2N_\mathrm{s}}$ is the scatterer to receiver, and $\mathbf{B}(f) \in \mathbb{C}^{2N_\mathrm{s} \times 2N_\mathrm{s}}$ is the scatterer to scatterer edge transfer function sub-matrix having spectral radius less than unity. The transfer function sub-matrices are given as:

$$\begin{aligned}
\mathbf{D}(f) &= \mathcal{X}_\mathrm{t}^T(\Omega_e)\mathcal{X}_\mathrm{r}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{d} \\
\mathbf{T}(f) &= \mathcal{X}_\mathrm{t}^T(\Omega_e)\mathbf{M}\mathbf{\Gamma}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{t} \\
\mathbf{B}(f) &= \mathbf{M}\mathbf{\Gamma}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{s} \\
\mathbf{R}(f) &= \mathcal{X}_\mathrm{r}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{r}
\end{aligned}$$

Here, $\mathcal{X}_\mathrm{t}(\Omega_e)$ and $\mathcal{X}_\mathrm{r}(\Omega_e)$ are the 2×1 transmit and receive polarimetric antenna array response vectors, respectively, and $\mathbf{\Gamma}(\Omega_e)$ is the 2×2 rotation matrix. The coupling between the two polarization states is represented by the 2×2 scattering matrix, $\mathbf{M}$. Assuming the same

scattering matrix for all the scatterers, $\mathbf{M}$ reads

$$\mathbf{M} = \frac{1}{1+\gamma}\begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}, \tag{17}$$

where $\gamma \in (0,1)$ is the polarization ratio. The scalar $G_e(f)$ captures the polarization-independent propagation characteristics, and is expressed as

$$G_e(f) = g_e(f)\exp[j(\psi_e - 2\pi\tau_e f)], \tag{18}$$

where ψ_e is the phase. The edge gain, $g_e(f)$ is calculated as:

$$g_e(f) = \begin{cases} \frac{1}{(4\pi f\tau_e)}; & e \in \mathcal{E}_{\mathrm{d}} \\ \frac{1}{\sqrt{4\pi\tau_e^2 f\mu(\mathcal{E}_{\mathrm{t}})\mathcal{S}(\mathcal{E}_{\mathrm{t}})}}; & e \in \mathcal{E}_{\mathrm{t}} \\ \frac{g}{\mathrm{odi}(e)}; & e \in \mathcal{E}_{\mathrm{s}} \\ \frac{1}{\sqrt{4\pi\tau_e^2 f\mu(\mathcal{E}_{\mathrm{r}})\mathcal{S}(\mathcal{E}_{\mathrm{r}})}}; & e \in \mathcal{E}_{\mathrm{r}} \end{cases} \tag{19}$$

Here, $g \in (0,1)$ is the reflection gain, $\mathrm{odi}(e)$ denotes the number of outgoing edges from the n^{th} scatterer, and

$$\mu(\mathcal{E}_{\mathrm{a}}) = \frac{1}{|\mathcal{E}_{\mathrm{a}}|}\sum_{e\subset\mathcal{E}_{\mathrm{a}}}\tau_e, \quad \mathcal{S}(\mathcal{E}_{\mathrm{a}}) = \sum_{e\subset\mathcal{E}_{\mathrm{a}}}\tau_e^{-2}, \quad \mathcal{E}_{\mathrm{a}} \subset \mathcal{E},$$

with $|A|$ denoting the cardinality of set A.

The propagation graph model can be used to simulate the the behavior of the radio channel in a given propagation environment. The transmitter and receiver locations are either fixed or can be drawn randomly in a given room or building, while the number of scatterers in the environment, N_s, needs to be set. An edge between $\mathcal{V}_t$ and $\mathcal{V}_r$ is drawn with probability P_{dir}, which is equal to one for an edge in $\mathcal{E}_d$, and zero in non-line-of-sight case. The edges between $\mathcal{V}_{\mathrm{t}}$ and $\mathcal{V}_{\mathrm{s}}$, $\mathcal{V}_{\mathrm{s}}$ and $\mathcal{V}_{\mathrm{s}}$, or $\mathcal{V}_{\mathrm{s}}$ and $\mathcal{V}_{\mathrm{r}}$ are drawn with the probability of visibility, P_{vis}. The phases ψ_e are drawn independently from a uniform distribution on $[0, 2\pi)$. The other parameters of the model include the reflection gain, g, and, in case of the polarimetric PG model, the polarization ratio, γ, resulting in the parameter vector $\theta = [g, N_s, P_{\mathrm{vis}}, \gamma]^\top$.

The likelihood function of the PG model has not been derived in the literature (I3), and so it could potentially be intractable due to I1 or I2 or both. First proposed in 2007, the PG model has since been modified and applied to several propagation scenarios such as outdoor-to-indoor [38], indoor-to-indoor [39, 40], high-speed railway [41, 42], and office [43], and corridor [44]. In spite of the growing popularity of the PG model, not a lot of research has gone into developing calibration methods for it. Most of the studies utilizing the PG model make use of a map of the environment to calibrate its parameters. A method of moments approach was proposed in [35] to calibrate the polarimetric PG model to measurements. The method involved fitting estimates of the second moments of the power delay spectrum and cross-polarization ratio to their derived expressions.

7 Observations on Calibration Methods

From the state-of-the-art calibration procedures of the aforementioned stochastic channel models, it is evident that there is no general methodology for calibrating the models. Even though the quantity modeled by each of the models is the same, their calibration approach seems to be specific to the model structure. As a result, every time a new stochastic channel model is proposed or the previous models are modified, their calibration techniques need amendment as well. This is also the case with the method of moments approach of the polarized PG model as any modification in the model means re-deriving the moment equations. The lack of a clear and established calibration method means that researchers often employ ad-hoc and heuristic techniques to calibrate the models.

Besides the lack of generality, the current calibration methods also suffer from the following issues:

Unknown performance The multi-step approach of combining the multipath extraction and clustering algorithms makes it challenging to assess the performance of the calibration methods. As a result, it is common practice to report the point estimates of the model parameters without uncertainty quantification or statements of confidence.

Implementation choices The individual steps themselves require the implementation of sophisticated high-resolution and clustering al-

gorithms which can be taxing. Moreover, certain choices need to be made in order to implement these algorithms such as the number of multipaths to resolve and the number of clusters in the data. The accuracy of these methods rely on these choices, thus compounding matters even further.

Errors in each step The multipath extraction and clustering methods themselves are prone to errors such as censoring effects [45], and the effect of faulty estimation in the intermediate steps is difficult to quantify in the final parameter estimates. Moreover, such errors are largely ignored in subsequent steps.

Identifiability In the case of the PG model, the method of moments approach of [35] suffers from identifiability issues, and hence requires manually fixing one of the parameters. The S-V model suffers from a similar issue if only the power delay spectrum is used for calibration.

An important observation to make here is that none of the stochastic channel models are in fact calibrated using standard estimation techniques such as maximum likelihood or Bayesian inference. This is owing to the likelihood function being intractable for the stochastic channel models. For the stochastic multipath models, the intractability of the likelihood manifests as the points of the underlying point process are unobservable, as demonstrated for the Turin model.

The problem of having models with intractable likelihoods is not unique to the field of radio channel modeling. In fact, such models are used in various science and engineering fields. This begs the question of how such models are calibrated in other fields? And whether those methods can be adapted and applied to address the calibration issues faced by the stochastic radio channel models? In the next chapter, we will explore the available literature on likelihood-free inference methods.

Calibrating Models with Intractable Likelihoods

In this chapter, we look at some of the more widely used likelihood-free inference methods present in the literature, with the aim of choosing a method that best fits our purpose of having a general calibration method for stochastic radio channel models.

8 Likelihood-free Calibration Methods

Researchers across various fields are now moving towards using complex models to fit data, thanks to the rising computational power of affordable computers and advanced statistical methods. Such models act as a simulator, defining a stochastic procedure that directly generates data. Naturally, these models are used in fields such as climate and weather, population genetics, ecology, and astrophysics, where the mechanistic understanding of the real-world phenomenon can be used to directly create a data simulator, as is the case for stochastic channel models in the field of radio propagation. These simulator-type, or generative, models are flexible enough to replicate complicated data-sets and are easy to simulate realizations from given any value of the parameters. However, they do not have a tractable likelihood function, making it an arduous task to infer their parameters from data using standard techniques such as maximum likelihood and Bayesian statistics. Overcoming this limitation therefore becomes crucial in order to use such models for simulation purposes. Unsurprisingly, a host of likelihood-free inference methods have been developed to calibrate the simulator-type models in those fields.

Needless to say, the likelihood-free inference methods are utilized in cases where the likelihood function of the model is intractable, thus

rendering maximization of the likelihood or Bayesian inference unrealizable. Some of the popular methods used for calibrating such models are described in the following. For a recent survey of such methods and future directions, see [46].

8.1 Method of Moments

This is a fairly well-known method of estimation used across a multitude of fields [47] including signal processing [8]. It begins by expressing the moments of the data, say $M(\cdot)$, as a function of the parameters of interest. Those analytical expressions are then fit to the moments obtained from the data, i.e., $M(\theta) = \hat{M}(y)$, with $\hat{M}(\cdot)$ estimated from the data by replacing the expectation $\mathbb{E}[\cdot]$ by averaging. Provided that $M(\cdot)$ is invertible, the parameters can be estimated as

$$\hat{\theta} = M^{-1}(\hat{M}(y)). \tag{20}$$

The number of moment expressions therefore needs to be equal to the number of parameters of the model. This method has the advantage of being fairly simple to implement with negligible computational load, although the resulting estimators can often be biased. There can be identifiability issues in case $M(\cdot)$ is not invertible. This can be resolved by considering more moment conditions than parameters and solving the overdetermined system of equations in a generalized method of moments approach [47].

For most generative models, deriving moment expressions is not feasible, hence the application of this method is fairly limited. However, in the context of radio channel models, method of moments has been used to calibrate the polarized PG model in [35], and the S-V model in [48], albeit with approximations.

8.2 Approximate Bayesian Computation

Approximate Bayesian computation (ABC) [49, 50] is a likelihood-free inference method that permits sampling from the approximate posterior of a generative model, see Fig. 7. Although methods with similar flavor had been proposed in 1984 by Rubin [51], ABC was first introduced in population genetics in the late 1990s by Tavare *et al.* [52] and Pritchard *et al.* [53].

As the name suggests, ABC falls under the Bayesian paradigm and hence assumes a prior distribution on the parameters of the model. ABC

relies on repeated model evaluations using parameter values sampled from the prior. Simulated data from the model is then compared to the observed data in some distance metric $\rho(\cdot, \cdot)$. Parameter samples that yield simulated data "close" to the observed data are taken as samples from the approximate posterior distribution. The closeness here is determined by the tolerance threshold ϵ. Given a model with parameters θ and data y, the most basic form of ABC, called rejection-ABC, proceeds as follows:

1. Draw θ^* from the prior distribution $p(\theta)$
2. Simulate y^* from the model using θ^*
3. Accept θ^* if $\rho(y, y^*) < \epsilon$

Repeating the above steps till N parameter values are accepted results in the sample $(\theta_1, \ldots, \theta_N)$ from the approximate posterior $p(\theta|\rho(y, y^*) < \epsilon) \approx p(\theta|y)$. Here, the approximation accuracy is controlled by ϵ. Taking ϵ very small leads to better posterior approximation but high rejection rates, while taking it too big gives low rejection rates but makes the approximate posterior look like the prior.

The rejection ABC method outlined above runs into issues when the data is high-dimensional, as is the case with impulse response measurements. Since the distances between points increase exponentially with the dimensions of the space, it is common practice to compute the distances between summary statistics of the data. The rejection criterion in the ABC algorithm is then replaced by $\rho(S(y), S(y^*)) < \epsilon$, where $S(\cdot)$ is the summarizing function that maps the data y into a low-dimensional vector of its statistics $\mathbf{s}$. Unless sufficient statistics are available, this summarization step leads to some loss of information, thereby adding to the approximation of the posterior.

Implementing an ABC method requires the choice of the three ingredients — the distance metric ρ, the summary statistics $S(\cdot)$, and the tolerance ϵ. The distance metric is typically taken to be the Euclidean distance in case informative summary statistics are available. Other alternative is to take ρ to be a distance defined on (empirical) probability distributions, such as the maximum mean discrepancy [54] in [55] or the Wasserstein distance [56] in [57]. Statistical methods exist in the ABC literature to obtain optimal summaries using projection techniques, see [58] for a review. In [59], it was shown that the posterior mean is the optimal summary statistic for a quadratic loss function. Such optimal

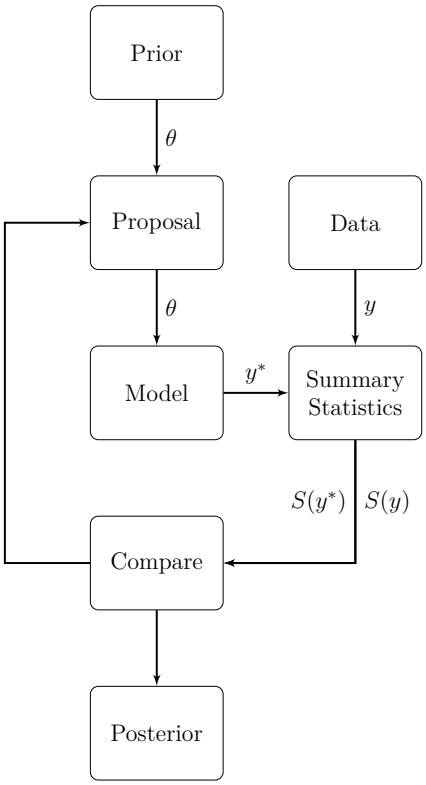


Fig. 7: Overview of ABC methods.

statistics were estimated using linear [59] and non-linear regression [60]. However, domain knowledge is crucial for designing relevant statistics to be used for fitting the model.

Several developments have been made in the ABC literature with regards to exploring the parameter space more efficiently than simply sampling independently from the prior as in rejection-ABC. An alternative method of embedding a Markov chain Monte Carlo sampler in the ABC framework was proposed in [61]. Another class of ABC methods are based on the sequential Monte Carlo techniques [62, 63], where the estimated posterior is successively refined by resampling from the obtained parameter values.

8.3 Minimum Distance Estimators

The minimum distance (or divergence) estimators are a class of estimators that minimize some notion of distance between an empirical distribution obtained from the data and the model $\mathbb{P}_\theta$ [64]. Considering the observed data $(y_1, \ldots, y_N)$ that yields an empirical distribution[2] $\mathbb{Q}^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i}$, where δ_{y_i} denotes a Kronecker delta with mass one at y_i, such an estimator can be written as

$$\hat{\theta}_{\mathcal{D}} = \text{argmin } \mathcal{D}\left(\mathbb{P}_\theta || \mathbb{Q}^N\right). \tag{21}$$

Note that taking $\mathcal{D}$ to be the Kullback-Leibler divergence results in the maximum likelihood estimation of θ. Solving Eq. 21 analytically is not possible in most practical cases, hence employment of gradient-based optimization techniques is common practice.

In the context of models with intractable likelihoods, minimum distance estimators have been proposed based on the maximum mean discrepancy in [65]. Minimum distance estimators are closely related to the method of simulated moments [66], where some norm of certain moments of the observed and simulated data are minimized for the parameter values. Note that this idea is similar to the summary-free ABC methods [55] where synthetic data-sets simulated from the model are compared to the observed data in some distance metric. Hence, the minimum distance estimators can be considered as the frequentist version of the ABC methods.

8.4 Surrogate Modeling

This is a classical approach where an approximate model, or surrogate, is used as a substitute for the data-generating process, especially when gathering data is expensive [67]. In the context of calibration, the idea is to replace the generative model by a tractable surrogate model, following which, frequentist or Bayesian inference can proceed as if the likelihood was tractable. An example is [68], where a model of the likelihood was created by estimating the distribution of simulated data with kernel density estimation. Block diagram depicting calibration using surrogate models is shown in Fig. 8.

The application of such methods has been limited as they do not scale well to high-dimensional data. However, the advent of the powerful and

[2]Equivalently, the probability density function for $\mathbb{Q}^N$ reads $\frac{1}{N}\sum_{i=1}^{N} \delta(y_i - y)$, where δ is the Dirac delta.

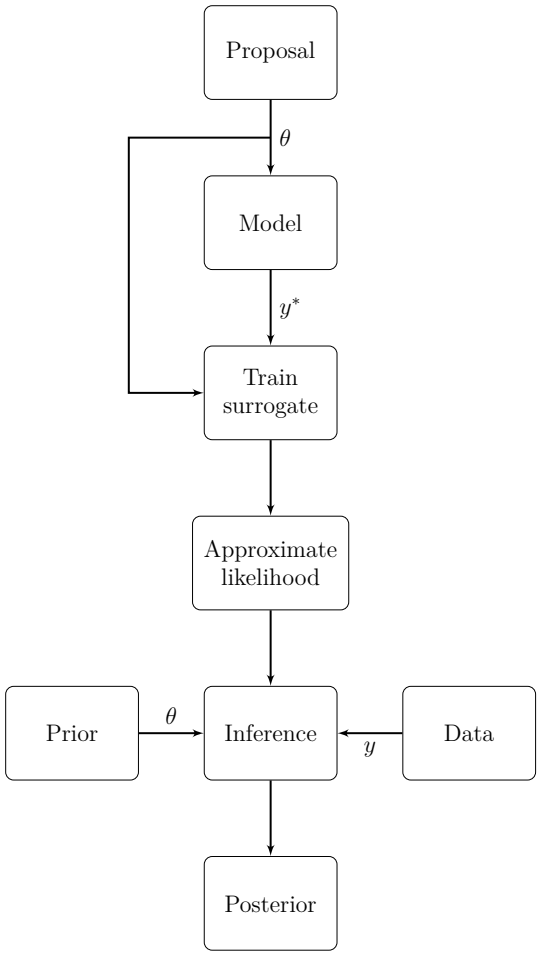


Fig. 8: Overview of calibration methods using surrogate modeling..

versatile machine learning tools such as the deep neural networks have made the task of density estimation feasible in high dimensions. As a result, neural networks are being used as surrogates for the generative model in many different scenarios, including for summaries of stochastic channel models [69]. The density estimation can be carried out both in a supervised or unsupervised manner. One of the major advantages of such methods is that after an upfront simulation and training phase, the surrogates can be used to evaluate new data points efficiently. A detailed survey and classification of various surrogate modeling techniques can be found in [46].

8.5 Synthetic likelihood

This is another popular likelihood-free inference method introduced in the field of ecology in 2010 [70]. Synthetic likelihood assumes that some summary statistics of the data, $\mathbf{s} = S(y)$, which are informative about the model parameters θ, can be well-modeled, or approximated, by a multivariate Gaussian distribution, i.e. $\mathbf{s} \sim \mathcal{N}(\mu, \Sigma)$. Simulations from the model are then used to estimate the mean vector μ and the covariance matrix Σ, thus approximating the likelihood function with the multivariate Gaussian density. This approximate likelihood can either be maximized to estimate the parameters in a frequentist manner, or used in a Bayesian setting to sample from the approximate posterior distribution [71]. The synthetic likelihood methods can be considered as a form of surrogate modeling on the summary statistics of the data. Since synthetic likelihood methods are limited by the normality assumption of the summary statistics, more flexible surrogate likelihoods for summaries have been proposed in [72–74].

9 Summaries for Radio Channel Measurements

Summary statistics, e.g. moments of the data, play an integral part in a large majority of likelihood-free inference methods. Therefore, we need to select or design summary statistics for radio channel measurements that are informative about the parameters of stochastic channel models in order to implement such methods. As alluded to before, there are statistical techniques developed in the ABC literature in particular to design optimal summaries. However, domain-specific summaries facilitate interpretability and knowing exactly what aspect of the channel is used for fitting the models.

The underlying marked point process Z forms the sufficient statistic for channel impulse response measurements. If Z is known, then the likelihood function takes the form of a complex Gaussian density due to the noise being the only random entity. However, we do not know how many multipath components are present in the data. Moreover, these need to be estimated using high-resolution algorithms. Alternatively, the channel could be summarized into other easy-to-compute statistics that are informative about the model parameters.

Some of the most widely used summaries for characterizing the radio channel are the received power, mean delay, and root mean square (rms) delay spread [75], see Fig. 9 for their interpretation. They are frequently

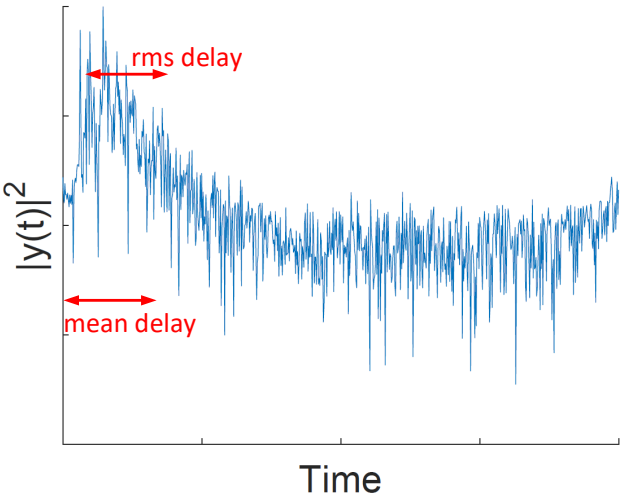


Fig. 9: Example of the power of a radio channel measurement $y(t)$ as a function of time delay. Mean delay is the 'centre of gravity' of the multipath components, while the rms delay quantifies their spread.

used in the design and analysis of communication and localization systems as they summarize important aspects of the radio channel and the multipath effect. Furthermore, cumulative distribution functions plots of these summaries are often used as means to report channel measurements.

The received power P_0, mean delay $\bar{\tau}$ and rms delay spread τ_{rms} are derived from transformations of the temporal moments of the channel measurement $y(t)$ according to

$$P_0 = m_0, \quad \bar{\tau} = \frac{m_1}{m_0}, \quad \text{and} \quad \tau_{\text{rms}} = \sqrt{\frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2}, \tag{22}$$

where the k^{th} temporal moment is defined as

$$m_k = \int t^k |y(t)|^2 \text{dt}, \quad k = 0, 1, 2, \ldots \tag{23}$$

The temporal moments can be seen as an expansion of $|y(t)|^2$ into the basis of monomials, which forms a complete basis for finite energy time-limited signals [76]. Their computation can be carried out by simply using a numerical integration for the integral in (23). Moreover, they are easy to interpret as it is evident what aspects of the channel is being

Table 1: The main choices and requirements for implementing the likelihood-free calibration methods described in Sec. 8.

Method	Choice	Requirement
Method of moments	Moment equations	Should be invertible
ABC	Summary statistics, Distance metric	Informative summaries
Minimum distance estimators	Distance metric	Solving Eq. 21
Surrogate modeling	Surrogate	Training the surrogate
Synthetic likelihood	Summary statistics	Estimating μ and Σ

used to fit the model. The mean of the temporal moments is informative about the power delay spectrum, $P(t)$, of the channel, that is

$$\mathbb{E}[m_k] = \int t^k \mathbb{E}[|y(t)|^2] dt = \int t^k P(t) dt, \tag{24}$$

while their covariance has been shown to be informative about the second-order intensity function of stochastic multipath models [77]. In addition, temporal moments are widespread in the field of channel modeling. For these reasons, the temporal moments appear to be a reasonable choice for the summary of the channel measurements to be used in calibration of stochastic radio channel models.

10 Choice of Method for Stochastic Channel Models

Now that we have reviewed some of the popular methods available in the literature for calibrating models with intractable likelihoods, see Table 1 for an overview, our task is to choose a method to be used for calibration of stochastic radio channel models. The aforementioned likelihood-free inference techniques can be broadly separated into two categories — the ones that use the model itself during inference, like ABC, and the methods which construct a surrogate model and use that for inference. The latter methods have the advantage that inference on new datasets can be obtained almost instantaneously once the surrogate model has been trained. However, our focus is on developing a calibration method which is general in the sense that it can be applied to stochastic channel models having different mathematical structure and parameters. With this aim of having a general calibration approach, it makes sense to include the model in the inference method. Hence, ABC methods seem to be an ideal choice in that case, considering the availability of informative summary statistics such as the temporal moments.

Design of Study

This chapter outlines the research questions being investigated in this thesis and the research methodology adopted to tackle those questions.

11 Research Hypothesis and Questions

As discussed in Chapter 2, the calibration procedure for stochastic radio channel models is specific to their mathematical structure, and there does not exist a standardized method to calibrate such models. In Chapter 3, we have identified ABC methods as a potential candidate to achieve our goal. Therefore, the research hypothesis being tested in this thesis can be summarized as:

Stochastic radio channel models can be calibrated in a general manner using approximate Bayesian computation methods.

The hypothesis motivates a number of research questions as outlined below. The research questions being answered in this thesis can be broadly categorized as follows:

- *Can stochastic multipath models, such as the Turin and the S-V model, be calibrated without multipath extraction? Can likelihood-free inference methods be used to calibrate these models?*

  Given the widespread use of high-resolution algorithms for calibration, the first question that we need to answer is whether it is even possible to calibrate stochastic multipath models without them. Answering this question entails calibrating such models using a different method, thus answering the other sub-question. Even though these questions are answered in all the papers in this thesis where we have proposed a new calibration procedure, this

has been one of the primary conclusions of the work in Papers A–C.

- *Can ABC methods be used to calibrate different stochastic radio channel models? If so, which summary statistics of the temporal moments are needed to implement such ABC methods? Can such statistics be automatically generated?*

  Before arriving at a general ABC method for stochastic models, we need to see if informative summary statistics can be designed to calibrate such models using ABC. As indicated in Chapter 3, the temporal moments could be a viable starting point for designing such statistics. In Paper D and E, we get an answer to these questions in the context of S-V and PG model, respectively. Paper F answers the sub-question related to the automatic generation of summary statistics.

- *How are the temporal moments of the radio channel distributed in different propagation scenarios? How can they be modeled jointly?*

  Since the temporal moments are widely used summaries for characterizing the radio channel and are found useful for calibration, it is straightforward to ask how can they be modeled jointly. Paper G and H provide an answer to this question.

- *Can different stochastic channel models be calibrated using the same ABC method? Can the distance metric in the ABC method be adapted in order to circumvent the need for summarization?*

  Finally, we arrive at the questions that directly deals with our stated hypothesis. These questions are answered in Paper I.

12 Research Methodology

We now outline the research methodology adopted while developing the calibration methods presented in this thesis. A summary of the methodology is presented in Alg. 1. Note that the following discussion is not pertinent to the modeling works of Paper G and H.

Before developing any calibration method, there are a couple of choices to make, namely the model being calibrated, and the assumptions regarding the parameters. The choice of the model readily impacts the calibration method. In this thesis, the three models described in

Chapter 2 serve as examples for the calibration methods. We begin by considering the most basic stochastic channel model available, i.e. the Turin model with simplifying assumptions of no noise. Once we develop a calibration method for such a simple model, we gradually ease the assumptions and move on to more complicated S-V and PG models. This approach brings clarity to the development process of calibration methods, as we can keep track of the challenges brought forth by each added level of complexity.

The other choice is between a frequentist or a Bayesian approach for estimating the parameters, which basically manifests in terms of choosing either a parameter space or a prior distribution. In this thesis, this choice is determined by the inference method being considered. Nevertheless, we have tried to rely solely on the data even in case of Bayesian inference, and hence selected uninformative, or flat, priors for the parameters in all of the papers barring Paper A.

Having selected the example model and defined the parameter space or prior range, we begin the process of developing calibration algorithms. We start by fixing a "true" value for the parameters, which we use to simulate a data-set from the model. We use this data-set as our pseudo-observed data, i.e. the data on which the model is being fitted. Use of simulated data enables us to evaluate the performance of the calibration method by measuring how accurately we are able to recover the true parameter values. The development of the calibration method is therefore an iterative process where the method is modified in each iteration based on the simulation results. Once peak performance of the calibration method is achieved, the loop is terminated and the method is deemed "ready".

We then validate the developed calibration method by applying the method on a set of real measurements and comparing the model's prediction with that of the measurements. Since no true parameters are available in the case of real data, we input the estimated parameter values into the model and make a qualitative comparison of the simulated and measured data in terms of certain relevant statistics. This serves as a reasonable way to qualify the estimated parameters, although in principle, the model's predictions should be compared to a separate set of validation data that was not used in the calibration procedure. Collecting data by conducting measurement campaigns is considered out of the scope of this thesis. We instead use the channel measurements collected at Lund University [23] by Prof. Fredrik Tufvesson's group to

Algorithm 1 Methodology for developing calibration algorithms

Input: Observed data, model, parameter space or prior distribution

Assumption: Data is informative about the model parameters

Set true parameter values
Generate pseudo-observed data using the model and the true parameters

while true parameters are not recovered accurately **do**
  Propose a calibration method
  Evaluate performance of the method on simulated data
end while
Apply the calibration method on observed data
Compare model prediction with observed data for validation

Output: Calibration method

validate our proposed methods.

We remark that applying the method on measured data is not essential for developing calibration methods. If the method is able to recover the true parameters accurately in simulation, that implies that the calibration method is working for that particular model. However, this additional step can be quite useful in pointing out any differences between the model assumptions and the measurement procedure. That is, if the calibration method works on simulated data but not on real data, it could very well mean that the model assumptions are conflicting with the data measurement process. If that is indeed found to be the case, then the model is adjusted accordingly and the simulation experiments are repeated again. Alternatively, this could be accounted for in the calibration method itself instead of adjusting the model. Validating the calibration method on real data also gives confidence to the readers regarding its applicability. Note that validation of the calibration method should not be mistaken for model validation. A poor fit of the model on real data could be due to the model not being right for the data at hand, instead of the calibration method being faulty.

Thesis Contributions

In this chapter, a summary and key contributions of all the papers included in this thesis are presented. The order of the papers reflects the gradual progression of research during the period of the study. A short summary of the key components of each paper related to calibration is provided in Table 2.

Paper A In this paper, we begin our investigation into likelihood-free calibration methods for stochastic channel models by considering the most basic homogeneous Turin model as our example. Moreover, we make simplifying assumptions of no additive noise and large bandwidth that yields an expression of the temporal moments in terms of the multipath components. We then use a synthetic likelihood approach and model the first three temporal moments as a multivariate Gaussian distribution. The expression for the mean vector and the covariance matrix are then derived in terms of the model parameters. With a closed form expression for the approximate likelihood of the temporal moments, sampling from the posterior distribution is carried out by assuming priors on the parameters. We find that the temporal moments are informative about the parameters of the Turin model, and that it can be calibrated without multipath extraction.

Paper B In this paper, we again consider the homogeneous Turin model as our example, but without the no noise and large bandwidth assumptions. This results in an additional parameter to be estimated, namely the noise variance. We then derive closed form expressions for the means of the first three temporal moments and the variance of the zeroth temporal moment. With four parameters to be estimated and four expressions, we use a method of moments approach to derive the estimator. Simulation experiments indicate

Table 2: Overview of Papers A–F and Paper I.

Paper	Method	Example	Approach
A	Synthetic likelihood	Turin	Bayesian
B	Method of Moments	Turin	Frequentist
C	Maximum likelihood	Turin	Frequentist
D	ABC	S-V	Bayesian
E	ABC and Surrogate modeling	PG	Both
F	ABC	PG	Bayesian
I	ABC	S-V and PG	Bayesian

accurate recovery of the true parameters. The calibration method applied on measured data provides good fit to the averaged power delay spectrum. The proposed estimator is therefore found to calibrate the homogeneous Turin model without multipath extraction using temporal moments. Moreover, the method is able to estimate the noise, thus eliminating any need to truncate the channel response to estimate the noise floor.

Paper C In this paper, we take a maximum likelihood estimation approach for calibrating the Turin model with both homogeneous and inhomogeneous arrival rate. The method relies on approximating the likelihood ratio using a Markov chain Monte Carlo sampler from the conditional distribution of the multipath components given the data. In addition to calibrating the model, we show that the likelihood ratio computation can be used for model selection. The method is tested on simulated as well as measured data. We conclude that in the case of the Turin model, calibration can be achieved via the well-established method of maximum likelihood estimation without resorting to multipath extraction.

Paper D The derivations of the moment equations and the likelihood function may become infeasible as we look at more complicated models such as the S-V model. Therefore, in this paper we look at a summary-based ABC solution to the calibration problem, with the Euclidean distance as our comparison metric. Taking the cluster-based S-V model as an example, we design summary statistics based on the temporal moments and the power delay profile of the channel. We use these summaries in a sequential Monte Carlo ABC framework, called population Monte Carlo (PMC) ABC, wherein

we iteratively converge towards the desired approximate posterior. A regression adjustment step is also employed at each iteration to improve the posterior approximation. The resulting algorithm, named PMC-ABC with regression adjustment, will also be utilized in rest of the papers dealing with ABC. From the simulation experiments, we find that the method is able to recover all the parameters of the S-V model fairly accurately except the cluster arrival rate. However, the cluster arrival estimate can be improved via an additional round of the ABC method. Similar results are obtained on applying the method to real measurements. We conclude that the S-V model can be calibrated without the need for multipath extraction and clustering, and that the designed summaries are informative about the model parameters.

Paper E In this paper, we apply the proposed PMC-ABC algorithm with regression adjustment on the polarized PG model by handcrafting summaries using the temporal moments. A methodology for qualifying the summaries is also provided. Additionally, we also propose a frequentist calibration method based on surrogate modeling for the polarized PG model. We use a deep neural network model as a surrogate for the same handcrafted summaries as used in the ABC method. The neural network is trained in a supervised manner with the handcrafted summaries as input and the parameters as output labels. The two methods are shown to estimate the parameters accurately in simulations, and provide a good fit to the averaged power delay profile when applied to measured data. The proposed methods also perform better than the state-of-the-art calibration method for the polarized PG model. We conclude that the summaries designed using the temporal moments are informative about the parameters of the polarized PG model. By calibrating the PG model using the same ABC method as the S-V model, we show that stochastic channel models can be calibrated using the same procedure as long as relevant summaries are available.

Paper F Applying the ABC method of Papers D and E to calibrate stochastic channel model requires handcrafting summary statistics that are informative about all the model parameters, which is a time-consuming task and may be non-trivial. Hence, in this paper, we propose a method to automatically generate summary statistics

using an autoencoder, which is a combination of two neural networks, namely the encoder and the decoder. The encoder projects the input data into a low-dimensional vector of features, and the decoder replicates the input data by decoding those features. We train the autoencoder with the first three temporal moments as input, and use the encoded feature vector as summary statistics in the PMC-ABC with regression adjustment algorithm. The proposed method is used to calibrate the PG model in simulations as well as on measured data. Although the performance of the auto-generated summaries is not as accurate as the handcrafted ones as in Paper E, we do get reasonably accurate results with fairly limited effort of training the autoencoder. Hence, the method is more generally applicable to different stochastic channel models, provided the autoencoder is trained properly.

Papers G & H With the temporal moments being informative summaries of the radio channel and useful for calibrating the stochastic channel models, it is natural to inquire about their joint behavior and how to characterize them. In these two papers, we investigate the joint distribution of the first three temporal moments based on empirical evidence. We propose to model them jointly using a multivariate log-normal distribution, and show that this model is able to fit the temporal moments data collected from different environments and measurement settings. Modeling the temporal moments as a log-normal random vector consequently provides a straightforward method to simulate the widely used moments of the channel such as the mean delay and the rms delay spread. We compare the proposed model with other competing model choices in the literature both qualitatively and quantitatively. We find that these moments are correlated random variables, and hence, should be simulated jointly. We conclude that joint modeling of the temporal moments, and thereby the received power, mean delay and rms delay spread, provides more accurate models in a wide range of scenarios as compared to independent modeling, which is prevalent in the literature. The proposed model of multivariate log-normal distribution is simple, easy to simulate from, and provides a reasonable fit in both indoor and outdoor settings.

Paper I In this paper, we propose a general ABC method that is able to calibrate different stochastic channel models using the same pro-

cedure. We use the MMD, which is a similarity measure between distributions or data-sets, as the distance metric in the PMC-ABC algorithm. Thus, our ABC rejection criteria involves comparing the joint distribution of the simulated temporal moments with the observed temporal moments using the MMD. Since the temporal moments are general purpose summaries of the radio channel, we are able to calibrate both the S-V and the PG model using this procedure, even though they have very distinct mathematical structures. This approach circumvents the need to design appropriate summaries or train an autoencoder when faced with a new channel model, thus imparting generality to the method. Moreover, the proposed method is able to detect and account for model misspecification, wherein the model is unable to replicate the observed data for any value of the parameters. Using the first four temporal moments, the proposed method gives accurate results for both simulated as well as measured data. Furthermore, the method is automatic in that no pre- or post-processing of the data or estimates is required from the user.

Conclusions and Outlook

This thesis concerns with the development of calibration methods for stochastic radio channel models, primarily using approximate Bayesian computation (ABC) methods. Based on the contributions of the thesis, we draw the following main conclusions:

- The stochastic radio channel models can be calibrated without extraction and clustering of multipath components, as is the norm in the literature, thus avoiding the need for implementing high-resolution and clustering algorithms.
- Likelihood-free inference methods, especially those based on ABC, are a feasible alternative to the state-of-the-art calibration methods. The temporal moments are informative summaries of the radio channel that are useful for implementation of these ABC methods. Summaries can either be handcrafted, as was the case in Paper D and Paper E for S-V and PG model, respectively, or can be automatically generated as shown in Paper F.
- The temporal moments are correlated random variables that can be well-modeled using a multivariate log-normal distribution. This provides a simple and straightforward method for jointly simulating the received power, mean delay and rms delay spread.
- We show that ABC methods enable the calibration of different stochastic channel models using the same general method in a statistically sound manner. This was achieved by using a distance metric on the data-sets, namely the maximum mean discrepancy (MMD), instead of specializing summaries as shown in Paper I.

The progressive development of calibration methods for stochastic channel models over the duration of this study has resulted in Paper I being the culmination of the conducted research. Hence, the calibration

method in Paper I retains the advantages of the previously proposed methods over the state-of-the-art, and at the same time overcomes their drawbacks or shortcomings. These advantages are summarized as follows:

1. Performance of the proposed calibration methods is straightforward to assess as information on the posterior distribution is obtained. The end-to-end performance evaluation of the state-of-the-art calibration methods is infeasible due to the multi-step approach that leads to a point estimate of the parameters without any uncertainty quantification.

2. As the proposed methods estimate the noise level in the measurements accurately, any pre-processing of the measurements to remove the effect of the noise before calibration is unnecessary.

3. Compared to the state-of-the-art approaches, the method proposed in Paper I has limited number of choices or settings.

4. The generality of the proposed method in Paper I, exemplified by the S-V and the PG model, enables comparison of the fit of different stochastic channel models to a particular data-set, as these models can be calibrated using the same method.

5. Compared to the other ABC methods, the method in Paper I achieves better approximation to the posterior distribution as it does not require summarization of the temporal moments.

6. The ability of the proposed method to detect model misspecification can be utilized by channel modeling experts to assess adequacy of a model before performing tedious analytical derivations of its likelihood function.

The bottle-neck in the computational cost of the proposed ABC methods is the repeated sampling from the model, which was not significant for the models considered in this thesis. This may not be the case for a computationally expensive model, making their calibration infeasible in reasonable time. Another limiting factor of the proposed method could be if the model has a large set of parameters to be calibrated, thus invoking the curse of dimensionality. These issues can potentially be addressed by employing Bayesian optimization in the ABC framework as done in [78].

As shown in Paper H, the temporal moments are very correlated random variables, indicating redundancies in the information contained in them. Potentially, the magnitude squared of the channel can be expanded using an orthogonal basis, which may lead to better efficiency of the proposed methods.

An obvious extension of this work is to develop calibration methods for directional channel models where the dispersion is in delay as well as direction. That would entail the use of certain "directional moments", analogous to the temporal moments, that are informative about the parameters governing the directional spread. Another research avenue to explore would be the use of signature moments [79, 80] in the proposed ABC methods.

Finally, the application of ABC and other methods described in this thesis are not constrained to radio propagation models, as generative models with intractable likelihoods possibly appear in other sub-fields of wireless communications. Similarly, the use of MMD as a similarity measure can potentially be utilized in other problems in communication or localization where there is a need to compare different distributions or data-sets. We hope that researchers in the field would be inspired to apply these methods and techniques on their problems.

References

[1] T. Pedersen, *Contributions in radio channel sounding, modeling, and estimation : Ph. D. thesis.* Aalborg: Department of Electronic Systems, Aalborg University, 2009.

[2] J. Proakis, *Digital communications.* Boston: McGraw-Hill, 2008.

[3] G. Steinböck, *Modeling of reverberation effects for radio localization and communications : Ph. D. thesis.* Aalborg: Department of Electronic Systems, Aalborg University, 2013.

[4] M. Jakobsen, *Modeling and analysis of stochastic radio channels : an application of the theory of spatial point processes : PhD thesis.* Aalborg: Navigation and Communications, Department of Electronic Systems, Aalborg University, 2013.

[5] R. Kouyoumjian and P. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proceedings of the IEEE*, vol. 62, no. 11, pp. 1448–1461, 1974.

[6] D. Gamerman, *Markov chain Monte Carlo : stochastic simulation for Bayesian inference.* Boca Raton: Taylor & Francis, 2006.

[7] A. Doucet, N. Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice.* Springer New York, 2001.

[8] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II.* Pearson Education (US), 1998. [Online]. Available: https://www.ebook.de/de/product/3239830/steven_kay_fundamentals_of_statistical_signal_processing_volume_ii.html

[9] S. Salous, *Radio Propagation Measurement and Channel Modelling*, 1st ed. Wiley Publishing, 2013.

[10] M. L. Jakobsen, T. Pedersen, and B. H. Fleury, "Analysis of the stochastic channel model by saleh & valenzuela via the theory of point processes," *Int. Zurich Seminar on Commun.*, 2012.

[11] 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. 21, no. 1, pp. 1–9, Feb 1972.

[12] J. A. Högbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *Astronomy and Astrophysics Supplement Series*, vol. 15, p. 417, Jun. 1974.

[13] 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.

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

[15] X. Yin and X. Cheng, *Propagation Channel Characterization, Parameter Estimation, and Modeling for Wireless Communications*. John Wiley & Sons Singapore Pte. Ltd, Feb 2018.

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

[17] J. Foerster, "Channel modeling sub-committee report final," Tech. Rep. IEEE doc.: IEEE P802.15-02/490r1-SG3a, Feb 2003.

[18] A. F. Molisch, "IEEE 802.154a channel model-final report," Tech. Rep. IEEE doc.: IEEE 802.15-04-0662-02-004a, 2005.

[19] L. Raschkowski, P. Kyösti, K. Kusume, and E. T. Jämsä, "METIS channel models, deliverable d1.4 v3," Tech. Rep. ICT-317669 METIS Project, 2015.

[20] P. Kyösti, "Winner II channel models, deliverables d1.1.2 v1.2, part I: Channel models," Tech. Rep. IST-4-027756 WINNER II Project, 2007.

[21] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, dec 2012.

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

[23] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mmwave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

[24] 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. on Wireless Commun.*, vol. 18, no. 8, pp. 3856–3870, 2019.

[25] M. Yang, B. Ai, R. He, G. Wang, L. Chen, X. Li, C. Huang, Z. Ma, Z. Zhong, J. Wang, Y. Li, and T. Juhana, "Measurements and cluster-based modeling of vehicle-to-vehicle channels with large vehicle obstructions," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 9, pp. 5860–5874, 2020.

[26] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. F. Molisch, "A measurement-based statistical model for industrial ultra-wideband channels," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 8, pp. 3028–3037, 2007.

[27] Chia-Chin Chong and Su Khiong Yong, "A generic statistical-based uwb channel model for high-rise apartments," *IEEE Trans. on Antennas and Propag.*, vol. 53, no. 8, pp. 2389–2399, 2005.

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

[29] 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.

[30] 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.

[31] R. He, Q. Li, B. Ai, Y. L. Geng, A. F. Molisch, V. Kristem, Z. Zhong, and J. Yu, "A kernel-power-density-based algorithm for channel multipath components clustering," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7138–7151, Nov 2017.

[32] C. Huang, R. He, Z. Zhong, B. Ai, Y.-A. Geng, Z. Zhong, Q. Li, K. Haneda, and C. Oestges, "A power-angle-spectrum based clustering and tracking algorithm for time-varying radio channels," *IEEE Trans. on Veh. Technol.*, vol. 68, no. 1, pp. 291–305, jan 2019.

[33] C. Huang, A. F. Molisch, Y. Geng, R. He, B. Ai, and Z. Zhong, "Trajectory-joint clustering algorithm for time-varying channel modeling," *IEEE Trans. on Veh. Technol.*, vol. 69, no. 1, pp. 1041–1045, 2020.

[34] T. Pedersen and B. H. Fleury, "Radio channel modelling using stochastic propagation graphs," in *IEEE ICC*, June 2007, pp. 2733–2738.

[35] R. Adeogun, T. Pedersen, C. Gustafson, and F. Tufvesson, "Polarimetric Wireless Indoor Channel Modelling Based on Propagation Graph," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 10, pp. 6585–6595, 2019.

[36] R. Adeogun and T. Pedersen, "Propagation graph based model for multi-polarized wireless channels," in *IEEE WCNC*, April 2018.

[37] ——, "Modelling polarimetric power delay spectrum for indoor wireless channels via propagation graph formalism," in *2nd URSI Atlantic Radio Sci. Meeting*, May 2018.

[38] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of outdoor-to-indoor radio channels via propagation graphs," in *URSI General Assembly and Scientific Symposium*, Aug 2014, pp. 1–4.

[39] R. O. Adeogun, A. Bharti, and T. Pedersen, "An iterative transfer matrix computation method for propagation graphs in multi-room environments," *IEEE Antennas and Wireless Propag. Lett.*, vol. 18, no. 4, pp. 616–620, April 2019.

[40] R. Adeogun, T. Pedersen, and A. Bharti, "Transfer function computation for complex indoor channels using propagation graphs," *IEEE PIMRC*, Sept. 2018.

[41] L. Tian, X. Yin, Q. Zuo, J. Zhou, Z. Zhong, and S. X. Lu, "Channel modeling based on random propagation graphs for high speed railway scenarios," in *IEEE PIMRC*, Sept 2012, pp. 1746–1750.

[42] T. Zhou, C. Tao, S. Salous, Z. Tan, L. Liu, and L. Tian, "Graph-based stochastic model for high-speed railway cutting scenarios," *IET Microwaves, Antennas Propagation*, vol. 9, no. 15, pp. 1691–1697, 2015.

[43] J. Chen, X. Yin, L. Tian, and M. Kim, "Millimeter-wave channel modeling based on a unified propagation graph theory," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 246–249, Feb 2017.

[44] Y. Liu, X. Yin, X. Ye, Y. He, and J. Lee, "Embedded propagation graph model for reflection and scattering and its millimeter-wave measurement-based evaluation," *IEEE Open Journal of Antennas and Propagation*, pp. 1–1, 2021.

[45] A. Karttunen, C. Gustafson, A. F. Molisch, J. Jarvelainen, and K. Haneda, "Censored multipath component cross-polarization ratio modeling," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2016.

[46] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 055–30 062, 2020. [Online]. Available: https://www.pnas.org/content/117/48/30055

[47] L. Matyas, Ed., *Generalized Method of Moments Estimation.* Cambridge University Press, apr 1999.

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

[49] S. A. Sisson, *Handbook of Approximate Bayesian Computation.* Chapman and Hall/CRC, Sep 2018.

[50] J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander, "Fundamentals and recent developments in approximate bayesian computation," *Syst. Biol.*, vol. 66, pp. 66–82, Jan 2017.

[51] D. B. Rubin, "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *The Annals of Statistics*, vol. 12, no. 4, pp. 1151 – 1172, 1984. [Online]. Available: https://doi.org/10.1214/aos/1176346785

[52] S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly, "Inferring coalescence times from dna sequence data," *Genetics*, vol. 145, no. 2, pp. 505–518, 1997.

[53] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, "Population growth of human y chromosomes: a study of y chromosome microsatellites," *Molecular Biology and Evolution*, vol. 16, no. 12, pp. 1791–1798, dec 1999.

[54] A. Gretton, K. Borgwardt, M. J. Rasch, and B. Scholkopf, "A kernel two-sample test," *J. of Mach. Learn. Res.*, vol. 13, pp. 723–773, 2012.

[55] M. Park, W. Jitkrittum, and D. Sejdinovic, "K2-ABC: approximate Bayesian computation with kernel embeddings," *Proc. of the 19th Int. Conf. on Artif. Intell. and Statistics*, vol. 51, pp. 398–407, 2015. [Online]. Available: http://arxiv.org/abs/1502.02558

[56] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Adv. in Neural Inform. Process. Syst.*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf

[57] E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, "Approximate Bayesian computation with the Wasserstein distance," *J. R. Stat. Soc. Ser. B: Stat. Methodol.*, vol. 81, no. 2, pp. 235–269, 2019.

[58] M. G. B. Blum, M. A. Nunes, D. Prangle, and S. A. Sisson, "A comparative review of dimension reduction methods in approximate bayesian computation," *Stat. Sci.*, vol. 28, no. 2, pp. 189–208, may 2013.

[59] P. Fearnhead and D. Prangle, "Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 3, pp. 419–474, may 2012.

[60] M. G. B. Blum and O. François, "Non-linear regression models for approximate bayesian computation," *Statistics and Computing*, vol. 20, no. 1, pp. 63–73, mar 2009.

[61] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, "Markov chain monte carlo without likelihoods," *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15 324–15 328, 2003. [Online]. Available: https://www.pnas.org/content/100/26/15324

[62] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate bayesian computation," *Biometrika*, vol. 96, no. 4, pp. 983–990, Oct 2009.

[63] S. A. Sisson, Y. Fan, and M. M. Tanaka, "Sequential monte carlo without likelihoods," *Proceedings of the National Academy of Sciences*, vol. 104, no. 6, pp. 1760–1765, 2007. [Online]. Available: https://www.pnas.org/content/104/6/1760

[64] D. J. Hand, "Statistical inference: The minimum distance approach by ayanendranath basu, hiroyuki shioya, chanseok park," *International Statistical Review*, vol. 81, no. 1, pp. 162–162, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12011_14

[65] F.-X. Briol, A. Barp, A. B. Duncan, and M. Girolami, "Statistical inference for generative models with maximum mean discrepancy," *arXiv:1906.05944*, 2019.

[66] A. Hall, *Generalized method of moments*. Oxford New York: Oxford University Press, 2005.

[67] R. Gramacy, *Surrogates : Gaussian process modeling, design, and optimization for the applied sciences*. Boca Raton, FL: CRC Press, 2020.

[68] P. J. Diggle and R. J. Gratton, "Monte carlo methods of inference for implicit statistical models," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 46, no. 2, pp. 193–212, jan 1984.

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

[70] S. N. Wood, "Statistical inference for noisy nonlinear ecological dynamic systems," *Nature*, vol. 466, no. 7310, pp. 1102–1104, aug 2010.

[71] L. F. Price, C. C. Drovandi, A. Lee, and D. J. Nott, "Bayesian synthetic likelihood," *Journal of Computational and Graphical Statistics*, vol. 27, no. 1, pp. 1–11, jul 2017.

[72] K. L. Mengersen, P. Pudlo, and C. P. Robert, "Bayesian computation via empirical likelihood," *Proceedings of the National Academy of Sciences*, vol. 110, no. 4, pp. 1321–1326, jan 2013.

[73] M. Fasiolo, S. N. Wood, F. Hartig, and M. V. Bravington, "An extended empirical saddlepoint approximation for intractable likelihoods," *Electronic Journal of Statistics*, vol. 12, no. 1, pp. 1544–1578, 2018.

[74] O. Thomas, R. Dutta, J. Corander, S. Kaski, and M. U. Gutmann, "Likelihood-free inference by ratio estimation," *Bayesian Analysis*, sep 2020.

[75] A. Goldsmith, *Wireless Communications*. Cambridge University Press, Aug 2005.

[76] L. E. Franks, *Signal Theory*. Englewood Cliffs, N. J., Prentice-Hall, 1969.

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

[78] M. U. Gutmann and J. Corander, "Bayesian optimization for likelihood-free inference of simulator-based statistical models," *J. Mach. Learn. Res.*, vol. 17, no. 1, p. 4256–4302, Jan. 2016.

[79] I. Chevyrev and A. Kormilitzin, "A primer on the signature method in machine learning."

[80] I. Chevyrev and H. Oberhauser, "Signature moments to characterize laws of stochastic processes," *arXiv:1810.10971*, 2018.

Part II

Papers

Paper A

Parameter Estimation for Stochastic Channel Models using Temporal Moments

Ayush Bharti, Ramoni Adeogun, Troels Pedersen

The paper has been published in the
IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting, 2019.

Abstract

This paper proposes a method to infer on the parameters of a stochastic channel model from observations of temporal moments without multipath extraction. The distribution of the temporal moments is approximated to be Gaussian, and sampling is carried out from the approximate posterior. The temporal moments are found to be informative about the model parameters, as the parameters can be recovered from the samples.

1 Introduction

Parameters of stochastic multipath models, since the early works in [1] and [2], have predominantly been estimated by first extracting multipath parameters (delays, gains, etc.) and then estimating model parameters in a second step. Multipath extraction requires sophisticated algorithms which can be cumbersome to use and prone to errors [3]. In statistical terms, the multipath parameters are used as summary statistics for estimating model parameters. Other summaries, e.g. the well-known temporal moments of the received signal, can potentially be used, thereby avoiding multipath extraction altogether. Here, we propose a sampling method to estimate parameters of a stochastic multipath model based on temporal moments.

2 Signal Model

Ignoring additive noise, the received signal in a multipath channel can be written in complex baseband notation as

$$y(t) = \sum_l \alpha_l s(t - \tau_l), \tag{A.1}$$

where $s(t)$ is the transmitted signal, α_l and τ_l are the complex gain and time-delay of the l^{th} multipath component, respectively. The k^{th} temporal moment of $y(t)$ is defined as

$$m_k = \int t^k |y(t)|^2 \mathrm{dt}, \quad k = 0, 1, 2, \ldots \tag{A.2}$$

Under the large bandwidth approximation, $|s(t)|^2 \to \delta(t)$, and the temporal moment reads

$$m_k = \sum_l |\alpha_l|^2 \tau_l^k, \quad k = 0, 1, 2, \ldots \tag{A.3}$$

Here we consider a variant of Turin's model [1] where delays and gains form a homogeneous Poisson point process with arrival rate λ_0. The mark density $p(\alpha|\tau)$ is circular complex Gaussian with variance $\sigma_\alpha^2(\tau)$. For this model the

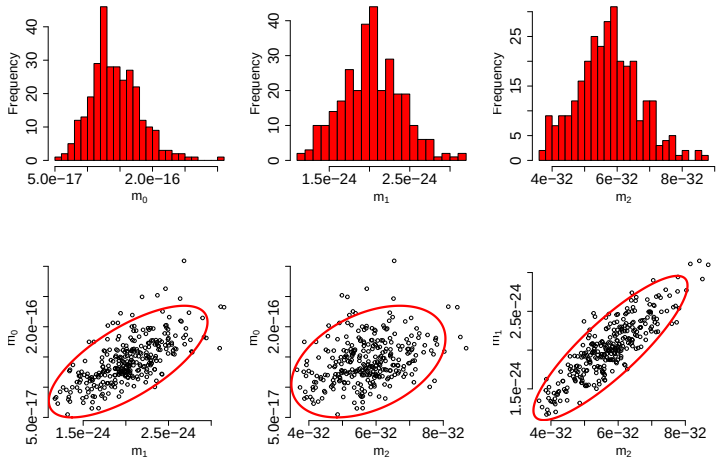


Fig. A.1: Scatter plots and histograms of synthetic data of $\mathbf{m}_0$, $\mathbf{m}_1$, and $\mathbf{m}_2$. Parameter settings: $g = 0.6$, $G_0 = 10^{-8}$, $\lambda_0 = 10^9$ s^{-1}, $N = 300$, $V = 36$ m^3, $S = 66$ m^2, $c = 3 \times 10^8$ ms^{-1}, and $\tau_{\max} = 200$ ns. Red ellipses are 95% probability contours of a Gaussian with parameters given by (A.5).

power delay spectrum reads $P(\tau) = \lambda_0 \sigma_\alpha^2(\tau)$, see [4]. For in-room scenarios, the power delay spectrum is well modelled by the reverberation model as [5]

$$P(\tau) = \begin{cases} G_0 \exp(-\frac{\tau}{T}), & \tau > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{A.4}$$

where G_0 is the reverberant power at delay zero, and $T = -4V/cS\ln(g)$ is the reverberation time, V is the volume, S is the surface area, c is the speed of light, and g is the reflection coefficient of the room. Fig. A.1 shows an example realisation drawn from the model.

The mean vector, μ, and the covariance matrix, Σ, of the first three temporal moments i.e. m_0, m_1, and m_2, can be found by invoking Campbell's theorem,

$$\mu = G_0 \begin{bmatrix} T \\ T^2 \\ 2T^3 \end{bmatrix}, \text{ and } \quad \Sigma = \frac{G_0^2}{\lambda_0} \begin{bmatrix} T & \frac{T^2}{2} & \frac{T^3}{2} \\ \frac{T^2}{2} & \frac{T^3}{2} & \frac{3T^4}{4} \\ \frac{T^3}{2} & \frac{3T^4}{4} & \frac{3T^5}{2} \end{bmatrix}. \tag{A.5}$$

3 Estimation Method

Let $\mathbf{m}_k$ be the N-dimensional vector of the k^{th} temporal moment, where $k = 0, 1, 2$. Samples from the posterior, $p(\Theta|\mathbf{m}_k)$, can be used to infer on the

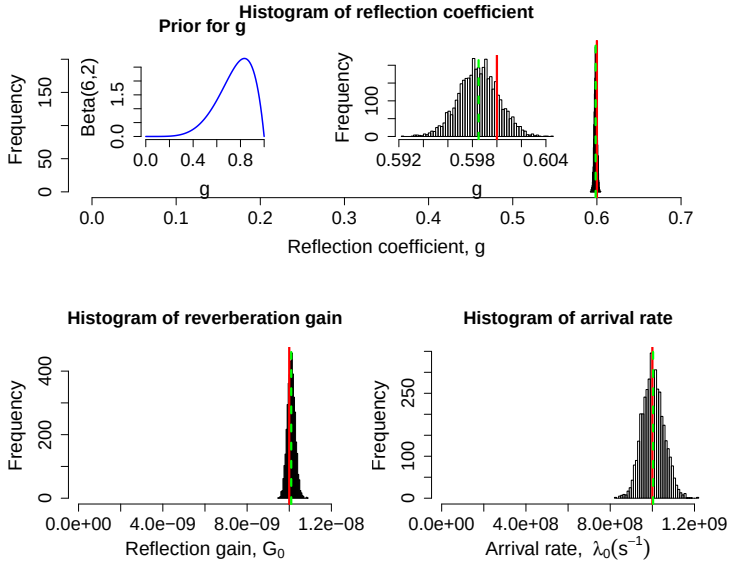


Fig. A.2: Histogram of the approximate posterior samples obtained from the data in Fig. A.1. A zoomed version of the histogram of g is inserted in the same plot. Red line: true value, dashed green line: MMSE estimate. Inset plot in blue represents the prior distribution, of g.

model parameters, $\Theta = [g, G_0, \lambda_0]^T$. Since the posterior and the likelihood, $p(\mathbf{m}_k|\Theta)$, are numerically unavailable, sampling is not possible. However, inspired by Fig. A.1, we approximate the likelihood as a Gaussian $\tilde{p}(\mathbf{m}_k|\Theta)$ with mean and covariance as in (A.5). Then we sample from the approximate posterior, $\tilde{p}(\Theta|\mathbf{m}_k) = \tilde{p}(\mathbf{m}_k|\Theta)p(\Theta)/p(\mathbf{m}_k)$, by using standard sampling techniques. Point estimates can then be obtained, e.g. averaging the posterior samples yield the minimum mean squared error (MMSE) estimate.

4 Simulation and Results

We run the default sampler in [6] on synthetic data from Fig. A.1 using a Beta prior for g and flat priors for G_0 and λ_0 to obtain 2000 samples from the approximate posterior. As shown in Fig. A.2, samples from all three posteriors are concentrated around their respective "true" values, resulting in very small estimation errors for MMSE estimator.

Root mean square errors (RMSE) of the estimator is computed using Monte Carlo experiment as follows. In each Monte Carlo run, N realisations of temporal moments are generated with the settings in Fig. A.1, and the MMSE

Table A.1: RMSE of the parameter estimates for different N with 500 Monte-Carlo runs each.

N	RMSE (RMSE/True Value)		
	$\hat{g}[10^{-3}]$	$\hat{G}_0[10^{-10}]$	$\hat{\lambda}_0$[MHz]
10	9 (1.5%)	10 (10%)	373 (37.3%)
50	4.6 (0.77%)	5.5 (5.5%)	140 (14%)
100	3.6 (0.6%)	4.3 (4.3%)	97 (9.7%)
300	2.4 (0.4%)	3.1 (3.1%)	56 (5.6%)

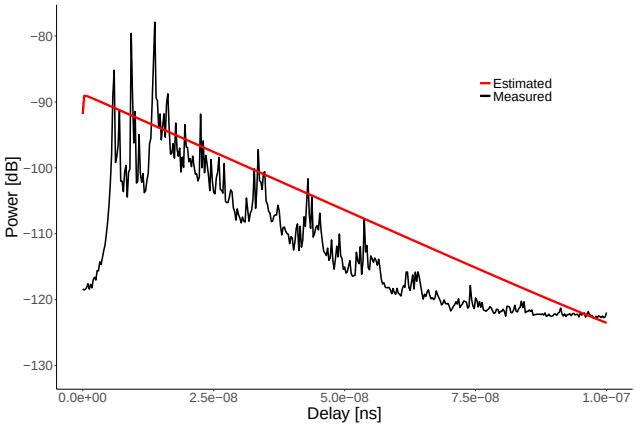


Fig. A.3: Measured averaged power delay profile (black) [3] for $N = 625$ and the power delay spectrum predicted from the parameter estimates (red).

estimate is computed. The RMSE, reported in Table A.1, decreases with increase in the size of the data. As expected from the relative widths of the posteriors in Fig. A.2, the RMSE of $\hat{g}$ is the smallest, followed by $\hat{G}_0$ and $\hat{\lambda}_0$. Note that even with a data size of $N = 10$, the RMSE is reasonably small.

To test the applicability of the estimator, we apply it to measured data from [3]. The estimated parameters are then used to predict the power delay spectrum. The result is compared with the averaged power delay profile from the measured data in Fig. A.3. Despite the employed assumptions of high bandwidth and no noise, the fit seems reasonable, indicating that the Gaussian approximation is sufficient.

5 Conclusion

We find that the parameters of the considered stochastic multipath model can be estimated by using temporal moments as summary statistics without the need for multipath extraction. Thus, temporal moments of the received signal are informative for estimating the parameters of the considered multipath model, i.e. arrival rate, reverberation gain and absorption coefficient. The proposed method is reasonably accurate despite the approximations involved (it ignores measurement noise, bandwidth limitations, and relies on a Gaussian approximation of the likelihood). Further work is needed to account for finite measurement bandwidth and noisy data.

Acknowledgments

This work is supported by the Danish Counsil for Independent Research, grant no. DFF – 7017-00265. This work was performed within the framework of the COST Action CA15104 IRACON.

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. 21, no. 1, pp. 1–9, Feb 1972.

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

[3] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

[4] T. Pedersen, "Modeling of path arrival rate for in-room radio channels with directive antennas," *IEEE Trans. on Antennas and Propag.*, vol. 66, no. 9, pp. 4791–4805, 2018.

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

[6] Stan Development Team, "RStan: the R interface to Stan, version 2.26," 2019.

Paper B

Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments

Ayush Bharti, Ramoni Adeogun, Troels Pedersen

The paper has been published in the
IEEE International Workshop on Signal Processing Advances in Wireless Communications, SPAWC, 2019.

Abstract

Stochastic channel models are usually calibrated after extracting the parameters of the multipath components from measurements. This paper proposes a method to infer on the underlying parameters of a stochastic multipath model, in particular the Turin model, without resolving the multipath components. Channel measurements are summarised into temporal moments instead of the multipath parameters. The parameters of the stochastic model are then estimated from the observations of temporal moments using a method of moments approach. The estimator is tested on real data obtained from in-room channel measurements. It is concluded that calibration of stochastic models can be done without multipath extraction, and that temporal moments are informative summary statistics about the model parameters.

1 Introduction

Realistic modelling of the radio channel is imperative to the design and analysis of any wireless communication system. Stochastic multipath models characterizing many different radio environments have been reported in the literature [1–4]. These models can be used to generate realizations of the channel in simulations, and to analyse the behaviour of communication systems. For a model to be useful for simulation, it should be calibrated, i.e. its parameters should be estimated such that the model fits to the measurement data.

Since the early works [1] and [2], parameters of a stochastic multipath channel model have usually been estimated in a two-step process as shown in Fig. B.1(a). First, the multipath parameters, for example the delays and their respective gains, are estimated from the channel measurements, followed by estimation of the model parameters, e.g. [4–7]. Resolving multipath components using high-resolution algorithms such as CLEAN [8], SAGE [9] or RIMAX [10] is not usually trivial, and the overall estimation accuracy of the model parameters relies on how accurately the multipath parameters are obtained. One particular problem which has been considered only recently [11] is the effect that some multipath components are undetected due to noise or bandwidth constraints. This censoring effect causes significant calibration errors.

The error introduced by this intermediate multipath extraction step can potentially be eliminated by methods that bypass this step and estimate the model parameters from other summaries, as depicted in Fig. B.1(b). Other summary statistics, apart from the multipath parameters, that can be found in the literature include power delay profile, root mean square delay spread, angular dispersion, temporal moments, and delay-Doppler function, among others [12]. Potentially these summaries are informative about the underlying model parameters and could thus be valuable for their estimation.

The choice of summary statistics is guided by a number of concerns. First, summary statistics should be informative about the model parameters. Ideally,

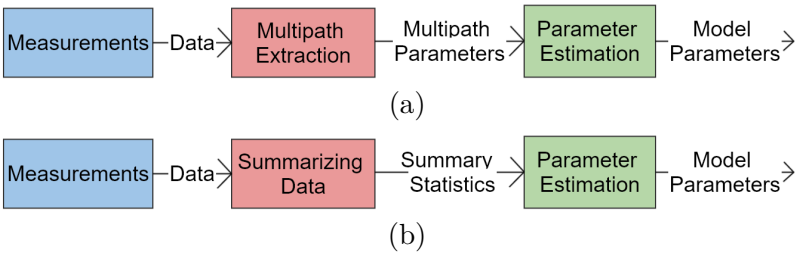


Fig. B.1: Calibration procedures involving (a) multipath extraction, and (b) general summaries.

a sufficient statistic should be chosen, but in most practical situations, this is not feasible or possible. In fact, the multipath parameters form a sufficient statistic for the underlying point process, provided that all multipath components are actually resolved without error, which is unlikely the case. Moreover, the summary should be easily computable. One such choice of summary statistics, that does not introduce significant computational overhead, are the temporal moments. These have been widely used in wireless communication literature since the 1970s, see e.g [13], but have, to the best of our knowledge hitherto not been considered for calibration of multipath models apart from the sampling approach proposed by the authors in [14].

In this contribution, we propose a method of moments parameter estimator for the Turin model [1] without multipath extraction. This model has recently attracted attention due to its simplicity, see e.g. [4, 20]. We compute the temporal moments from the transfer function measurements, without transforming it to time domain. Expressions for the means and covariances of the temporal moments are derived. Inserting sample means and covariances from the measurements to these expressions yield the parameter estimates. The performance of the proposed estimator is evaluated by a simulation experiment and validated using measurement data.

2 Signal Model

Consider a single-input, single-output (SISO) multipath propagation scenario where the received signal is measured using a transmit and a receive antenna at a certain frequency bandwidth, B. The received signal, Y_k, is modelled in the frequency domain as

$$Y_k = H_k + N_k, \quad k = 0, 1, \ldots, (N_s - 1), \tag{B.1}$$

where k is the frequency index, H_k is the transfer function, N_k is the noise contribution, and N_s is the total number of sample points. The major source of noise in such measurements is noise from the measurement equipment itself, which can be assumed independent and identically distributed (iid) at each measurement point. Here the noise is modelled as iid complex Gaussian variables, $N_k \sim \mathcal{CN}(0, \sigma_N^2)$, $k = 0, 1, \ldots, N_s - 1$.

The temporal moments can be computed in the frequency domain as follows. Discrete-frequency, continuous-time inverse Fourier transform of Y_k gives the signal in the time domain, $y(t)$

$$y(t) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} Y_k \exp\left(j2\pi k\Delta f t\right), \tag{B.2}$$

where Δf is the frequency separation between two samples. The time domain signal is periodic with period

$$t_{\max} = \frac{1}{\Delta f} = \frac{(N_s - 1)}{B}. \tag{B.3}$$

Define the i^{th} temporal moment of $y(t)$ as

$$m_i = \int_0^{t_{\max}} t^i |y(t)|^2 \mathrm{dt}, \quad i = 0, 1, 2, \ldots \tag{B.4}$$

Consequently, in the frequency-domain, we have

$$m_i = \frac{1}{N_s^2} \sum_k \sum_{k'} Y_k Y_{k'}^* a_i(k - k'), \tag{B.5}$$

with the definition

$$a_i(k - k') = \int_0^{t_{\max}} t^i \exp\left(j2\pi\Delta f(k - k')t\right) \mathrm{dt}. \tag{B.6}$$

A number of properties can be shown for this function. Since $m_i \geq 0$ for all Y_k, $a_i(k - k')$ is a positive semidefinite function. Also, $a_i(k) = a_i^*(-k)$. Note that $a_i(0) = t_{\max}^{i+1} / (i + 1)$. The magnitude of $a_i(k - k')$ decreases rapidly as $|k - k'|$ increases. In particular, $a_0(k - k') = 0$ for $k \neq k'$. Therefore, the diagonal terms in (B.5) are the most relevant for computing the temporal moments.

The frequency domain method in (B.5) is preferred over (B.4) here since we work with frequency domain measurements and it leads to low complexity of the estimator derived in Section 3. It should be noted that no attempt has been made to remove noise and effects of finite measurement bandwidth. Thus, we avoid the employment of heuristics to set the noise floor and truncate the time domain signal at an arbitrary point. This is advantageous here, since the derived estimator and its performance does not depend on such arbitrary choices, and the results are more easily reproduced.

The temporal moments $m_i, i = 0, 1, 2, \ldots$, are random variables with means

$$\mathbb{E}[m_i] = \frac{1}{N_s^2} \sum_k \sum_{k'} \mathbb{E}\left[Y_k Y_{k'}^*\right] a_i(k - k'), \tag{B.7}$$

and covariances

$$\mathrm{Cov}(m_i, m_j) = \frac{1}{N_s^4} \sum_{k,k',n,n'} \mathrm{Cov}(Y_k Y_{k'}^*, Y_n Y_{n'}^*) a_i(k-k') a_j(n-n'). \quad \text{(B.8)}$$

It should be noted that the first and the second moment of the temporal moments depend upon the second and fourth moment of the received signal, respectively.

Employing the uncorrelated scattering assumption [15], the second moment of Y_k can be written as

$$\mathbb{E}\left[Y_k Y_{k'}^*\right] = \int P_y(t) \exp\left(-j2\pi\Delta f t(k-k')\right) \mathrm{dt} + \sigma_N^2 \delta(k-k'), \quad \text{(B.9)}$$

where $\delta(\cdot)$ is the Kronecker delta function, and $P_y(t)$ is the power delay spectrum of the received signal[1]. The power delay spectrum reads from (B.2) as

$$P_y(t) = \frac{1}{N_s^2} \sum_k \sum_{k'} \mathbb{E}\left[Y_k Y_{k'}^*\right] \exp\left(j2\pi\Delta f(k-k')t\right). \quad \text{(B.10)}$$

Considering high bandwidths, $P_y(t)$ can be approximated as

$$P_y(t) \approx E_s P_h(t) + \sigma_N^2 / N_s, \quad \text{(B.11)}$$

where E_s is the energy of the transmitted signal, and $P_h(t)$ is the power delay spectrum of the channel. With the transmitted signal in frequency domain being the rectangular window of unit magnitude over the bandwidth, that gives $E_s = B$.

3 Method of Moments Estimator for Turin's Model

3.1 Channel model description

As an example of how the temporal moments can be utilized to calibrate stochastic channel models, we derive a method of moments estimator for the

[1] With $s(t)$ as the transmitted signal, $P_y(t)$ is defined as in the noise free case:

$$P_y(t) = \mathbb{E}[|y(t)|^2] = \int P_h(\tau)|s(t-\tau)|^2 \mathrm{d}\tau,$$

where $P_h(t)$ may be informally interpreted as $P_h(t) = \mathbb{E}[|h(t)|^2]$, where $|h(t)|^2$ is the (instantaneous) power delay profile of the channel. [16]

seminal model by Turin [1], applied to an in-room setting. For a multipath channel, the transfer function, H_k, reads

$$H_k = \sum_l \alpha_l \exp\left(-j2\pi\Delta f k \tau_l\right), \tag{B.12}$$

where α_l and τ_l are the complex gain and time-delay of the l^{th} multipath component, respectively. The delays form a homogeneous Poisson point process with arrival rate $\lambda(t) = \lambda_0$. The gains, conditioned on the delays, are modelled as independent circular complex Gaussian variables with variance $\sigma_\alpha^2(t)$. Therefore, $\{(\tau_l, \alpha_l)\}$ forms a marked Poisson point process of intensity λ_0, with points $\{\tau_l\}$ and marks $\{\alpha_l\}$. For this type of model, $P_h(t) = \lambda(t)\sigma_\alpha^2(t)$, see [16].

Typically for in-room channel measurements, the power delay spectrum has an exponentially decaying behaviour, and can be approximated using the reverberation model in [17] as

$$P_h(t) = \begin{cases} G_0 \exp(-\frac{t}{T}), & t > t_0 \\ 0 & \text{otherwise,} \end{cases} \tag{B.13}$$

where G_0 is the power at delay zero called reverberation gain, T is the reverberation time, and t_0 is the delay of the first multipath component. The arrival rate, however, does not enter in (B.13). Substituting this model in (B.9) and carrying out the integration gives the autocorrelation function of Y as

$$\begin{aligned} R_Y(k,k') &= \mathbb{E}\left[Y_k Y_{k'}^*\right] \\ &= \frac{G_0 T B e^{-t_0\left(\frac{1}{T} + j2\pi\Delta f(k-k')\right)}}{1 + j2\pi\Delta f T(k-k')} + \sigma_N^2 \delta(k-k'). \end{aligned} \tag{B.14}$$

As to be expected, Y is a wide-sense stationary process in the frequency domain.

3.2 Estimator derivation

We follow a method of moments approach to estimate the four parameters, G_0, T, σ_N^2, and λ_0, from N observations of the summary statistic consisting of three temporal moments (m_0, m_1, m_2). Out of the four minimum equations required to solve for the parameters, three of them are taken to be the equations for the mean. Using (B.14) in (B.7) gives the three mean equations as:

$$\mu_0 = \frac{G_0 T B}{N_s^2}\beta_0(T) + \frac{\sigma_N^2}{B}, \tag{B.15}$$

$$\mu_1 = \frac{G_0 T B}{N_s^2}\beta_1(T) + \frac{\sigma_N^2 t_{\max}^2}{2N_s}, \tag{B.16}$$

$$\mu_2 = \frac{G_0 T B}{N_s^2}\beta_2(T) + \frac{\sigma_N^2 t_{\max}^3}{3N_s}, \tag{B.17}$$

where $\mu_i = \mathbb{E}[m_i]$, and $\beta_i(T)$ is defined as

$$\beta_i(T) = \sum_{\tilde{k}=-N_s+1}^{N_s-1} \frac{(N_s - |\tilde{k}|)e^{-t_0\left(\frac{1}{T}+j2\pi\Delta f\tilde{k}\right)}a_i(\tilde{k})}{1+j2\pi\Delta fT\tilde{k}}, \tag{B.18}$$

with $\tilde{k} = k - k'$ and $i = 0, 1, 2$. Substituting σ_N^2 from (B.15) into (B.16) and (B.17), and then dividing the two equations gives

$$\left(\hat{\mu}_1 - \frac{\hat{\mu}_0}{2\Delta f}\right)\left(\frac{B}{N_s^2}\beta_2(T) - \frac{e^{-\frac{t_0}{T}}}{3\Delta f^2}\right) - \left(\hat{\mu}_2 - \frac{\hat{\mu}_0}{3\Delta f^2}\right)\left(\frac{B}{N_s^2}\beta_1(T) - \frac{e^{-\frac{t_0}{T}}}{2\Delta f}\right) = 0. \tag{B.19}$$

Here, $\hat{\mu}_i$ are the estimates of μ_i for $i = \{0, 1, 2\}$, found by taking the sample mean of the temporal moments. Solving (B.19) numerically for T (which is easily done using standard numerical solvers) gives the estimate for the reverberation time, $\hat{T}$. The estimate for the reverberation gain, $\hat{G}_0$, is then

$$\hat{G}_0 = \frac{N_s^2\left(2\Delta f\hat{\mu}_1 - \hat{\mu}_0\right)}{2\Delta fB\beta_1(\hat{T}) - N_s^2 e^{-\frac{t_0}{\hat{T}}}}. \tag{B.20}$$

The noise variance is then estimated by inserting $\hat{G}_0$, $\hat{T}$ and $\hat{\mu}_0$ in (B.15).

The arrival rate, λ_0, does not appear in the mean equations (B.15)-(B.17), but in the covariances derived in Appendix. Any of the covariance equations can be used. The simplest is the equation for the variance of m_0:

$$\mathrm{var}(m_0) = \frac{G_0^2 T e^{-\frac{2t_0}{T}}}{\lambda_0} + \underbrace{\frac{t_{\max}^2}{N_s^4}\sum_{p=-N_s+1}^{N_s-1}(N_s - |p|)\,|R_Y(p)|^2}_{\gamma}, \tag{B.21}$$

where $R_Y(p)$ is the autocorrelation function of Y at lag p. Estimating γ requires the estimation of $R_Y(p)$, which is done by using $\hat{G}_0$, $\hat{T}$, and $\hat{\sigma}_N^2$ in (B.14). The arrival rate is then estimated as

$$\hat{\lambda}_0 = \frac{\hat{G}_0^2\hat{T}e^{-\frac{2t_0}{\hat{T}}}}{\widehat{\mathrm{var}}(m_0) - \hat{\gamma}}, \tag{B.22}$$

where $\widehat{\mathrm{var}}(m_0)$ is the sample variance of the zeroth temporal moment and $\hat{\gamma}$ is the estimate of γ.

3.3 Estimation procedure

Let the measurements of a SISO channel be stored in a matrix $\mathbf{Y} \in \mathbb{C}^{N\times N_s}$, where N is the number of realizations of the received signal. The first three

Algorithm 2 Method of moments estimator

Input: $\mathbf{Y}$, t_0

1: Compute $\mathbf{D}$ from $\mathbf{Y}$ using (B.5)
2: Compute the sample means of $\{\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2\}$, and the sample variance of $\mathbf{m}_0$
3: Solve (B.19) numerically to find $\hat{T}$
4: Use $\hat{T}$ in (B.20) to find $\hat{G}_0$
5: Obtain $\hat{\sigma}_N^2$ using $\hat{G}_0$ and $\hat{T}$ in (B.15)
6: Estimate $R_Y(p)$ from (B.14) using $\hat{G}_0$, $\hat{T}$, $\hat{\sigma}_N^2$, and then compute $\hat{\gamma}$
7: Obtain $\hat{\lambda}_0$ from (B.22) using $\hat{G}_0$, $\hat{T}$ and $\hat{\gamma}$

Output: $\hat{T}$, $\hat{G}_0$, $\hat{\sigma}_N^2$, $\hat{\lambda}_0$

temporal moments can then be computed for each realization, resulting in a data matrix $\mathbf{D} = [\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2]$, where $\mathbf{m}_i \in \mathbb{R}^N$ for $i = 0, 1, 2$. The algorithm for estimating the parameters using method of moments is given in Alg. 2. Note that t_0 is considered as an input to the estimator. This could either be obtained as side information from the measurement set-up or estimated by finding the first peak in the data. No further assumptions, such as the number of multipath components to extract, are required here.

4 Performance Evaluation

4.1 Numerical experiment

To evaluate the accuracy of the estimator, we perform Monte Carlo simulations for different values of SNR, defined as SNR $= G_0 TB/\sigma_N^2$, and number of realizations, N, using synthetic data generated from the model. The root-mean-squared error (RMSE) of the different parameter estimates are normalized by their "true" values and shown in Fig. B.2.

It can be seen from Fig. B.2 that the RMSE of the estimates decrease as the number of realizations (and hence the number of data points of each temporal moment) increase. This illustrates that the method of moments estimate is consistent. The normalized RMSE is fairly small in all four cases with the arrival rate estimator showing the largest error. This is expected since $\hat{\lambda}$ is obtained as a combination of a variance estimate and the three other estimates. The estimation accuracy also improves with SNR. However, the improvement is less pronounced for reverberation gain due to the high bandwidth. Further analysis of the simulation data (not shown here) shows that the parameter estimates are unbiased.

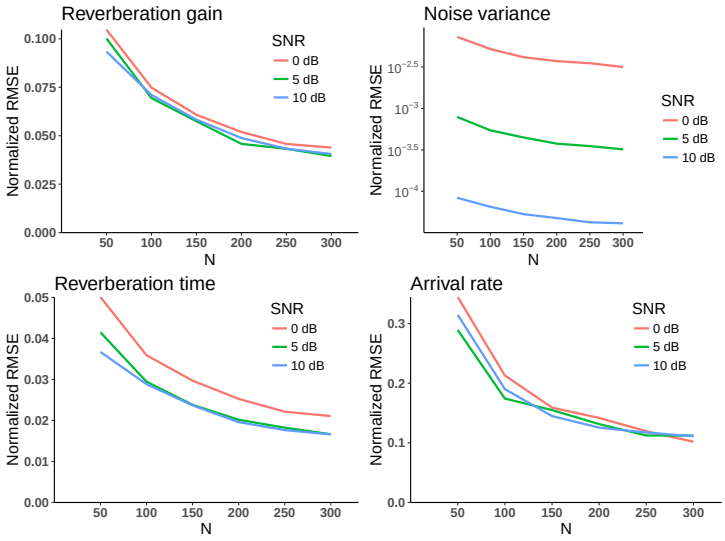


Fig. B.2: Normalized RMSE of the different parameter estimates as a function of number of channel realizations, plotted for different SNR values. Parameter settings: $G_0 = 10^{-8}$, $T = 10$ ns, $\lambda_0 = 1$ ns^{-1}, $t_0 = 5$ ns, $B = 4$ GHz, $N_s = 801$, Number of Monte Carlo runs = 500. Note that the RMSE of noise variance is in log-scale.

4.2 Application to measured data

To test the estimator's applicability, we now apply it to a set of measurement data described in [11]. The dataset consists of channel transfer functions obtained using a vector network analyser (VNA) in a room of dimensions $3\times4\times3$ m^3. The set-up is SISO, with a virtual planar array of 25×25, resulting in $N = 625$ realizations of the channel. The bandwidth of the measured signal is 4 GHz, with $N_s = 801$ samples in each channel measurement, resulting in $\Delta f = 5$ MHz and $t_{\max} = 200$ ns. The delay of the first peak is $t_0 = 6$ ns, found through visual inspection of the data. The estimator was not observed to be sensitive to variations in t_0 of the order of $1/B$. For this dataset, our implementation of Alg. 2, programmed in R version 3.4.3, took around 8 s when run on a notebook with a dual-core Intel i7 processor and 24 GB RAM.

To further demonstrate the soundness of the proposed method, we now use the obtained estimates to estimate the power delay spectrum using (B.10). This is then compared with the averaged power delay profile (PDP) of the measurement data as shown in Fig. B.3. The averaged PDP of simulated signals, i.e. the simulated power delay spectrum from the model using the estimates, is also shown. Both the theoretical and simulated power delay spectrum obtained from the estimates seem to fit the averaged PDP of the measurements, although there is a slight discrepancy between the estimated noise floor and the noise floor observed in the measurements. This demonstrates that the estima-

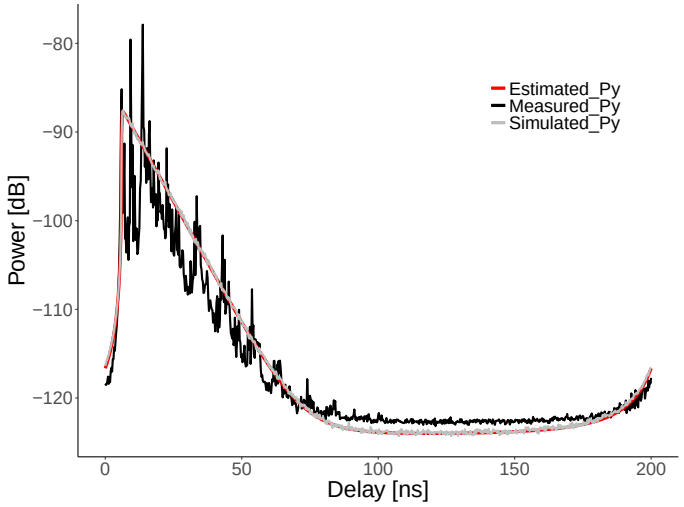


Fig. B.3: Figure showing the averaged PDP of $N = 625$ realizations of the measurement data (black) and the simulated data (grey) versus the delay. The estimated $P_y(t)$ computed from (B.10) is shown in red. The estimates obtained are: $\hat{G}_0 = -83.9$ dB, $\hat{T} = 7.8$ ns, $\hat{\sigma}_N^2 = 2.8 \times 10^{-10}$, and $\hat{\lambda}_0 = 10.0$ ns^{-1}. The estimated noise floor is $\hat{\sigma}_N^2/N_s = -124.5$ dB.

tor is able to extract the model parameters accurately from real measurements without the need for multipath extraction.

5 Conclusion

The proposed method of moments estimator can be used to calibrate the Turin model with a constant arrival rate without multipath extraction. The performance evaluation shows that the parameters of the considered stochastic multipath channel model can be estimated with good accuracy. The temporal moments, used in the estimation procedure, are easy-to-compute summary statistics of the measurement data. Moreover, they are informative about the parameters of the considered model, i.e. reverberation gain, reverberation time, arrival rate and noise variance. This estimation procedure bypasses the need to truncate the measured impulse response via some heuristics to estimate the decay of the power delay spectrum.

Comparison with performance bounds, such as the Cramer-Rao bound, is hindered by the lack of a likelihood function. Furthermore, comparison with calibration procedures involving multipath extraction with a number of heuristic choices is non-trivial. These considerations are left for future work.

Appendix

The characteristic function for the received signal evaluated at an arbitrary point $\mathbf{v}$ is defined as

$$\mathcal{C}(\mathbf{v}) = \mathbb{E}\left[\exp\left(j\Re\mathbf{v}^H\mathbf{Y}\right)\right], \tag{B.23}$$

where $\Re$ denotes the real part. The joint moment of four frequencies are obtained by letting $\mathbf{v} = [\nu_1, \nu_2, \nu_3, \nu_4]^T$ and $\mathbf{Y} = [Y_{k_1}, Y_{k_2}, Y_{k_3}, Y_{k_4}]^T$. Since marked point process $\{(\tau_l, \alpha_l)\}$ forms a two-dimensional Poisson point process with rate $p(\alpha|\tau)\lambda(\tau)$ [18], we obtain by Campbell's theorem the cumulant generating function (log characteristic function) as

$$K(\mathbf{v}) = \int \left[\mathcal{C}_{\alpha|\tau}\left(\sum_{j=1}^{4}\nu_j \exp\left(-j2\pi\Delta f k_j\tau\right)\right) - 1\right]\lambda(\tau)\mathrm{d}\tau, \tag{B.24}$$

where $\mathcal{C}_{\alpha|\tau}(\cdot)$ is the characteristic function for the circular symmetric complex Gaussian $p(\alpha|\tau)$ [20].

The covariance of the temporal moments is related to the fourth moment of the received signal, Y_k, which in turn can be written in terms of the fourth cumulant of the signal [19, Eq. (2.118)] as

$$\mathrm{Cov}(Y_kY_{k'}^*, Y_nY_{n'}^*) = \mathrm{Cum}(Y_kY_{k'}^*Y_nY_{n'}^*) + \mathbb{E}[Y_kY_{n'}^*]\mathbb{E}[Y_nY_{k'}^*]. \tag{B.25}$$

The fourth cumulant of Y is found by complex differentiation of the cumulant generating function [19]

$$\begin{aligned}\mathrm{Cum}(Y_kY_{k'}^*Y_nY_{n'}^*) &= \frac{2^4}{j^4}\cdot\frac{\partial^4 K(\mathbf{v})}{\partial\nu_1\partial\nu_2^*\partial\nu_3\partial\nu_4^*}\bigg|_{\nu_1,\nu_2,\nu_3,\nu_4=0}\\ &= 2\int \sigma_\alpha^4(t)\lambda(t)e^{-j2\pi\Delta f(k_1-k_2+k_3-k_4)t}\mathrm{d}t.\end{aligned} \tag{B.26}$$

With $\lambda(t) = \lambda_0\mathbb{1}(t > t_0)$, we get

$$\begin{aligned}\mathrm{Cum}(Y_kY_{k'}^*Y_nY_{n'}^*) &= \frac{2G_0^2B^2}{\lambda_0}\int_{t_0}^{\infty} e^{-\frac{2t}{T}}e^{-j2\pi\Delta f(k-k'+n-n')t}\mathrm{d}t\\ &= \frac{2G_0^2TB^2\psi_{(k-k'+n-n')}}{\lambda_0\left[2 + j2\pi\Delta f(k-k'+n-n')T\right]},\end{aligned} \tag{B.27}$$

where $\psi_{(k-k'+n-n')} = e^{-t_0\left(2/T + j2\pi\Delta f(k-k'+n-n')\right)}$. Now, let $\mathrm{Cov}(m_i, m_j) = \varrho_1^{ij} + \varrho_2^{ij}$. Then ϱ_1^{ij}, being the quadruple sum over the cumulant expression, can be expressed in the form of a double sum as

$$\varrho_1^{ij} = \frac{G_0^2TB^2}{\lambda_0 N_s^4}\sum_{\tilde{k},\tilde{n}=-N_s+1}^{N_s-1}\frac{\left(N_s - |\tilde{k}|\right)\left(N_s - |\tilde{n}|\right)\psi_{(\tilde{k}+\tilde{n})}a_i(\tilde{k})a_j(\tilde{n})}{1 + j\pi\Delta f(\tilde{k}+\tilde{n})T}. \tag{B.28}$$

Since $\mathbb{E}[Y_k Y_{n'}^*] = R_Y(k-n')$, ϱ_2^{ij} can be written as

$$\begin{aligned}\varrho_2^{ij} &= \frac{1}{N_s^4} \sum_{k,k',n,n'} \mathbb{E}\left[Y_k Y_{n'}^*\right] \mathbb{E}\left[Y_n Y_{k'}^*\right] a_i(k-k') a_j(n-n') \\ &= \frac{1}{N_s^4} \sum_{k,k',n,n'} R_Y(k-n') R_Y(n-k') a_i(k-k') a_j(n-n') \\ &= \frac{1}{N_s^4} \sum_{p=-N_s+1}^{N_s-1} (N_s - |p|) \left(a_j * R_Y\right)[p] \left(a_i * R_Y\right)[-p], \quad \text{(B.29)}\end{aligned}$$

where the convolution defined as

$$(f * g)[n] := \sum_l f[l] g^*[n-l].$$

Noticing that $a_0(k-k') = t_{\max}\delta(k-k')$, $\mathrm{var}(m_0)$ is given as in (B.21).

Acknowledgements

The authors would like to thank Dr. Carl Gustafson and Prof. Fredrik Tufvesson (Lund University) for providing the measurement data. This work is supported by: (1) VIRTUOSO, funded by Intel Mobile Communications, Keysight, Telenor, Aalborg University, and Denmark Innovation Foundation; (2) the Danish Counsil for Independent Research, grant no. DFF 7017-00265 and performed within the framework of the COST Action CA15104 IRACON.

References

[1] George L. Turin, Fred D. Clapp, Tom L. Johnston, Stephen B. Fine, Dan Lavry, "A statistical model of urban multipath propagation," in IEEE Trans. Veh. Technol., vol. 21, February 1972.

[2] Adel A. M. Saleh, Reinaldo A. Valenzuela, "A statistical model for indoor multipath propagation," in IEEE J. Sel. Areas Commun., vol. 5, pp. 128-137, February 1987.

[3] Quentin H. Spencer, Brian D. Jeffs, Michael A. Jensen, A. Lee Swindlehurst, "Modelling the statistical time and angle of arrival characteristics of an indoor multipath channel," in IEEE J. Sel. Areas Commun., vol. 18, no. 3, pp. 347-360, March 2000.

[4] Katsuyuki Haneda, Jan Järveläinen, Aki Karttunen, Mikko Kyrö, Jyri Putkonen, "A statistical spatio-temporal radio channel model for large indoor environments at 60 and 70 GHz," in IEEE Trans. Antennas Propag., vol. 63, no. 6, pp. 2694-2704, June 2015.

[5] P. Kyösti et al., "WINNER II channel models, deliverables D1.1.2 V1.2, part I: Channel models," Tech. Rep. IST-4-027756 WINNER II Project, 2008.

[6] Juho Poutanen, Katsuyuki Haneda, Lingfeng Liu, Claude Oestges, Fredrik Tufvesson, Pertti Vainikainen, "Parameterization of the COST 2100 MIMO channel model in indoor scenarios," in EuCAP, 2011.

[7] Carl Gustafson, Katsuyuki Haneda, Shurjeel Wyne, Fredrik Tufvesson, "On mm-wave multipath clustering and channel modeling," in IEEE Trans. Antennas Propag., vol. 62, no. 3, pp. 1445-1455, March 2014.

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

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

[10] A. Richter, M. Landmann, R. Thomä, "RIMAX-A maximum likelihood framework for parameter estimation in multidimensional channel sounding," in Proc. ISAP, Sendai, Japan, pp. 53-56, 2004.

[11] Carl Gustafson, David Bolin, Fredrik Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," IEEE GLOBECOM, 2016.

[12] Sana Salous, "Radio Propagation Measurement and Channel Modelling." Wiley, April 2013.

[13] Donald C. Cox, " Delay Doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment," in IEEE Trans. Antennas Propag., vol. 20, no. 5, pp. 625-635, September 1972.

[14] A. Bharti, R. Adeogun, 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.

[15] P. A. Bello, "Characterization of randomly time-variant linear channels," in IEEE Trans. Commun., vol. 11, pp. 360-396, 1963.

[16] Troels Pedersen, "Modelling of path arrival rate for in-room radio channels with directive antennas," in IEEE Trans. Antennas Propag., vol. 66, no. 9, pp. 4791-4805, September 2018.

[17] G. Steinböck, T. Pedersen, B. Fleury, W. Wang, R. Raulefs, "Distance dependent model for the delay power spectrum of in-room reverberant channels," in IEEE Trans. Commun., 2013.

[18] J. F. C. Kingman, “Poisson Processes,” Chapter 3, 1993.

[19] P. J. Schreier, L. L. Scharf, “Statistical Signal Processing of Complex-Valued Data.” Cambridge University Press, 2010.

[20] Troels 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, 2019.

Paper C

Maximum Likelihood Calibration of Stochastic Multipath Radio Channel Models

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

The paper has been published in the
IEEE Transactions on Antennas and Propagation, 2020.

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.

1 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. C.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 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 algorithms is typically non-trivial,

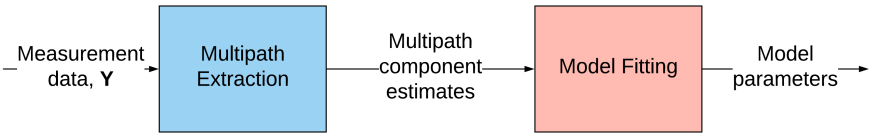


Fig. C.1: Calibration procedure usually followed where estimates of multipath components are used as summaries.

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 2 introduces the inhomogeneous Turin model as the stochastic multipath model studied in this paper. Next, in Section 3, 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 4 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 5, where we analyze a real dataset of indoor channel measurements originally considered in [25]. Finally, Section 6 concludes the paper with a discussion and indications to avenues of future research.

2 Stochastic Multipath Model

2.1 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

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

where H_k is the transfer function sampled at the k^{th} 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)

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

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

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

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].

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

2.2 Stochastic multipath model

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

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

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]

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

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,

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

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 (C.5) and (C.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]

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

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 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 (C.7). For computational convenience, we reparametrize the model as

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

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 (C.5), (C.6) and (C.8), we set the second conditional moment of the magnitude as

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

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 3.3.

2.3 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

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

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

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

[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 $\lambda(t) = c \exp(\kappa_1 t)$.

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

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

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

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

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,

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

The likelihood function (C.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 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 (C.7) this situation is very relevant in our study.

3 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 Sec. 3.1, be evaluated using conditional samples of Z_{m} given $\mathbf{y}$. Furthermore, these samples can be generated using an MCMC algorithm detailed in Section 3.2. Finally, the approximated likelihood ratio is maximized with respect to θ as discussed in Sec. 3.3. We comment in Sec. 3.4 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 (C.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].

3.1 Monte Carlo approximations of likelihood

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

$$\frac{L(\theta;\mathbf{y})}{L(\theta_0;\mathbf{y})} = \mathbb{E}_{Z_{\mathsf{m}}|\mathbf{y};\theta_0}\Big[\frac{L(\theta;\mathbf{y},Z_{\mathsf{m}})}{L(\theta_0;\mathbf{y},Z_{\mathsf{m}})}\Big], \tag{C.12}$$

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 (C.10). The right-hand side of (C.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 (C.12) can be approximated by an empirical average based on samples[4] $Z_{\mathsf{m},1},\ldots,Z_{\mathsf{m},N}$ from the

[3] Replacing the right-hand side of (C.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 3.2.

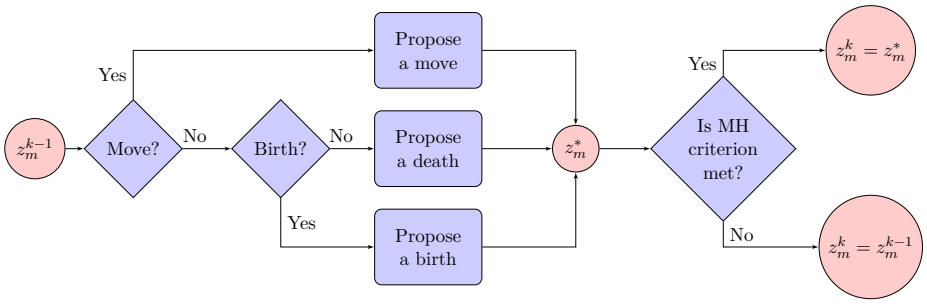


Fig. C.2: Diagram of one iteration of the birth-death MCMC algorithm.

conditional distribution of Z_{m} given $\mathbf{y}$,

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

with the notation

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

A delicate issue of the estimator (C.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 (C.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]

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

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 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.

3.2 Birth/death MCMC sampling with parallel tempering

Algorithm 3 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` $< p_{\text{move}}$ **then**
    Pick (τ, α_τ) uniformly at random from z'_{m}
    if `rand` $< p_{\text{delay}}$ **then**
      $\tau \leftarrow \tau + \sigma_{\text{delay}} \cdot$ `randn`
      $\text{mhr} \leftarrow L(\theta; \mathbf{y}, z'_{\mathsf{m}}) / L(\theta; \mathbf{y}, z_{\mathsf{m}}^{(k-1)})$
    else
      if `rand` $< p_{\text{phase}}$ **then**
        $\text{phase}(\alpha_\tau) \leftarrow \text{phase}(\alpha_\tau) + \sigma_{\text{phase}} \cdot$ `randn`
        $\text{mhr} \leftarrow L(\theta; \mathbf{y}, z'_{\mathsf{m}}) / L(\theta; \mathbf{y}, z_{\mathsf{m}}^{(k-1)})$
      else
        $v :=$ `randn`
        $\text{magn}(\alpha_\tau) \leftarrow \text{magn}(\alpha_\tau) \cdot \exp(\sigma_{\text{magnitude}} \cdot v)$
        $\text{mhr} \leftarrow \exp(v) L(\theta; \mathbf{y}, z'_{\mathsf{m}}) / L(\theta; \mathbf{y}, z_{\mathsf{m}}^{(k-1)})$
      end if
    end if
  else
    if `rand` $< p_{\text{birth}}$ **then**
      Add marked point (τ, α_τ) to z'_{m} with τ uniform in I and $\alpha_\tau \sim f$
      $\text{mhr} \leftarrow \frac{L(\theta;\mathbf{y},z'_{\mathsf{m}})(1-p_{\text{birth}})|I|}{L(\theta;\mathbf{y},z_{\mathsf{m}}^{(k-1)})p_{\text{birth}} f(\alpha_\tau)(n(z'_{\mathsf{m}})+1)}$
    else
      Pick (τ, α_τ) uniformly at random from z'_{m} and delete it from z'_{m}.
      $\text{mhr} \leftarrow \frac{L(\theta;\mathbf{y},z'_{\mathsf{m}})p_{\text{birth}} f(\alpha_\tau) n(z_{\mathsf{m}}^{(k-1)})}{L(\theta;\mathbf{y},z_{\mathsf{m}}^{(k-1)})(1-p_{\text{birth}})|I|}$
    end if
  end if

  if `rand` $< \min\{1, \text{mhr}\}$ **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

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 3 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. C.2.

The MCMC updates of Algorithm 3 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 described in Section 2. 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

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

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 $i < K$.

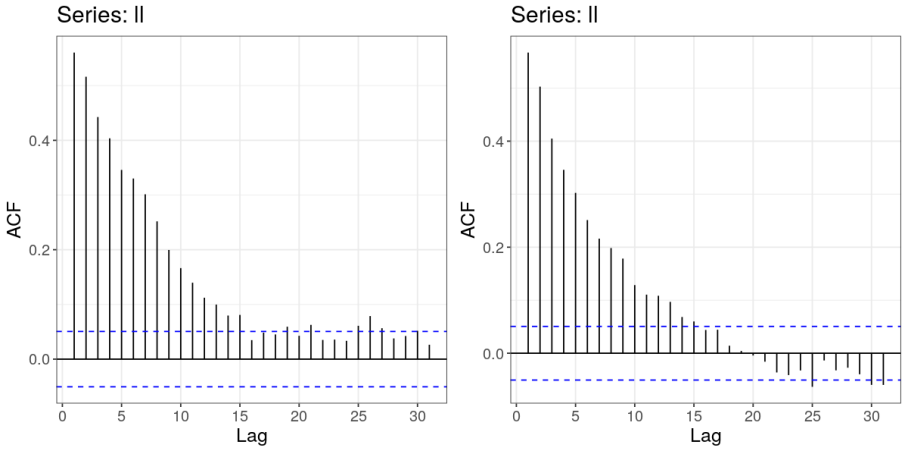


Fig. C.3: Estimated autocorrelation for MCMC samples of the log full data likelihood in case of the constant (left) and quadratic (right) rate model. The dashed blue lines indicate lag-wise 95% probability intervals for the estimated auto-correlations under the null-hypothesis of zero autocorrelation.

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].

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 ns apart.

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 auto-

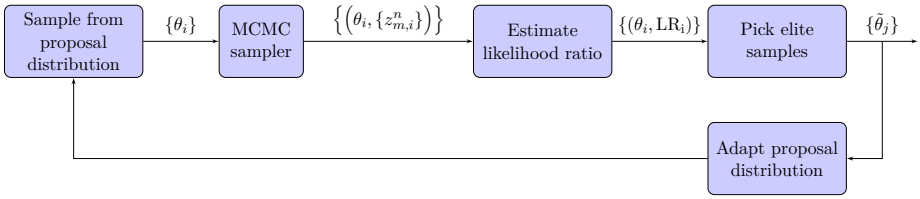


Fig. C.4: An overview of the maximization procedure.

correlation 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 4 below, the autocorrelation plots in Figure C.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.

3.3 Optimization methods

To maximize the Monte Carlo approximation of the likelihood (C.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 3.3. 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 and Hessian matrix required for this is discussed in Section 3.3. Algorithm 4 summarizes the resulting procedure and Fig. C.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 (C.13). Finally these approximations are multiplied to get the approximation of the full likelihood to be maximized.

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

Algorithm 4 CEM Maximization of $L(\theta; \mathbf{y})$

Input. data $\mathbf{y}$, initial guess $\theta_{\max}$, initial CEM proposal distribution.
Output. ML estimate $\hat{\theta}_{\mathsf{ML}}$
$\theta_0 := \theta_{\max}$
Draw sample $\{Z_{\mathsf{m},n}\}_{n \leq N}$ from $p(z_{\mathsf{m}}|\mathbf{y}; \theta_0)$ using Algorithm 3
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 $\theta_{\max}$

Table C.1: 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

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 (C.13) of the likelihood ratio (C.12) becomes unreliable. This situation 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.

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 (C.14).

3.4 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 (C.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 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 (C.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.

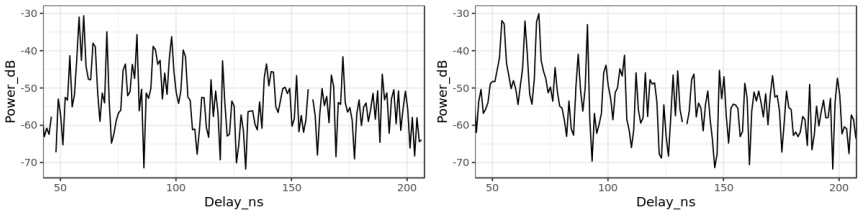


Fig. C.5: Power-delay profile for realizations of the constant (left) and quadratic (right) rate model.

4 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 5. 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 C.5 shows the power-delay profiles simulated from the two models. The settings of the CEM algorithm are given in Tab. C.1.

4.1 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. C.2 shows the estimated log likelihood-ratios

Table C.2: 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

$\log\left(L(\theta_{i+1};\mathbf{y})/L(\theta_i;\mathbf{y})\right)$ 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.

4.2 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.

In general, Figure C.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.

4.3 Unknown multipath components – fixed κ_1

After these initial findings, we now rely on Algorithm 4 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 4.

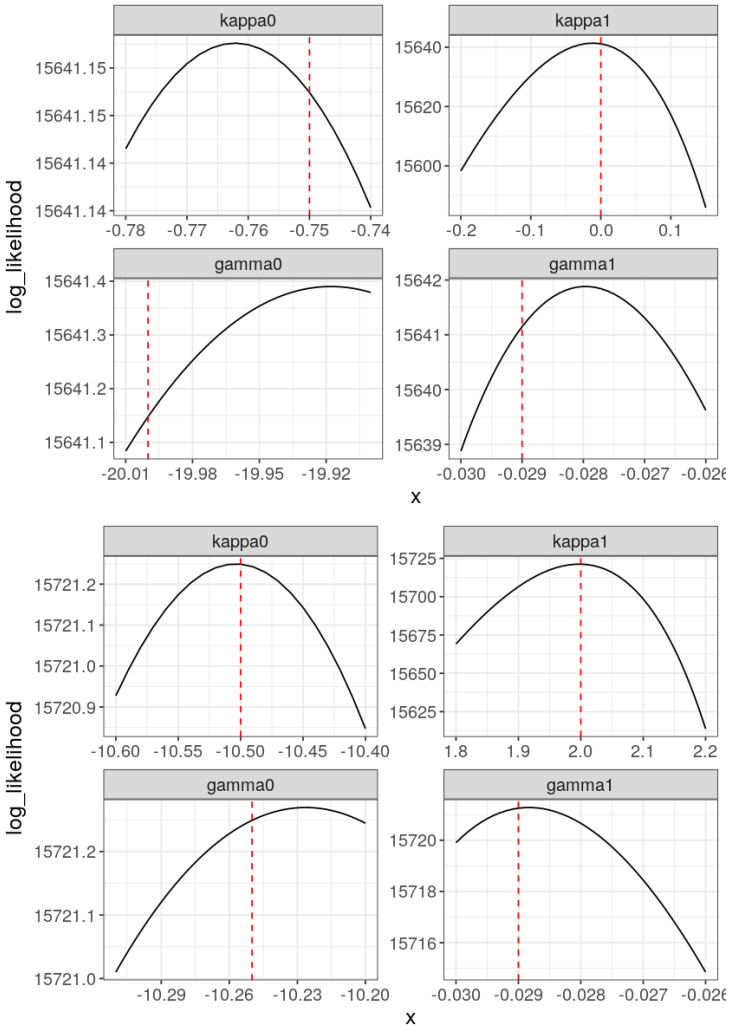


Fig. C.6: Profile plots of the log-likelihood for known multipath components in the constant (top) and quadratic (bottom) rate model. Red dashed lines indicate the true parameters.

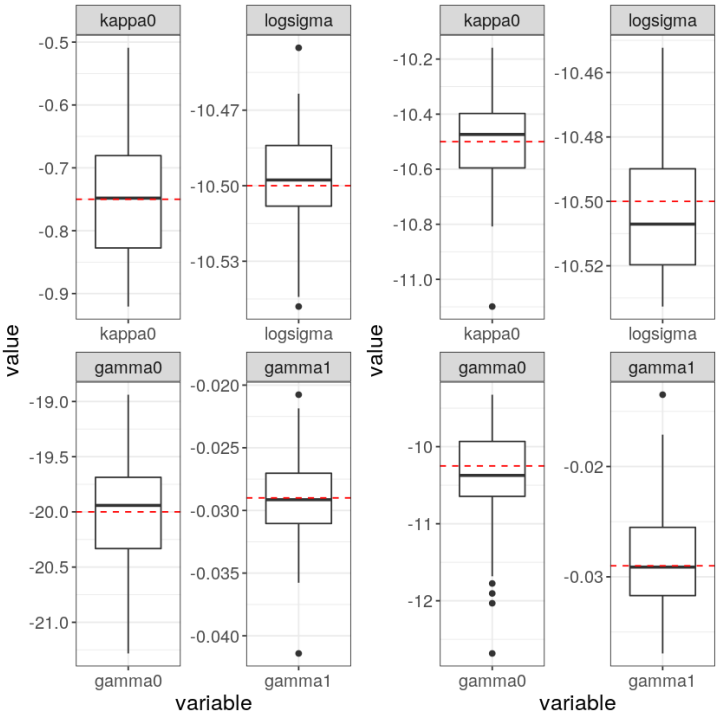


Fig. C.7: Fixed κ_1. Box plots for estimated parameters in 50 realizations of the constant (left) and quadratic (right) rate model. Red dashed lines indicate the true parameters.

In 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 C.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.

4.4 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 re-

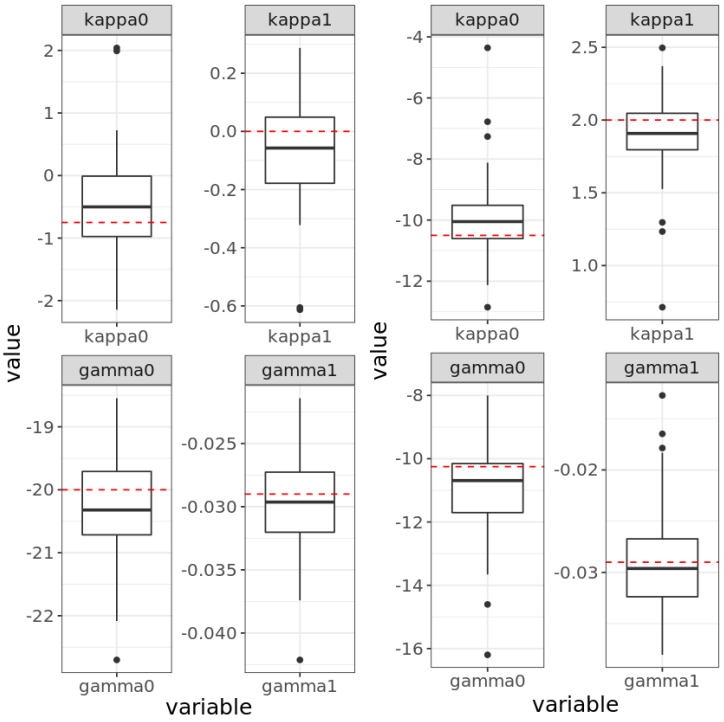


Fig. C.8: Variable κ_1. Box plots for estimated parameters in 50 realizations of the constant (left) and quadratic (right) rate model. Red dashed lines indicate the true parameters. Box plot for $\log \sigma^2$ omitted due to space constraints.

alizations, Figure C.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.

4.5 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 C.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.

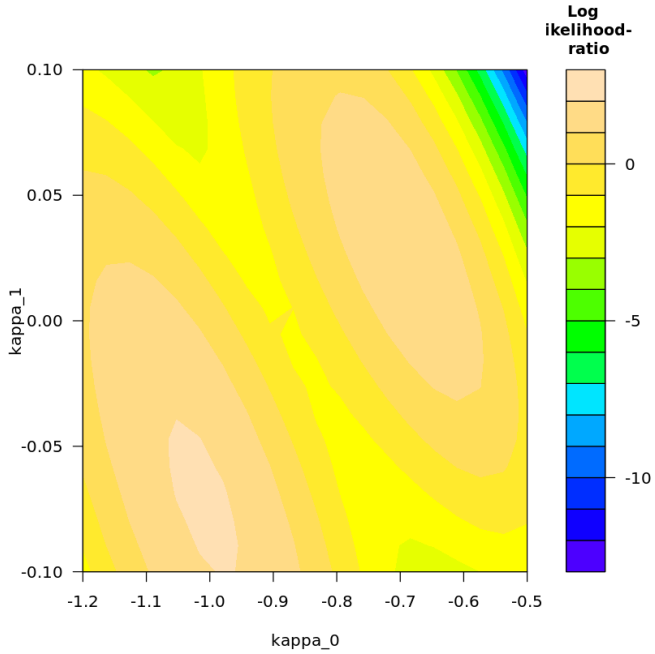


Fig. C.9: Contour plot of the log likelihood-ratios for varying κ_0 and κ_1 based on samples from an MCMC run of length 2,000 in the constant rate model.

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 C.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.

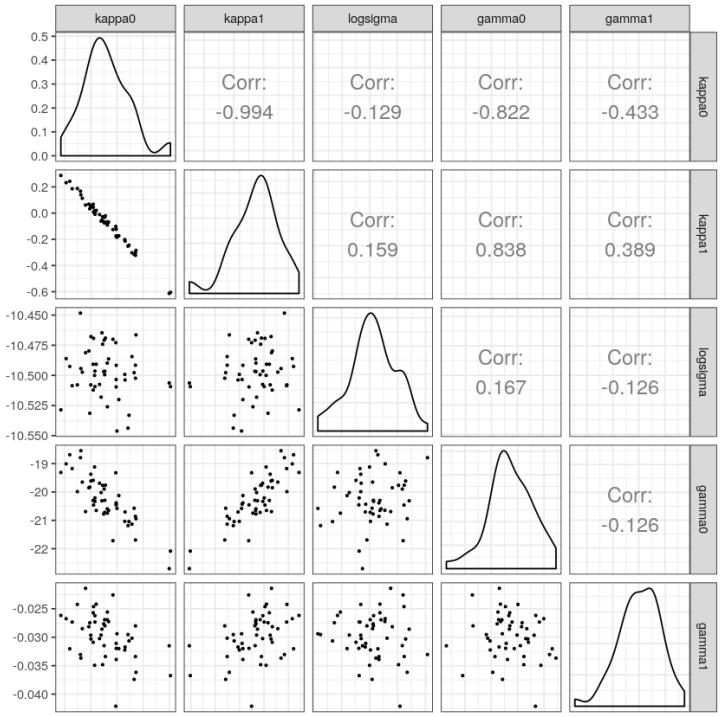


Fig. C.10: Pairwise scatter plots for the parameter estimates of 50 realizations of the constant rate model.

5 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 GHz, 3 GHz]. In particular, the impulse response lies in the interval [0 ns; 750 ns] and decays rapidly after a strong peak close to $\tau_0 = 50$ ns. Therefore, we henceforth work with a window of size $|I| = 150$.

5.1 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

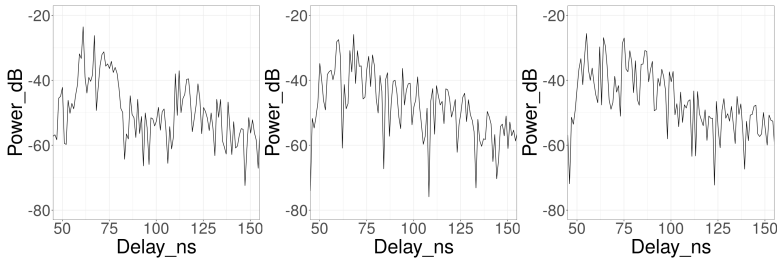


Fig. C.11: Power-delay profile for measurement data (left), a realization of the constant rate model (center) and a realization of the quadratic rate model (right).

the quadratic rate model. Since random fluctuations were stronger than for the synthetic data, we stabilized the optimization by increasing the number of MCMC samples from 1,500 to 2,500. Figure C.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 C.12.

Let $\hat{\theta}_{\mathsf{const}}$ and $\hat{\theta}_{\mathsf{quad}}$ 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 $L(\hat{\theta}_{\mathsf{quad}}, \mathbf{y})/L(\hat{\theta}_{\mathsf{const}}, \mathbf{y})$ via bridge sampling, see Section 2. 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 C.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, for the considered data set, the constant rate model seems more appropriate than the quadratic rate model.

6 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

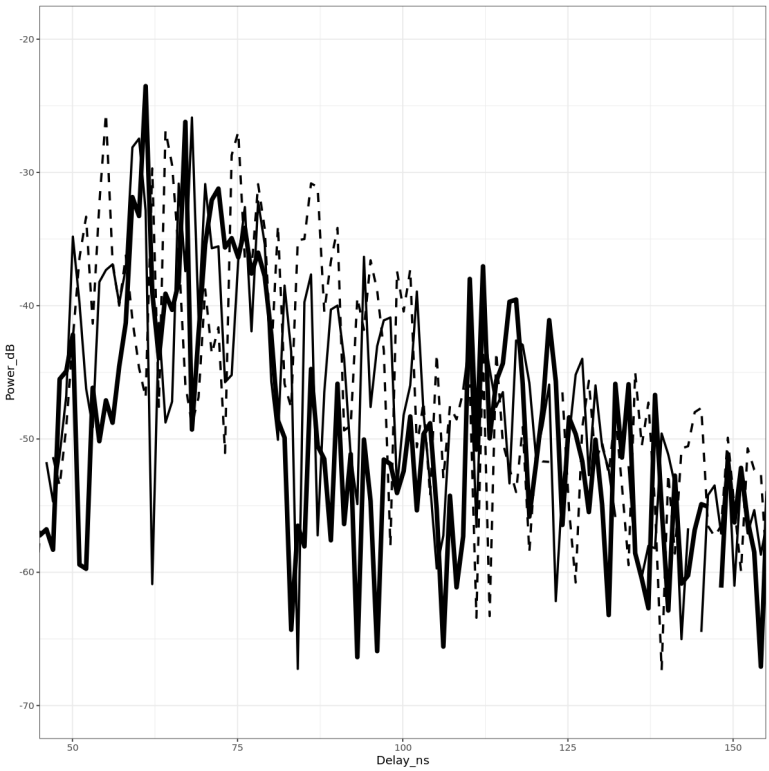


Fig. C.12: Power-delay profile for measurement data (thick), a realization of the constant rate model (thin) and a realization of the quadratic rate model (dashed).

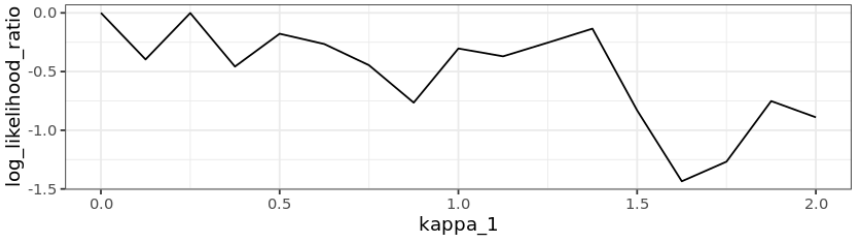


Fig. C.13: Log likelihood ratios $\log(L(\theta_{i+1}, \mathbf{y})/L(\theta_i, \mathbf{y}))$ plotted against κ_1.

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 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*, ser. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, FL, 2004, vol. 100.

[3] 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. 21, no. 1, pp. 1–9, Feb 1972.

[4] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, February 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 *Eur. Conf. on Antennas and Propag.*, 2011, pp. 3606–3610.

[9] P. Kyösti, "Winner II channel models, deliverables d1.1.2 v1.2, part I: Channel models," Tech. Rep. IST-4-027756 WINNER II Project, 2007.

[10] L. Raschkowski, P. Kyösti, K. Kusume, and E. T. Jämsä, "METIS channel models, deliverable d1.4 v3," Tech. Rep. ICT-317669 METIS Project, 2015.

[11] J. A. Högbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *Astronomy and Astrophysics Supplement Series*, vol. 15, p. 417, Jun. 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*, 2004, pp. 53–56.

[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] Chia-Chin Chong and Su Khiong Yong, "A generic statistical-based uwb channel model for high-rise apartments," *IEEE Trans. on Antennas and 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 parameteric MIMO channel data including path powers," in *Proc. IEEE 64th Veh. Technol. Conf.-Fall*, 2006, pp. 1–5.

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

[20] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter Estimation for Stochastic Channel Models using Temporal Moments," in *Proc. 2019 IEEE Int. Symp. on Antennas and Propag. and USNC-URSI Radio Sci. Meeting*, 2019.

[21] ——, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

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

[23] A. Bharti, R. Adeogun, and T. Pedersen, "Learning parameters of stochastic radio channel models from summaries," *IEEE Open J. of Antennas and Propag.*, pp. 1–1, 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.* Interface Foundation of North America, 1991, pp. 156–163.

[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. Steinbock, T. Pedersen, B. H. Fleury, W. Wang, and R. Raulefs, "Distance dependent model for the delay power spectrum of in-room radio channels," *IEEE Tran.s on Antennas and Propag.*, vol. 61, no. 8, pp. 4327–4340, aug 2013.

[28] T. Pedersen, "Modeling of path arrival rate for in-room radio channels with directive antennas," *IEEE Trans. on Antennas and Propag.*, vol. 66, no. 9, pp. 4791–4805, 2018.

[29] ——, "Stochastic Multipath Model for the In-Room Radio Channel Based on Room Electromagnetics," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 4, pp. 2591–2603, April 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*, 2002, pp. 19–23.

[31] T. Pedersen and B. H. Fleury, "A realistic radio channel model based in stochastic propagation graphs," in *MATHMOD*, Feb. 2006, pp. 324–331.

[32] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of reverberant radio channels using propagation graphs," vol. 60, no. 12, pp. 5978–5988, Dec 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. Boca Raton, Florida: Chapman & Hall/CRC, 1999, pp. 79–140.

[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, no. 4, pp. 473–483, 11 1992. [Online]. Available: https://doi.org/10.1214/ss/1177011137

[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. Elsevier, 2013, pp. 35 – 59.

[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.

Paper D

Calibration of Stochastic Channel Models using Approximate Bayesian Computation

Ayush Bharti, Troels Pedersen

The paper has been published in the
IEEE Global Communications Workshop, Globecom, 2019.

Abstract

Calibration of stochastic radio channel models is the process of fitting the parameters of a model such that it generates synthetic data similar to the measurements. The traditional calibration approach involves, first, extracting the multipath components, then, grouping them into clusters, and finally, estimating the model parameters. In this paper, we propose to use approximate Bayesian computation (ABC) to calibrate stochastic channel models so as to bypass the need for multipath extraction and clustering. We apply the ABC method to calibrate the well-known Saleh-Valenzuela model and show its performance in simulations and using measured data. We find that the Saleh-Valenzuela model can be calibrated directly without the need for multipath extraction or clustering.

1 Introduction

Stochastic multipath models are indispensable tools for characterizing the radio channel and analysing the performance of communication systems. Typically, the models are generative, so synthetic data can be generated and used in simulation studies. Such models are particularly useful if calibrated with measurement data. Calibrating a model simply means estimating its parameters such that the model fits to the measurements in some metric. Traditionally, calibration problems are solved by maximizing the likelihood function of the data with respect to the parameters, or by finding the posterior distribution of the parameters (in a Bayesian approach). In either case, access to the likelihood function is required. Unfortunately, the likelihood is intractable here as marginalization with respect to the hidden multipath components is not possible.

Instead of relying on inference frameworks, typically we resort to splitting the inference problem into sub-problems and solve these independently. For example, calibration is typically done in three steps. First, multipath parameters such as delays, gains, etc., of each multipath component is extracted from the channel impulse response measurements. High-resolution algorithms such as SAGE and CLEAN, among others, are used for this step. Then, the extracted multipath components are grouped together into clusters, either manually or using automated clustering algorithms, e.g. [1]. The third step then involves fitting the extracted and clustered multipath components to the model and estimating the parameters. Such multi-step procedures involve complex algorithms that can be cumbersome to use and prone to errors such as estimation artifacts and censoring effects. Moreover, a number of arbitrary choices need to be made in order to implement these algorithms. The estimation accuracy of the parameters then relies on each of the intermediate steps, and therefore the overall performance of these methods is difficult to investigate.

Recently, calibration methods have been proposed in [2, 3] for the Turin model [4] (with constant arrival rate) that bypasses the multipath extraction

step. They summarise the channel measurements into temporal moments and estimate the model parameters directly. While [2] assumes a multivariate Gaussian model for the temporal moments and samples from the approximate posterior of the parameters, [3] fits the expressions of means and covariances of the temporal moments to the parameters using a method of moments approach. However, in cases where the Gaussian assumption fails or when the model is too complicated to derive the theoretical expressions, these methods cannot be utilised. Hence, there is a need for methods to calibrate complex models that involve clustering of multipath components.

The problem of calibrating generative models with unavailable likelihoods appears in many other sciences. One approach to solving this problem is ABC, which was first introduced in population genetics, and further developed and used in many other fields of science as well [5, 6]. ABC relies on comparing summary statistics of simulated and measured data in some distance metric. A potential solution to calibrating stochastic multipath models could be to use ABC methods combined with informative summary statistics such as those studied in [2, 3]. To the best of our knowledge, ABC methods have not been used in radio channel characterization.

In this paper, we propose an ABC algorithm based on population Monte Carlo [7] and regression adjustment [8] for calibration of stochastic channel models. As an example, we develop appropriate summary statistics for calibrating the well-known Saleh-Valenzuela [9] model using ABC. Performance evaluation of this calibration method is carried out via a simulation experiment as well as using measurement data.

2 Estimation Problem Formulation

Consider single-input, single-output (SISO) frequency domain measurement where the received signal sampled at K equidistant frequency points within the bandwidth B reads

$$Y_k = H_k + W_k, \quad k = 0, 1, \ldots, (K-1), \tag{D.1}$$

where H_k is the transfer function, and W_k is iid circular Gaussian measurement noise with variance σ_W^2. Discrete-frequency, continuous-time inverse Fourier transform gives the time domain signal,

$$y(t) = \frac{1}{K} \sum_{k=0}^{K-1} Y_k \exp\left(j2\pi k \Delta f t\right). \tag{D.2}$$

The frequency separation, Δf, is related to the period of the time domain signal as

$$t_{\max} = \frac{1}{\Delta f} = \frac{K-1}{B}. \tag{D.3}$$

A measurement campaign where N_{obs} observations of the channel transfer function are recorded, results in the data matrix $\mathbf{Y}_{\text{obs}} \in \mathbb{C}^{N_{\text{obs}} \times K}$.

A general multipath model of the transfer function may read

$$H_k = \sum_l \alpha_l \exp\left(-j2\pi\Delta k \tau_l\right), \tag{D.4}$$

where τ_l and α_l are the time delay and complex gain of the l^{th} multipath component. The delays and their corresponding gains form a marked point process, $X = \{(\tau_l, \alpha_l)\}$. A stochastic channel model driven by parameters θ is a mechanism that outputs such a marked point process, i.e. $X \sim \mathcal{M}(\theta)$, where $\mathcal{M}(\cdot)$ is the generative model. The model is calibrated by estimating θ from the measurements, $\mathbf{Y}_{\text{obs}}$.

Typically, stochastic multipath models are proposed with the purpose of simulation, i.e. they are generative in nature. However, in the general case of unknown, potentially infinite, number of multipath components, the likelihood $p(X|\theta)$, and consequently, $p(\mathbf{Y}|\theta)$, are analytically and numerically intractable. Therefore, standard inference and sampling techniques methods which rely on the likelihood (or posterior) cannot be applied.

3 Approximate Bayesian Computation

ABC can be used to sample from an approximate posterior when the likelihood is numerically unavailable. Based on these samples, we can approximate the standard point estimates, e.g. the minimum mean squared error (MMSE) estimate.

ABC relies on sampling parameter values from the prior distribution, $p(\theta)$, and simulating datasets, $\mathbf{Y}$, from the model, $\mathcal{M}(\cdot)$. Summary statistics of $\mathbf{Y}$ and $\mathbf{Y}_{\text{obs}}$ are computed using a function, $S(\cdot)$. Let the vector of simulated and observed summary statistics be $\mathbf{s} = S(\mathbf{Y})$ and $\mathbf{s}_{\text{obs}} = S(\mathbf{Y}_{\text{obs}})$, respectively. The summaries are compared in some "distance" measure, $\rho(\cdot, \cdot)$. The basic accept-reject ABC method involves accepting values of θ for which $\rho(\mathbf{s}, \mathbf{s}_{\text{obs}}) < \epsilon$, where ϵ is some pre-defined tolerance threshold. Thus, samples from $\tilde{p}(\theta|\mathbf{s}_{\text{obs}})$ can be obtained, which is an approximation of the desired posterior, $p(\theta|\mathbf{Y}_{\text{obs}})$. The approximation arising here due to the substitution of $\mathbf{Y}_{\text{obs}}$ with $\mathbf{s}_{\text{obs}}$ and the use of ϵ.

Implementation of an ABC algorithm requires the specification of three quantities: the distance measure $\rho(\cdot, \cdot)$, the summary statistics $\mathbf{s}$, and the threshold ϵ. In this paper, we will use the Euclidean distance as our distance function, therefore $\rho(\cdot, \cdot) = ||\cdot||$. Moreover, we consider ϵ in terms of percentile and specify the number of accepted samples, M_ϵ instead. That is, for M samples of θ from the prior, $\epsilon = M/M_\epsilon$. Note that, an ϵ in terms of the distance measure can be used, but it leads to unpredictable run times for a certain number of accepted samples. The selection of appropriate summary

statistics, however, relies heavily on the application (and model) at hand, and will be addressed later with regards to an example channel model.

There exists more advanced ABC methods in the literature than the accept-reject algorithm [5, 6], and in principle, any of them could be used for calibrating stochastic channel models. We propose to use the population Monte Carlo (PMC) ABC [7] with regression adjustment [8].

3.1 Proposed ABC method

The accept-reject ABC algorithm weights each of the accepted samples equally without taking into account the distance from the observed summary statistics. In this paper, we use the method proposed in [8] which improves the approximation of the posterior by weighting the accepted parameter samples, θ_j, according to $||\mathbf{s}_{\text{obs}} - \mathbf{s}_j||$ and adjusting θ_j by using linear regression model applied locally in the vicinity of $\mathbf{s}_{\text{obs}}$. Specifically, the optimisation problem being solved is [8]

$$\underset{a,\mathbf{b}}{\operatorname{argmin}} \sum_{j=1}^{M_\epsilon} \left[\theta_j - a - (\mathbf{s}_{\text{obs}} - \mathbf{s}_j)^T \mathbf{b}\right]^2 K_\epsilon \left(||\mathbf{s}_{\text{obs}} - \mathbf{s}_j||\right) \tag{D.5}$$

Here, $K_\epsilon(\cdot)$ is the Epanechnikov kernel. The samples, $\tilde{\theta}_j$, are then adjusted as

$$\tilde{\theta}_j = \theta_j - (\mathbf{s}_{\text{obs}} - \mathbf{s}_j)^T \hat{\mathbf{b}}. \tag{D.6}$$

This regression-ABC algorithm is described in Alg. 5. For details regarding the solution of (D.5), see [8]. The estimate of the intercept, $\hat{a}$, gives the estimate of the posterior mean, $\mathbb{E}[\theta|\mathbf{s}_{\text{obs}}]$.

After the regression adjustment, some samples of $\tilde{\theta}$ may go beyond the prior range. For example, getting a negative value after adjustment of a strictly positive parameter does not make sense. To avoid this problem, samples of θ outside its prior range are replaced by the extreme points (entrywise).

We combine the regression-ABC with the population Monte Carlo (PMC)-ABC method of [7]. PMC-ABC is a sequential Monte Carlo method that works with a population of θ values instead of one sample at a time. At each iteration, M samples of θ are generated from the M_ϵ accepted samples after performing regression-ABC. The overall algorithm is given in Alg. 6, where θ is assumed univariate. In the general case of θ being a vector of parameters, the same algorithm applies on each entry of θ independently. Note that $\tilde{\mathcal{N}}(\cdot,\cdot)$ denotes a Gaussian truncated to be in the prior range of θ, and $q(\cdot,\cdot)$ is a Gaussian kernel. The overall computational time of Alg. 6 depends on how many observations of the channel, say N_{sim}, are simulated to compute $\mathbf{s}$ from each value of θ.

Algorithm 5 Regression ABC algorithm

Input: Parameter values $(\theta_1, \ldots, \theta_M)$ and corresponding simulated statistics $(\mathbf{s}_1, \ldots, \mathbf{s}_M)$, observed statistics $\mathbf{s}_{\text{obs}}$, number of accepted samples M_ϵ,

1: Accept $(\theta_1^*, \ldots, \theta_{M_\epsilon}^*) \sim \{\theta_i\}_{i=1}^{M}$ with the smallest $\|\mathbf{s}_{\text{obs}} - \mathbf{s}_i\|$, $i = 1, \ldots, N$

2: Solve optimisation problem (D.5) with $\left\{\theta_j^*\right\}_{j=1}^{M_\epsilon}$ and corresponding $\left\{\mathbf{s}_j^*\right\}_{j=1}^{M_\epsilon}$ to get $\hat{\mathbf{b}}$

3: Adjust accepted samples $\left\{\theta_j^*\right\}_{j=1}^{M_\epsilon}$ using (D.6) to get $\left\{\tilde{\theta}_j\right\}_{j=1}^{M_\epsilon}$ in the prior range

Output: Samples $(\tilde{\theta}_1, \ldots, \tilde{\theta}_{M_\epsilon})$ from approximate posterior

4 Application to Saleh-Valenzuela model

The seminal model proposed by Saleh-Valenzuela [9] has been the basis of standardized models such as IEEE 802.15.3a and IEEE 802.15.4a. It has been extended further to include the spatial properties of the channel and also been applied in millimetre wave scenarios using a MIMO system. Calibration methods for this model commonly rely on multipath extraction and clustering. For this reason, we demonstrate the utility of the proposed ABC method by applying it to the Saleh-Valenzuela model.

4.1 Estimation problem

The Saleh-Valenzuela model can be formulated in the frequency domain as

$$H_k = \sum_l \sum_p \beta_{pl} \exp\left(-j2\pi\Delta f k(T_l + \tau_{pl})\right), \tag{D.7}$$

where T_l is the delay of the l^{th} cluster, while τ_{pl} and β_{pl} are the delay and complex gain of the p^{th} ray within the l^{th} cluster, respectively. By definition in [9], $T_0 = 0$ and $\tau_{0l} = 0$, $l \in \{0, 1, \ldots\}$. The arrival time of the clusters and that of the rays within the clusters are modelled as one-dimensional homogeneous Poisson point processes, i.e., $T_l \sim \text{PPP}(\mathbb{R}_+, \Lambda)$ and $\tau_{kl} \sim \text{PPP}(\mathbb{R}_+, \lambda)$ with parameters $\Lambda, \lambda > 0$. The gains β_{kl}, conditioned on T_l and τ_{kl}, are modelled as iid zero-mean complex Gaussian random variables. Their conditional variance is modelled as

$$\mathbb{E}\left[|\beta_{kl}|^2 | T_l, \tau_{kl}\right] = Q \exp(-T_l/\Gamma) \exp(-\tau_{kl}/\gamma), \tag{D.8}$$

with Q being the average power of the first arriving multipath component, and $\Gamma, \gamma > 0$ being the cluster and ray decay constants, respectively. The expression for the power delay spectrum is given in [10]. To calibrate this model, the parameter vector, $\theta = [Q, \Lambda, \lambda, \Gamma, \gamma]^T$, should be estimated based on $\mathbf{Y}_{\text{obs}}$.

Algorithm 6 PMC-ABC with regression adjustment

Input: Prior $p(\theta)$, model $\mathcal{M}(\theta)$, observed statistics $\mathbf{s}_{\text{obs}}$, M_ϵ, M, T

At iteration $t = 1$,
Sample M samples of θ from the prior, i.e. $(\theta_1, \ldots, \theta_M) \sim p(\theta)$
Generate $\mathbf{s}_i \sim \mathcal{M}(\theta_i)$, $i = 1, \ldots, M$
Perform **Algorithm** 5 on $(\theta_1, \ldots, \theta_M)$ and $(\mathbf{s}_1, \ldots, \mathbf{s}_M)$ to generate $(\theta_1^{(1)}, \ldots, \theta_{M_\epsilon}^{(1)})$
Set $w_i^{(1)} = 1/M_\epsilon$ for $i = 1, \ldots, M_\epsilon$ and take $\sigma^{(1)}$ to be twice the sample variance of $(\theta_1^{(1)}, \ldots, \theta_{M_\epsilon}^{(1)})$
for $t = 2, \ldots, T$ **do**
  for $j = 1, \ldots, M$ **do**
    Sample $\theta_j^* \sim (\theta_1^{(1)}, \ldots, \theta_{M_\epsilon}^{(1)})$ with probability $(w_1^{(1)}, \ldots, w_{M_\epsilon}^{(1)})$
    Perturb θ_j^* by sampling $\theta_j^{**}|\theta_j^* \sim \tilde{\mathcal{N}}\left(\theta_j^*, \sigma^{(t-1)}\right)$
    Generate $\mathbf{s}_j \sim \mathcal{M}(\theta_j^{**})$
  end for
  Perform **Algorithm** 5 on $(\theta_1^{**}, \ldots, \theta_M^{**})$ and $(\mathbf{s}_1, \ldots, \mathbf{s}_M)$ to generate $(\theta_1^{(t)}, \ldots, \theta_{M_\epsilon}^{(t)})$
  Set $w_i^{(1)} \propto p(\theta_i^{(t)}) / \sum_{j=1}^{M_\epsilon} w_j^{(t-1)} q\left(\theta_j^{(t-1)}|\theta_i^{(t)}, \sigma^{(t-1)}\right)$, and take $\sigma^{(t)}$ to be twice the sample variance of $(\theta_1^{(t)}, \ldots, \theta_{M_\epsilon}^{(t)})$
end for

Output: Samples $(\theta_1^{(T)}, \ldots, \theta_{M_\epsilon}^{(T)})$ from the approximate posterior $\tilde{p}(\theta|\mathbf{s}_{\text{obs}})$

4.2 Development of summary statistics

In order to use the ABC method to infer on θ, informative summary statistics of $\mathbf{Y}$ are needed about the five model parameters. The selection of summaries determines the degree to which we approximate the posterior distribution by replacing $\mathbf{Y}$ with $\mathbf{s}$.

Temporal moments are widely used statistics that have been recently used to estimate the parameters of the Turin model [4] without multipath extraction in [2, 3]. We define the generalised i^{th} temporal moment as

$$m_i^{(n)} = \int_0^{t_{\max}} t^i |y(t)|^n \mathrm{d}t, \quad i = 0, 1, 2, \ldots. \tag{D.9}$$

Most commonly, only temopral moments with $n = 2$ are considered. The temporal moments are random variables that summarise each channel realisation into one instance for a particular n and i. So, N_{obs} realisations would lead to a vector of temporal moments, $\mathbf{m}_i^{(n)}$ for each choice of i and n. We use the first three temporal moments, i.e. $D^{(n)} = [\mathbf{m}_0^{(n)}, \mathbf{m}_1^{(n)}, \mathbf{m}_2^{(n)}]$. We further summarise the n^{th} order temporal moments by computing the sample mean vector, $\mu_D^{(n)}$, and covariance matrix, $\Sigma_D^{(n)}$. This results in nine distinct sum-

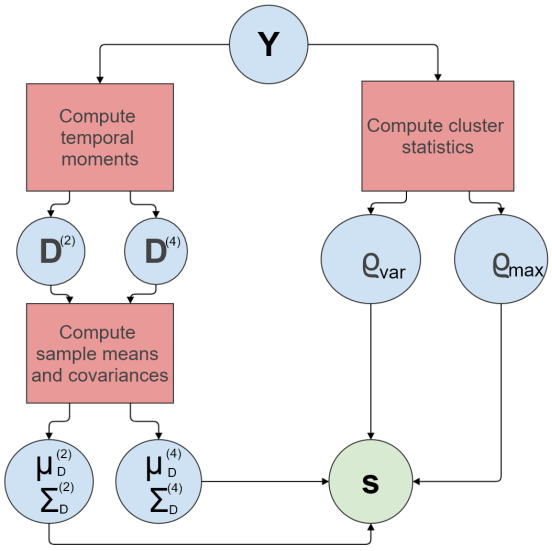


Fig. D.1: Data flow from measurements to proposed summary statistics.

mary statistics for each n. Since the kurtosis of channel measurements is known to be informative about the arrival rate [11, 12], we also include $\mu_D^{(4)}$ and $\Sigma_D^{(4)}$ as summaries.

We observed through simulation (not shown due to space limitations) that the temporal moments alone were not as informative about the cluster parameters (Λ, Γ). For this reason, we include two additional summaries, $\varrho_{\max}$ and ϱ_{var}, to measure the "unorderedness" of the averaged power delay profile (APDP). Consider the discrete-time APDP with index sequence $\mathbf{v} = [1, \ldots, K]$. Sorting the APDP in descending order leads to a permutation of indices, say $\mathbf{v}'$. Then the two summaries are computed as

$$\varrho_{\max} = \max\left(\mathbf{v} - \mathbf{v}'\right), \quad \text{and} \tag{D.10}$$

$$\varrho_{\text{var}} = \text{var}\left(\nu\right), \tag{D.11}$$

where ν contains all positive entries of $(\mathbf{v} - \mathbf{v}')$.

The resulting collection of summaries yields a twenty dimensional vector, $\mathbf{s}$, as outlined in Fig. D.1. The choice of summary statistics is not limited to the ones discussed here. In principle, other summaries could also be used, as long as they are informative about the model parameters.

4.3 Setting priors for parameters

The priors are used to sample θ values. Here we use flat priors for all the parameters as a way limit the search space in the first iteration of Alg. 6. The

Table D.1: Parameter settings and estimates.

θ	Prior range	Simulated			Measured	
		θ_{true}	$\hat{\theta}$	θ_{sd}	$\hat{\theta}$	θ_{sd}
Q	$(10^{-9}, 10^{-9})$	5×10^{-8}	4.4×10^{-8}	5.5×10^{-9}	9.6×10^{-9}	1.4×10^{-9}
Λ[s^{-1}]	$(5 \times 10^6, 10^8)$	10^7	5.5×10^7	4.7×10^7	2.0×10^7	1.4×10^7
λ[s^{-1}]	$(5 \times 10^6, 4 \times 10^9)$	10^9	1.2×10^9	2.8×10^8	7.2×10^8	3.0×10^8
Γ	$(0, 10^{-7})$	10^{-8}	8.4×10^{-9}	5.7×10^{-10}	2.8×10^{-8}	4.2×10^{-9}
γ	$(0, 10^{-7})$	10^{-9}	4.5×10^{-9}	1.2×10^{-9}	1.2×10^{-8}	4.4×10^{-9}

range should be wide enough to capture all the plausible values of the parameter, but not so wide that the computational time of the algorithm becomes excessive. The specific ranges of the priors, given in Table D.1, are obtained by the following reasoning.

The prior for Q is easily set from the PDP of measurements. Here, we set it to be ± 10 dB around the power of the first observed component. For Λ, on an average we expect to see a minimum of one cluster in the data. This gives the lower bound of the prior as $1/t_{\max}$. We limit our search to at most 20 clusters arriving on an average, giving the upper bound as $20/t_{\max}$. Similar argument is applied for λ, where the minimum and maximum number of rays arriving on an average range from one to K. The decay constants are positive parameters, so their lower bound is set to zero. The upper bound is limited to a number that will lead to a 100 dB drop in the power over half the observation window, i.e. $0.5t_{\max}/100$.

5 Performance Evaluation

5.1 Simulation Experiment

We first test the proposed algorithm by applying it to simulated data. A synthetic data set is drawn from the Saleh-Valenzuela model with "true" parameters (denoted as θ_{true}) given in Table D.1, and summarised as per Fig. D.1. The results obtained by applying the PMC-ABC with regression adjustment to this data are reported in Fig. D.2. The approximate MMSE estimate, defined as

$$\hat{\theta} = \frac{1}{M_\epsilon} \sum_{j=1}^{M_\epsilon} \tilde{\theta}_j^{(T)}, \tag{D.12}$$

is also reported in Table D.1, along with the standard deviation of the posterior after T iterations, θ_{sd}.

It appears that the (approximate) posterior converges for all the parameters. However, the marginal posterior for Λ converges more slowly than for the other four parameters. The accepted samples are around the true value of the parameter, indicating that the method is able to estimate the parameters.

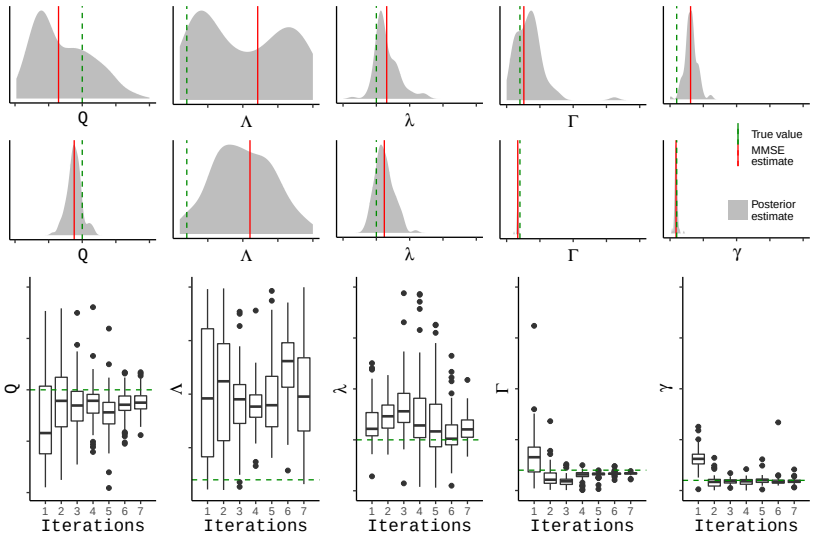


Fig. D.2: Density estimate of approximate posteriors of the parameters obtained after first (top panel) and seventh (middle panel) iteration of Alg. 6 on simulated data. The bottom panel shows boxplot of accepted samples as a function of iterations. The black dots denote the outliers. The parameter axes are in their prior range. Simulation settings: $N_{\text{sim}} = 50$, $B = 4$ GHz, $K = 801$, $M = 2000$, $M_\epsilon = 100$, $T = 7$.

5.2 Application to measured data

We now apply the proposed method to a set of in-room transfer function measurements obtained using a vector network analyser and described in [13]. The set-up is SISO, with a virtual planar array of 25×25, giving $N_{\text{obs}} = 625$. The dimensions of the room was $3 \times 4 \times 3$ m^3. The signal was measured at $K = 801$ frequency points in the range 58 GHz to 62 GHz. Thus, $B = 4$ GHz, $\Delta f = 5$ MHz, and $t_{\max} = 200$ ns. The results are shown in Fig. D.3. The convergence behaviour of the approximate posterior for measurements is similar to that observed in simulations. The posteriors for all the parameters but Λ are concentrated in a small region of the prior range.

5.3 Refining the posterior for Λ

As observed from Fig. D.2 and D.3, $\tilde{p}(\Lambda|\mathbf{s}_{\text{obs}})$ converges more slowly than the posterior of other parameters. One method to speed up the convergence is to limit the parameter search space by fixing a subset of the parameters to particular values, and only sampling from the remaining variables. Here, we fix all but Λ to their MMSE estimate, $\hat{\theta}$, and sample only from $p(\Lambda|\mathbf{s}_{\text{obs}})$ by applying Alg. 6 with $T = 1$. Fig. D.4 shows the approximate posterior obtained with and without the regression adjustment.

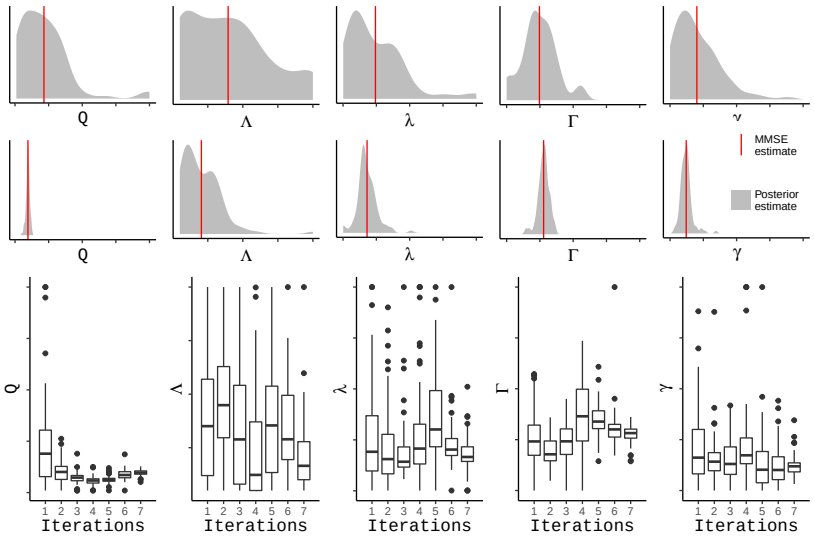


Fig. D.3: Density estimate of approximate posteriors of parameters obtained after first (top panel) and seventh (middle panel) iteration of Alg. 6 on measured data. The bottom panel shows boxplot of accepted samples as a function of iterations. Simulation settings as in Fig. D.2.

For the simulated data, the posterior obtained without regression adjustment is much narrower compared to Fig. D.2, and the MMSE estimate agrees well with the true value. Regression adjustment narrows the posterior even further. Similar behaviour is observed in the measurements. We also observe $\hat{\Lambda}$ to be very small in the measurements. This is particularly true when the regression adjustment is applied, which shifts the posterior below the prior range. This indicates that clusters are rare in the measured data. The rarity of clusters is confirmed via visual inspection of the PDP of the data. Consequently, the lower bound of $p(\Lambda)$ could be chosen even smaller.

6 Conclusion

We proposed the PMC-ABC with regression adjustment method for calibrating stochastic channel models with intractable likelihoods. The method is effective in calibrating the Saleh-Valenzuela model from measurement data without the need for multipath extraction and clustering. The temporal moments, combined with the proposed cluster statistics, are found to be informative about the model parameters. The summaries are observed to be less responsive to the cluster arrival rate than the other parameters. As a result, the marginal posterior of cluster arrival rate converges slower than others. However, its estimate has been shown to improve using an additional round of the ABC

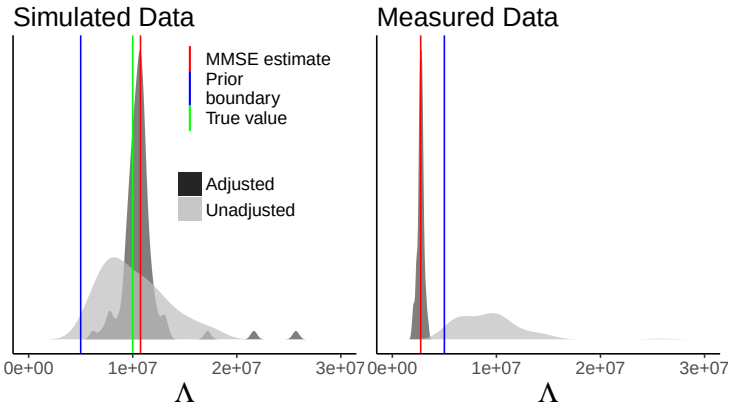


Fig. D.4: Approximate posterior of Λ with and without the regression adjustment for simulated and measured data.

method. Similar convergence behaviour is observed on applying the method to measurement data. From the measurements, it seems that the cluster arrival rate is outside the prior range, suggesting the absence of clusters in this data. Non-clustered models might be a better fit for this data. In any case, a smaller lower bound for the prior of this parameter should be chosen for future studies.

Despite the fact that the method is developed without considering the measurement noise, it performs well on the noisy measurements. We conjecture that the method can be improved by including the noise variance as a parameter in the method, which may require an additional summary. This method can also be extended to calibrate spatial models, provided appropriate summaries are available.

Acknowledgment

The authors would like to thank Dr. Carl Gustafson and Prof. Fredrik Tufvesson (Lund University) for providing the measurement data. This work is supported by the Danish Council for Independent Research, grant no. DFF 7017-00265 and performed within the framework of the COST Action CA15104 IRACON.

References

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

[2] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter Estimation for Stochastic Channel Models using Temporal Moments," in *Proc. 2019 IEEE Int. Symp. on Antennas and Propag. and USNC-URSI Radio Sci. Meeting*, 2019.

[3] ——, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

[4] 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. 21, no. 1, pp. 1–9, Feb 1972.

[5] B. M. Turner and T. V. Zandt, "A tutorial on approximate bayesian computation," *J. Math. Psychol.*, vol. 56, no. 2, pp. 69–85, apr 2012.

[6] M. A. Beaumont, "Approximate bayesian computation in evolution and ecology," *Annu. Rev. Ecol. Evol. Syst.*, vol. 41, no. 1, pp. 379–406, dec 2010.

[7] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate bayesian computation," *Biometrika*, vol. 96, no. 4, pp. 983–990, Oct 2009.

[8] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002. [Online]. Available: https://www.genetics.org/content/162/4/2025

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

[10] M. L. Jakobsen, T. Pedersen, and B. H. Fleury, "Analysis of the stochastic channel model by saleh & valenzuela via the theory of point processes," *Int. Zurich Seminar on Commun.*, 2012.

[11] T. Pedersen, "Stochastic Multipath Model for the In-Room Radio Channel Based on Room Electromagnetics," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 4, pp. 2591–2603, April 2019.

[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, 2013.

[13] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mmwave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

Paper E

Learning Parameters of Stochastic Radio Channel Models from Summaries

Ayush Bharti, Ramoni Adeogun, Troels Pedersen

The paper has been published in the
IEEE Open Journal of Antennas and Propagation, 2020.

Abstract

Estimating parameters of stochastic radio channel models based on new measurement data is an arduous task usually involving multiple steps such as multipath extraction and clustering. We propose two different machine learning methods, one based on approximate Bayesian computation (ABC) and the other on deep learning, for fitting data to stochastic channel models directly. The proposed methods make use of easy-to-compute summary statistics of measured data instead of relying on extracted multipath components. Moreover, the need for post-processing of the extracted multipath components is omitted. Taking the polarimetric propagation graph model as an example stochastic model, we present relevant summaries and evaluate the performance of the proposed methods on simulated and measured data. We find that the methods are able to learn the parameters of the model accurately in simulations. Applying the methods on 60 GHz indoor measurement data yields parameter estimates that generate averaged power delay profile from the model that fits the data.

1 Introduction

Stochastic models of the radio channel are indispensable tools in the design and analysis of communication and localization systems. Stochastic radio channel models are used for characterizing and simulating realizations of the channel in different environments. However, for the model to generate data similar to what is observed in the measurements, its parameters need to be learned from the data. The process of learning or estimating the parameters of a model from new measurements is termed as calibration. Calibration could be obtained by deriving the parameters theoretically, e.g. in room electromagnetics or in ray tracing. In fact, some parameters such as speed of light or room geometry are set not using the data. Standard calibration technique using data would be to either maximize the likelihood function of the data with respect to the parameters, or to characterize the posterior distribution of the parameters in a Bayesian sense. However, most stochastic channel models suffer from intractability of the likelihood function, and therefore, calibrating them given a new set of measurement data is challenging [1].

Typically, stochastic multipath radio channel models are calibrated in steps, as described in Fig. E.1(a). This calibration methodology has been followed since the early works of Turin [2] and Saleh-Valenzuela [3] till more recent stochastic channel models [4–8]. First, the data is reduced to a set of multipath components, each having their own gain, delay, etc., by applying high-resolution algorithms such as SAGE (Space Alternating Generalized Expectation maximization) [9], MUSIC (MUltiple SIgnal Classification), ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques), and RiMAX [10], among others. The extracted multipath components are then used to estimate the model parameters. In case of cluster-based channel models, an additional

clustering step can be applied. Alternatively, the presence of clusters, or other multipath effects, can be included in the derivation of the high-resolution estimator, as in [11–13], to obtain the cluster parameters directly. In a final step, the model parameters are estimated from the extracted multipath components and clusters.

Even though such multi-step calibration approaches are widely used, they suffer from a range of issues. It can be cumbersome and labor-intensive to derive, implement, and test sophisticated multipath extraction and clustering algorithms that require a number of heuristic choices to be made. Moreover, the estimation of the multipath components is prone to errors due to censoring [14]. Therefore, the overall estimation accuracy of the model parameters is difficult to determine due to this step-by-step calibration approach.

There may exist statistics other than the multipath components that are easier to obtain, and still hold enough information to be able to learn the model parameters. Potentially, the parameters of the channel models can be estimated without the multipath extraction step by relying on these easy-to-compute summary statistics. Such an estimator for the Saleh-Valenzuela model [3] was proposed in [15] where the estimation problem was framed as an optimization problem that fitted summary statistics of the data with approximate analytical expressions. More recently, multipath extraction-free calibration methods based on sampling [16] and method of moments [17] have been developed and applied to the Turin model. These methods summarize the data into certain statistics, and rely on explicit derivation of equations linking their means and covariances to the model parameters. Drawback of these methods is that such equations need to be derived for every stochastic channel model, which is either non-trivial or oftentimes not possible.

In the present contribution, we further advance the idea of using other summaries than the multipath parameters for model calibration. This leads to the calibration methodology outlined in Fig. E.1(b) where the data is first summarized into a set of statistics from which the model parameters are obtained. We extend our previous work on learning parameters of stochastic channel models using approximate Bayesian computation (ABC) [18] and deep learning (DL) [19], where we had applied the methods on the cluster-based model of Saleh-Valenzuela [3].

ABC is a framework for performing likelihood-free inference on generative models with intractable likelihoods [20, 21], as is the case for stochastic channel models. It relies on sampling parameter values from a prior distribution, simulating data from the model, and comparing the simulated summaries to those obtained from the measurements. Parameter values that yield summaries similar to the observed ones are used to approximate the posterior distribution. Initially developed in the field of population genetics [22], ABC has since been applied in various other fields of research such as ecology [23], astrophysics [24], and structural dynamics [25], to name a few. To the best of our knowledge, ABC has not been applied in wireless communications except for our previous conference paper [18].

The DL method utilizes a neural network (NN) to establish a functional relationship between the summaries and the parameters, and uses the trained NN to estimate the parameters given the observed summaries. Neural networks have been proven to exhibit capability for universal approximation of any continuous real-valued function [26]. These networks have been successfully applied in fields such as computer vision and image processing over the last several years. Recently, the wireless communications community has also explored avenues for application of DL. While efforts have been directed towards DL-enabled physical layer design, only a few applications to radio channel modeling and calibration have been proposed. In [27], the authors utilized Deep Neural Network for uplink-downlink channel calibration in massive MIMO. Similar network is utilized for predicting path-loss exponent from millimeter wave channel measurements in [28]. The DL method proposed here is a generalization of the framework introduced in [19] where a single layer neural network is applied to estimate parameters of propagation models. Similar DL-based likelihood-free inference framework has been applied in population genetics [29].

In this paper, we present two machine learning methods based on ABC and DL to calibrate stochastic radio channel models without multipath extraction. We show the applicability of the methods by calibrating the polarimetric propagation graph (PG) model [30] as an example, since multipath extraction cannot be directly applied to calibrate it. We also present a number of summary statistics for representing channel measurements that are used as input for the learning methods. The chosen statistics are qualified via simulation study of the PG model. Simulation results illustrate the capability of the proposed learning methods to accurately estimate the model parameters. The methods are also applied to calibrate the PG model using real indoor channel measurements. Reasonable fits of the averaged power delay profile were seen between the measurements and the model, thus validating the proposed learning methods.

The remainder of the article is organized as follows. We describe the ABC and DL calibration methods in section II. The description of the polarimetric PG model is given in section III, along with the choice of summary statistics. In section IV, we apply the ABC and DL methods on the PG model and show the results. The discussion regarding the proposed methods is presented in section V, and section VI provides our concluding remarks.

2 Summary-Based Calibration Methods

Consider a stochastic generative model, $\mathcal{M}(\boldsymbol{\theta})$, which is easy to simulate from given any value of the parameter, $\boldsymbol{\theta}$. Each time the model is called for a given parameter vector, it generates independent realizations of simulated data, $\mathbf{Y}$. Here, $\mathbf{Y}$ could be a vector or a matrix. Let $\mathbf{Y}_{\text{obs}}$ be a set of measurement data obtained experimentally. The calibration problem then involves estimat-

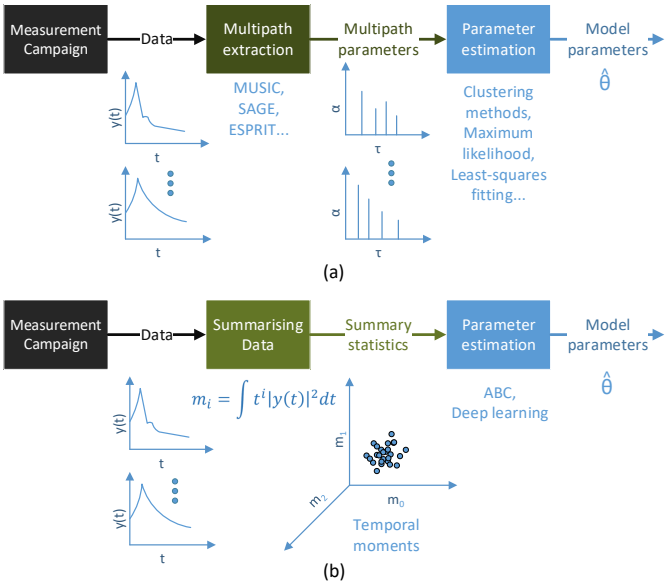


Fig. E.1: State-of-the-art calibration methodology (a), versus the proposed method (b).

ing $\boldsymbol{\theta}$ such that the model, $\mathcal{M}(\boldsymbol{\theta})$, fits to the measured data, $\mathbf{Y}_{\text{obs}}$. However, the likelihood function of $\mathbf{Y}$ given $\boldsymbol{\theta}$ is intractable, and so standard estimation techniques are not applicable. Typically, $\mathbf{Y}$ is a high-dimensional data matrix, as is the case with radio channel transfer function or impulse response measurements for multiple independent realizations of the channel. Therefore, we use a function, $S(\cdot)$, that summarizes $\mathbf{Y}$ into a set of q statistics, $\mathbf{s} \in \mathbb{R}^q$, such that $\mathbf{s} = S(\mathbf{Y})$. We then use these statistics as data in our calibration methods to estimate $\boldsymbol{\theta}$ given the observations $\mathbf{s}_{\text{obs}} = S(\mathbf{Y}_{\text{obs}})$. Ideal choice for $\mathbf{s}$ would be sufficient statistics of $\mathbf{Y}$, but those are unavailable in most practical cases.

2.1 Approximate Bayesian Computation

ABC is a likelihood-free inference method that samples from the approximate posterior distribution of the parameters by finding values that lead to simulated datasets from the model that are similar to the observed data. The method involves sampling from the prior distribution of the parameters, $p(\boldsymbol{\theta})$, and then generating datasets from the model. These simulated datasets are then compared to the observed set of measurements in some distance metric, $\rho(\cdot, \cdot)$, and the values of $\boldsymbol{\theta}$ that yield a distance smaller than a pre-defined tolerance threshold, ϵ, form samples from the approximate posterior distribution. These samples can then be used to approximate standard point estimates of $\boldsymbol{\theta}$ such as the minimum mean square error (MMSE) or the maximum *a posteriori*

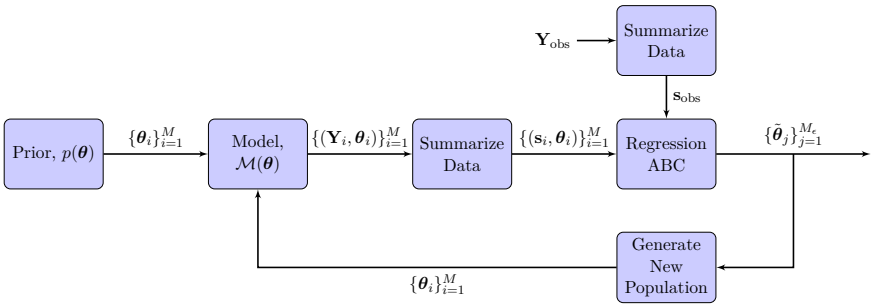


Fig. E.2: Block diagram depicting data flow in the proposed PMC-ABC algorithm with regression adjustment.

estimate. The basic rejection-ABC algorithm can be summarized as follows:

1. Sample from prior, $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$
2. Simulate $\mathbf{Y}^* \sim \mathcal{M}(\boldsymbol{\theta}^*)$
3. Compute summary statistics $\mathbf{s}^* = S(\mathbf{Y}^*)$
4. Accept $\boldsymbol{\theta}^*$ if $\rho(\mathbf{s}^*, \mathbf{s}_{\text{obs}}) < \epsilon$

The approximation here arises on account of summarizing the data into a set of statistics, and accepting samples within a tolerance threshold. Choosing sufficient statistics to be used in ABC mitigates the former approximation, however, finding sufficient statistics is typically not feasible. Choosing a small ϵ improves the posterior approximation but increases the rejection rate significantly, whereas a large ϵ leads to sampling from almost the prior. Therefore, the distance metric, the tolerance threshold, and the summary statistics are the necessary ingredients required to implement an ABC algorithm.

The choice of appropriate summary statistics is crucial to the quality of approximation [31], and it depends upon the application and the model at hand. Given a pool of summary statistics, there exist statistical methods to e.g. select best subset among them, or construct a much smaller set of highly informative statistics through projection techniques, among others [32]. However, domain knowledge is vital for constructing appropriate summaries that are not only informative about the model parameters, but also relevant from an application perspective. Thus, domain experts can choose meaningful summaries that they wish to fit to the model. The specific statistics used in this paper will be addressed in Sec. 3.4 with regards to a specific example radio channel model.

Typically, the Euclidean distance between the observed and the simulated statistics is used as the distance metric in summary-based ABC methods, i.e. $\rho(\cdot, \cdot) = \|\cdot\|$. Other distance metrics can also be used, however, the Euclidean distance seems appropriate in the context of summary-based ABC methods for the application at hand. In cases where a set of statistics are used that differ in their units and order of magnitude, it is important to normalize them

before computing the distance [33]. The normalization aims to bring all the individual distances to the same scale, such that the total distance will be their sum. In this paper, we take $\rho(\cdot,\cdot)$ to be the Euclidean distance between the normalized statistics, but still denote it as $\|\mathbf{s}-\mathbf{s}_{\text{obs}}\|$, with the normalization of the statistics assumed to be implicit.

Specifying an appropriate value of ϵ in terms of the distance may prove to be difficult. Setting ϵ too low leads to unknown run time of the algorithm to get a certain number of accepted samples. Therefore, it is usual to employ a k-nearest neighbor approach and specify ϵ as a percentile of the total simulated samples. That is, out of M samples of $\boldsymbol{\theta}$ from the prior, we accept the first $M_\epsilon = \epsilon M$ samples leading to the smallest $\|\mathbf{s}-\mathbf{s}_{\text{obs}}\|$.

The basic ABC method is simple but can be rather slow. Instead, we propose to use a sequential sampling method, specifically the Population Monte Carlo ABC (PMC-ABC) [34], and supplement it by employing the local-linear regression adjustment method proposed in [31]. The resulting algorithm, named PMC-ABC with regression adjustment, is detailed in Alg. 8. A block diagram summarizing the proposed method is shown in Fig. E.2. In the following, we describe the two ABC techniques.

Regression ABC

We supplement the rejection-ABC algorithm by employing the local-linear regression adjustment method proposed in [31]. The regression adjustment improves the posterior approximation by 1) weighting the accepted parameter samples according to their corresponding distance value and 2) adjusting them using a linear regression model applied locally in the vicinity of $\mathbf{s}_{\text{obs}}$. For $\boldsymbol{\theta}_j$ being the j^{th} accepted parameter sample and $\mathbf{s}_j$ the corresponding simulated statistics vector, the linear model reads

$$\boldsymbol{\theta}_j = \boldsymbol{\alpha} + (\mathbf{s}_j - \mathbf{s}_{\text{obs}})^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j, \quad j = 1, \ldots, M_\epsilon, \tag{E.1}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^p$ and $\boldsymbol{\beta} \in \mathbb{R}^{q\times p}$ are the intercept and regression coefficients, respectively, and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{M_\epsilon}$ are uncorrelated noise variables with zero mean. The least-squares estimate of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are obtained by solving the optimization problem

$$\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{j=1}^{M_\epsilon} \left[\boldsymbol{\theta}_j - \boldsymbol{\alpha} - (\mathbf{s}_j - \mathbf{s}_{\text{obs}})^T \boldsymbol{\beta}\right]^2 \mathcal{K}_\epsilon\left(||\mathbf{s}_j - \mathbf{s}_{\text{obs}}||\right). \tag{E.2}$$

The Epanechnikov kernel, $\mathcal{K}_\epsilon(\cdot)$, depends on the maximum accepted distance based on the chosen ϵ, and ensures that the regression model is applied locally. The samples are then adjusted as

$$\tilde{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_j - (\mathbf{s}_j - \mathbf{s}_{\text{obs}})^T \hat{\boldsymbol{\beta}}, \quad j = 1, \ldots, M_\epsilon, \tag{E.3}$$

Algorithm 7 Regression ABC algorithm

Input: Parameter values $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ and corresponding simulated statistics $(\mathbf{s}_1, \ldots, \mathbf{s}_M)$, observed statistics $\mathbf{s}_{\text{obs}}$, number of accepted samples M_ϵ,

Accept $(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_{M_\epsilon}^*) \sim \{\boldsymbol{\theta}_i\}_{i=1}^M$ with the smallest $\|\mathbf{s}_i - \mathbf{s}_{\text{obs}}\|$

Solve optimisation problem (E.2) with $\left\{\theta_j^*\right\}_{j=1}^{M_\epsilon}$ and corresponding $\left\{\mathbf{s}_j^*\right\}_{j=1}^{M_\epsilon}$ to get $\hat{\boldsymbol{\beta}}$

Adjust accepted samples $\left\{\boldsymbol{\theta}_j^*\right\}_{j=1}^{M_\epsilon}$ using (E.3) to get $\left\{\tilde{\boldsymbol{\theta}}_j\right\}_{j=1}^{M_\epsilon}$

Output: Samples $(\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon})$ from approximate posterior

thus improving the approximation to the posterior distribution. Note that the adjustment is applied on each entry of the parameter vector independently. This regression-ABC algorithm is described in Alg. 7.

The adjustment in (E.3) is done disregarding the prior range of the parameters. Therefore, the regression method may adjust the samples to fall outside the support of the prior. This issue can be addressed by transforming the parameters before adjustment [31]. A log transformation can be used for positive parameters, and a logit transformation for parameters with bounded priors [35].

Algorithm 8 PMC-ABC with regression adjustment

Input: Prior $p(\boldsymbol{\theta})$, model $\mathcal{M}(\boldsymbol{\theta})$, observed statistics $\mathbf{s}_{\text{obs}}$, M_ϵ, M, T

Initialization: $t = 1$,

for $i = 1$ to M **do**
Sample $\boldsymbol{\theta}_i^{(1)} \sim p(\boldsymbol{\theta})$
Simulate $\mathbf{Y}_i^{(1)} \sim \mathcal{M}(\boldsymbol{\theta}_i^{(1)})$ and compute $\mathbf{s}_i^{(1)} = S(\mathbf{Y}_i^{(1)})$
end for
Perform regression adjustment by applying **Algorithm** 7 on $\left\{\left(\mathbf{s}_i^{(1)}, \boldsymbol{\theta}_i^{(1)}\right)\right\}_{i=1}^{M}$ to obtain $\left\{\tilde{\boldsymbol{\theta}}_j^{(1)}\right\}_{j=1}^{M_\epsilon}$
Set weights

$$\mathbf{w}_j^{(1)} = 1/M_\epsilon, \quad j = 1, \ldots, M_\epsilon, \quad \text{and variance}$$

$$\boldsymbol{\sigma}_{(1)}^2 = 2\widehat{\text{Var}}\left(\left\{\tilde{\boldsymbol{\theta}}_j^{(1)}\right\}_{j=1}^{M_\epsilon}\right)$$

for $t = 2$ to T **do**

for $i = 1, \ldots, M$ **do**
Sample $\boldsymbol{\theta}_i^* \sim \left\{\tilde{\boldsymbol{\theta}}_j^{(t-1)}\right\}_{j=1}^{M_\epsilon}$ with probabilities $\mathbf{w}_j^{(t-1)}$
Generate $\boldsymbol{\theta}_i^{(t)} \sim \mathcal{K}_t\left(\boldsymbol{\theta}|\boldsymbol{\theta}_i^*; \boldsymbol{\sigma}_{(t-1)}^2\right)$
Simulate $\mathbf{Y}_i^{(t)} \sim \mathcal{M}\left(\boldsymbol{\theta}_i^{(t)}\right)$ and compute $\mathbf{s}_i^{(t)} = S\left(\mathbf{Y}_i^{(t)}\right)$
end for
Perform regression adjustment by applying **Algorithm** 7 on $\left\{\left(\mathbf{s}_i^{(t)}, \boldsymbol{\theta}_i^{(t)}\right)\right\}_{i=1}^{M}$ to obtain $\left\{\tilde{\boldsymbol{\theta}}_j^{(t)}\right\}_{j=1}^{M_\epsilon}$
Set weights

$$\mathbf{w}_j^{(t)} \propto \frac{p\left(\boldsymbol{\theta}_j^{(t)}\right)}{\sum_{j=1}^{M_\epsilon} \mathbf{w}_j^{(t-1)} \mathcal{K}_t\left(\boldsymbol{\theta}_j^{(t)}|\tilde{\boldsymbol{\theta}}_i^{(t-1)}; \boldsymbol{\sigma}_{(t-1)}^2\right)}, \quad 1 \leq j \leq M_\epsilon$$

$$\text{and variance} \quad \boldsymbol{\sigma}_{(t)}^2 = 2\widehat{\text{Var}}\left(\left\{\tilde{\boldsymbol{\theta}}_j^{(t)}\right\}_{j=1}^{M_\epsilon}\right)$$

end for

Output: Samples $\left(\tilde{\boldsymbol{\theta}}_1^{(T)}, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon}^{(T)}\right)$ from the approximate posterior

Population Monte Carlo ABC

In cases where the parameter vector is high-dimensional, rejection-ABC method needs a large number of simulations of the model to reasonably explore the parameter space. Therefore, more advanced ABC methods have been introduced that rely on Markov chain Monte Carlo and Sequential Monte Carlo techniques that sample the parameter space efficiently. One such sequential technique is the Population Monte Carlo (PMC)-ABC method [34] that iteratively converges towards the approximate posterior distribution.

In the initialization of PMC-ABC, M_ϵ closest parameter samples out of $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ are retained, similar to rejection-ABC. These accepted samples, $\{\boldsymbol{\theta}_j\}_{j=1}^{M_\epsilon}$, form an approximation to the posterior distribution. A new population of M parameter samples is then drawn from the density kernel

$$\varphi_t(\boldsymbol{\theta}) = \sum_{j=1}^{M_\epsilon} \mathbf{w}_j^{(t-1)} \mathcal{K}_t \left(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}_j^{(t-1)}; \boldsymbol{\sigma}^2_{(t-1)} \right), \tag{E.4}$$

where t is the iteration index, $\mathbf{w}_j^{(t-1)}$ is the importance sampling weight associated with the accepted sample $\boldsymbol{\theta}_j^{(t-1)}$, and $\boldsymbol{\sigma}^2_{(t-1)}$ is a variance vector with each entry associated with a kernel $\mathcal{K}_t$. Note that $\mathbf{w}_j$ and $\boldsymbol{\sigma}^2$ are vectors of the same dimension as $\boldsymbol{\theta}$, and the new population for each parameter is drawn independently from the kernel with the corresponding variance. Typical choice for $\mathcal{K}_t$ is a Gaussian kernel, although other distributions may also be useful. A good choice for the variance of $\mathcal{K}_t$ is shown to be twice the empirical variance of the accepted samples [34]. Data and statistics are again simulated from the newly generated population, and the M_ϵ closest parameter samples are accepted and assigned weights $\mathbf{w}^{(t)} \propto p(\boldsymbol{\theta})/\varphi_t(\boldsymbol{\theta})$, where the division is taken entry-wise. This sequence of steps is repeated for T iterations, till the approximate posterior distributions converge.

To improve the posterior approximation and speed up the convergence, we combine the aforementioned two methods by applying the regression adjustment step on the accepted parameters after each iteration of PMC-ABC. The sample mean of the accepted samples after T^{th} iteration, $\left(\tilde{\boldsymbol{\theta}}_1^{(T)}, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon}^{(T)}\right)$, gives the approximate MMSE estimate.

2.2 Deep Learning

Given a set of summary statistics and corresponding model parameters, $\{(\mathbf{s}_i, \boldsymbol{\theta}_i)\}_{i \in \{1,\cdots,M\}}$, the calibration problem described above can be expressed as a mapping from $\mathbf{s}$ to $\boldsymbol{\theta}$. Denoting f as the mapping function, the model parameters can be expressed as

$$\boldsymbol{\theta}_i = f(\mathbf{s}_i) + \varepsilon_i \quad i = 1, \cdots, M, \tag{E.5}$$

where ε_i denotes the approximation error. Given the expression in (E.5), the calibration problem is equivalent to finding a representation of f such that ε_i

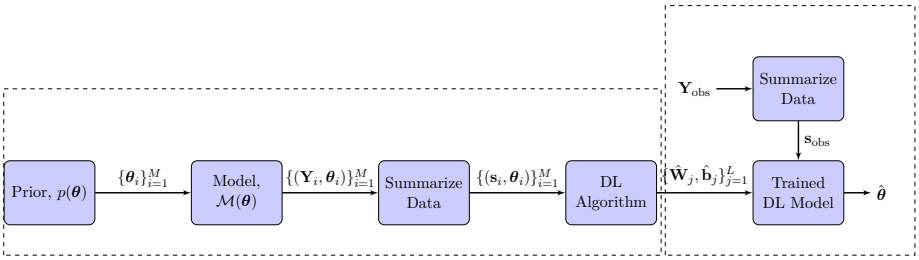


Fig. E.3: Block diagram for the proposed DL method. The left and right dashed boxes represent the training and the test phase, respectively.

is minimized for all $\mathbf{s}_i$. The function f can, e.g., be defined as a linear or polynomial function of $\mathbf{s}$ which can be obtained via a least square fit to the simulated data, $\{\mathbf{s}_i\}_{i=1}^M$. It may however be difficult to find such functions for the multi-dimensional and potentially complex relationship between the statistics and the model parameters. We, therefore, propose using a deep neural network architecture which has been shown to exhibit universal approximation property [26].

The DL based calibration method is illustrated in Fig. E.3. We approximate f using a deep NN architecture illustrated in Fig. E.4. The deep NN model can be defined using a hypothesis, $\hat{f}(\cdot; \boldsymbol{\Phi})$, with parameter $\boldsymbol{\Phi}$. For a network with R hidden layers, E_r neurons in the the r^{th} hidden layer, and p neurons in the output layer (since $\boldsymbol{\theta} \in \mathbb{R}^p$), the DL hypothesis for the i^{th} statistic can be expressed as [26]

$$\begin{aligned} \mathbf{h}_r(\mathbf{s}_i) &= a_h(\mathbf{W}_r \mathbf{h}_{r-1}(\mathbf{s}_i) + \mathbf{b}_r), \quad r = 1, \cdots, R \\ \hat{f}(\mathbf{s}_i; \boldsymbol{\Phi}) &= a_{\text{out}}(\mathbf{W}_{\text{out}} \mathbf{h}_R(\mathbf{s}_i) + \mathbf{b}_{\text{out}}) \end{aligned} \tag{E.6}$$

where $\mathbf{h}_r(\cdot)$ is the output of the r^{th} layer, $\mathbf{W}_r \in \mathbb{R}^{E_r \times E_{r-1}}$ and $\mathbf{W}_{\text{out}} \in \mathbb{R}^{p \times E_R}$ denote the weights matrix for connections terminating at the r^{th} hidden layer and output layer, respectively, with $\mathbf{b}_r \in \mathbb{R}^{E_r}$ and $\mathbf{b}_{\text{out}} \in \mathbb{R}^p$ being the corresponding bias terms. The activation function at the nodes of each intermediate layer is a_h, and that at the output layer is a_{out}. Note that $\mathbf{h}_0(\mathbf{s}_i) = \mathbf{s}_i$, with the number of neurons in the input layer being the dimensionality of $\mathbf{s}$. The DL hypothesis in (E.6) is parameterized by the set $\boldsymbol{\Phi} = \{\{(\mathbf{W}_r, \mathbf{b}_r)\}_{r=1}^R, (\mathbf{W}_{\text{out}}, \mathbf{b}_{\text{out}})\}$. The network parameters are estimated by training the network using the simulated data-set, $\{(\mathbf{s}_i, \boldsymbol{\theta}_i)\}_{i \in \{1, \cdots, M\}}$, from the model. Typically, the training is done by minimizing a loss function, $\mathcal{L}(\boldsymbol{\Phi})$, defined as

$$\mathcal{L}(\boldsymbol{\Phi}) = \frac{1}{2M} \sum_{i=1}^{M} ||f(\boldsymbol{\Phi}, \mathbf{s}_i) - \boldsymbol{\theta}_i||_2. \tag{E.7}$$

Minimization of the loss function is typically performed via stochastic gradient

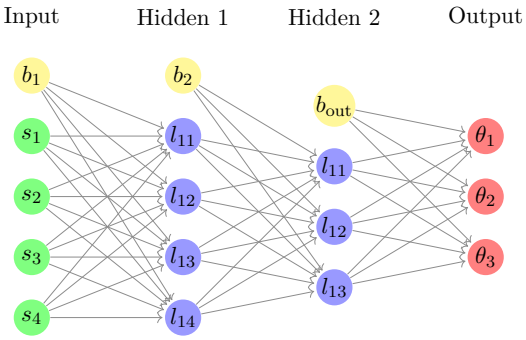


Fig. E.4: Description of a general deep neural network architecture with two hidden layers.

descent with back propagation of the error gradients viz:

$$\mathbf{\Phi}_n = \mathbf{\Phi}_{n-1} - \zeta \nabla \mathcal{L}(\mathbf{\Phi}_{n-1}), \tag{E.8}$$

where ζ denotes the step-size (also referred to as the learning rate) and ∇ is the gradient operator. Due to its fast convergence and good generalization for small data-sets, we utilize the Levenberg – Marquardt algorithm [36, 37] for network training in this paper. The network parameters are thus updated as

$$\mathbf{\Phi}_n = \mathbf{\Phi}_{n-1} - \left[\mathbf{J}^T\mathbf{J} + \lambda\mathbf{I}\right]^{-1} \mathbf{J}^T \mathcal{L}(\mathbf{\Phi}_{n-1}), \tag{E.9}$$

where $\mathbf{J}$ is the Jacobian matrix with $J_{m\ell} = \delta(f(\mathbf{\Phi}, \mathbf{s}_m) - \boldsymbol{\theta}_m)/\delta\Phi_\ell$, $\mathbf{I}$ is an identity matrix and λ is the adaptive damping factor. The damping factor controls the learning rate and is increased by λ_{inc} or decreased by λ_{dec} with increasing or decreasing error, respectively. This procedure is repeated until a termination criterion is achieved. Implementation of such a NN architecture can be achieved through standard toolboxes available in MATLAB, R, or Python.

Denoting the trained network parameters as $\hat{\mathbf{\Phi}}$, the calibration is done by applying the trained network on $\mathbf{s}_{\text{obs}}$ as

$$\hat{\boldsymbol{\theta}} = \hat{f}(\mathbf{s}_{\text{obs}}; \hat{\mathbf{\Phi}}). \tag{E.10}$$

The accuracy of the estimated model parameter, $\hat{\boldsymbol{\theta}}$, is affected by how well the trained hypothesis approximates the relationship between the summary statistics and the model parameters. This is dependent upon a number of factors such as the selected network structure, activation functions, and training method. Therefore, adequate care has to be taken in selecting the NN model in order to obtain reasonable estimates.

3 Calibration of Polarimetric Propagation Graph Model

The proposed ABC and DL methods are applied for estimating parameters of a polarimetric propagation graph (PG) model [30, 38, 39]. First proposed in [40], the PG offers a simple and efficient approach for modeling propagation channels that account for both specular and dense multipath components. The model also has the ability to capture reverberation effects. Several studies have applied or modified the PG model in indoor [41], outdoor-to-indoor [42], high-speed railway [43, 44], indoor-to-indoor [45, 46] and millimetre-wave systems [47, 48]. Despite the growing interest in the PG model, study on its calibration based on measurements is severely limited. A vast majority of works utilizing the PG model are either based on the stochastic generation procedure in [49] or in combination with a map of the environment. The polarimetric PG model in [30] was calibrated with measurements using method of moments. However, the method requires manually fixing one parameter due to identifiability issues. Moreover, the measurement noise variance is not estimated, necessitating manual truncation of the power delay profile (PDP) prior to fitting.

3.1 Model Description

Consider a time-invariant radio channel in a multi-input, multi-output (MIMO) set-up with N_t and N_r output ports at the transmit and receive antennas, respectively. In the PG framework , the radio channel is modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ [49]. The vertex set $\mathcal{V} = \mathcal{V}_t \cup \mathcal{V}_r \cup \mathcal{V}_s$ is a union of a set $\mathcal{V}_t$ of N_t transmitters, a set $\mathcal{V}_r$ of N_r receivers, and a set $\mathcal{V}_s$ of N_s scatterers in the environment. The edges $\mathcal{E} = \mathcal{E}_{\mathrm{d}} \cup \mathcal{E}_{\mathrm{t}} \cup \mathcal{E}_{\mathrm{s}} \cup \mathcal{E}_{\mathrm{r}}$ model the wave propagation between the vertices, where $\mathcal{E}_{\mathrm{d}}$ is a set of direct edges, $\mathcal{E}_{\mathrm{t}}$ is a set of transmitter to scatterer edges, $\mathcal{E}_{\mathrm{s}}$ is a set of scatterer to scatterer edges and $\mathcal{E}_{\mathrm{r}}$ is a set of scatterer to receiver edges.

To each vertex v we associate a position $\mathbf{r}_v \in \mathbb{R}^3$. From these positions, the length of an edge (v, w) is $||\mathbf{r}_v - \mathbf{r}_w||$. This results in a propagation delay from v to w of $\tau_e = \|(\mathbf{r}_w - \mathbf{r}_v)\|/c$, where c is the speed of light in vacuum and $\|\cdot\|$ denotes the Euclidean norm. Accordingly, the direction of propagation is specified by a unit vector Ω_e associated with edge e, pointing in the direction of propagation.

The transfer function matrix at a particular frequency, $\mathbf{H}(f)$, of the polarimetric PG is given as

$$\mathbf{H}(f) = \mathbf{D}(f) + \mathbf{R}(f)[\mathbf{I} - \mathbf{B}(f)]^{-1}\mathbf{T}(f), \tag{E.11}$$

where $\mathbf{D}(f) \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ is the transmitter to receiver, $\mathbf{T}(f) \in \mathbb{C}^{2N_{\mathrm{s}} \times N_{\mathrm{t}}}$ is the transmitter to scatterer, $\mathbf{R}(f) \in \mathbb{C}^{N_{\mathrm{r}} \times 2N_{\mathrm{s}}}$ is the scatterer to receiver, and

$\mathbf{B}(f) \in \mathbb{C}^{2N_\mathrm{s}\times 2N_\mathrm{s}}$ is the scatterer to scatterer edge transfer function sub-matrix. Then, the transfer function sub-matrices are given as:

$$\begin{aligned}
\mathbf{D}(f) &= \mathcal{X}_\mathrm{t}^T(\Omega_e)\mathcal{X}_\mathrm{r}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{d}\\
\mathbf{T}(f) &= \mathcal{X}_\mathrm{t}^T(\Omega_e)\mathbf{M}\mathbf{\Gamma}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{t}\\
\mathbf{B}(f) &= \mathbf{M}\mathbf{\Gamma}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{s}\\
\mathbf{R}(f) &= \mathcal{X}_\mathrm{r}(\Omega_e)G_e(f), \quad e \in \mathcal{E}_\mathrm{r}
\end{aligned}$$

Here, $\mathcal{X}_\mathrm{t}(\Omega_e)$ and $\mathcal{X}_\mathrm{r}(\Omega_e)$ are the 2×1 transmit and receive polarimetric antenna array response vectors, respectively, and $\mathbf{\Gamma}(\Omega_e)$ is the 2×2 rotation matrix. The 2×2 scattering matrix, $\mathbf{M}$, represents the coupling between the two polarization states. Assuming it is equal for all the scatterers, $\mathbf{M}$ reads

$$\mathbf{M} = \frac{1}{1+\gamma}\begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}, \tag{E.12}$$

where $\gamma \in (0,1)$ is the polarization power coupling ratio. Finally, $G_e(f)$ is the scalar that captures polarization-independent propagation characteristics, and is expressed as

$$G_e(f) = g_e(f)\exp[j(\psi_e - 2\pi\tau_e f)], \tag{E.13}$$

where ψ_e is the phase. The edge gain, $g_e(f)$ is calculated as:

$$g_e(f) = \begin{cases} \frac{1}{(4\pi f\tau_e)}; & e \in \mathcal{E}_\mathrm{d} \\ \frac{1}{\sqrt{4\pi\tau_e^2 f\mu(\mathcal{E}_\mathrm{t})\mathcal{S}(\mathcal{E}_\mathrm{t})}}; & e \in \mathcal{E}_\mathrm{t} \\ \frac{g}{\mathrm{odi}(e)}; & e \in \mathcal{E}_\mathrm{s} \\ \frac{1}{\sqrt{4\pi\tau_e^2 f\mu(\mathcal{E}_\mathrm{r})\mathcal{S}(\mathcal{E}_\mathrm{r})}}; & e \in \mathcal{E}_\mathrm{r} \end{cases} \tag{E.14}$$

Here, $g \in (0,1)$ is the reflection gain, $\mathrm{odi}(e)$ denotes the number of outgoing edges from the n^{th} scatterer, and

$$\mu(\mathcal{E}_\mathrm{a}) = \frac{1}{|\mathcal{E}_\mathrm{a}|}\sum_{e\subset\mathcal{E}_\mathrm{a}}\tau_e, \quad \mathcal{S}(\mathcal{E}_\mathrm{a}) = \sum_{e\subset\mathcal{E}_\mathrm{a}}\tau_e^{-2}, \quad \mathcal{E}_\mathrm{a} \subset \mathcal{E},$$

with $|\cdot|$ denoting set cardinality.

To draw a random graph and simulate transfer function from the model, the positions of the transmit and receive antennas need to be specified. An edge between $\mathcal{V}_\mathrm{t}$ and $\mathcal{V}_\mathrm{r}$ is drawn with probability P_dir. Note that for line-of-sight case, $P_\mathrm{dir} = 1$, while for non-line-of-sight (NLOS) case $P_\mathrm{dir} = 0$. Edges between $\mathcal{V}_\mathrm{t}$ and $\mathcal{V}_\mathrm{s}$, $\mathcal{V}_\mathrm{s}$ and $\mathcal{V}_\mathrm{s}$, or $\mathcal{V}_\mathrm{s}$ and $\mathcal{V}_\mathrm{r}$ are drawn with probability P_vis. The phase ψ_e is drawn uniformly between 0 and 2π.

3.2 Calibration problem formulation

To calibrate the PG model based on measured data, $\mathbf{Y}_{\text{obs}}$, we need to estimate the parameters of the model such that the model fits the data. We consider measurements conducted in NLOS conditions, resulting in $P_{\text{dir}} = 0$. Apart from the model parameters, we would also like to estimate the noise variance. The parameter vector to be estimated from $\mathbf{Y}_{\text{obs}}$ thus becomes $\boldsymbol{\theta} = [g, N_s, P_{\text{vis}}, \gamma, \sigma_N^2]^T$.

3.3 Measurement data description

Let the MIMO channel transfer function be measured at K equidistant points in the bandwidth B, giving a frequency separation of $\Delta f = B/(K-1)$. The resulting measured signal matrix at each frequency point, $\mathbf{Y}_k \in \mathbb{C}^{N_r \times N_t}$, reads

$$\mathbf{Y}_k = \mathbf{H}(f_k) + \mathscr{N}_k, \quad k = 0, 1, \cdots, K-1 \tag{E.15}$$

where $\mathbf{H}(f_k)$ is the sampled transfer matrix, and $\mathscr{N}_k$ is the measurement noise. Assuming independent and identically distributed (iid) noise at each measurement point and for each transmitter-receiver link, we model it as iid zero-mean complex Gaussian variables with variance σ_N^2. The entire polarized observed data-set, denoted as $\mathbf{Y}_{\text{obs}}$, thus becomes an $N_r \times N_t \times K$ matrix. Let Y_k^{ij} be the measurement for the k^{th} frequency sample between the i^{th} receiver and the j^{th} transmitter. The received signal in time-domain, $y^{\text{ij}}(t)$, is computed as

$$y^{\text{ij}}(t) = \frac{1}{K} \sum_{k=0}^{K-1} Y_k^{\text{ij}} \exp(j2\pi k \Delta f t). \tag{E.16}$$

3.4 Summary statistics

Implementation of the ABC and the DL learning method necessitates a choice of appropriate summary statistics of observed data that are informative about the model parameters. The first three temporal moments, computed as

$$m_l^{\text{ij}} = \int_0^{1/\Delta f} t^l |y^{\text{ij}}(t)|^2, \quad l = 0, 1, 2, \tag{E.17}$$

have been used previously for calibration of stochastic radio channel models and found to be informative about model parameters [16–19]. Here, we compute the sample mean of the l^{th} temporal moment as

$$\bar{m}_l = \frac{1}{N_r N_t} \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} m_l^{\text{ij}}, \tag{E.18}$$

and the sample covariance between l^{th} and l'^{th} temporal moment as

$$\widehat{\text{cov}}\left(m_l^{\text{ij}}, m_{l'}^{\text{ij}}\right) = \frac{1}{N_r N_t - 1} \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} \left(m_l^{\text{ij}} - \bar{m}_l\right)\left(m_{l'}^{\text{ij}} - \bar{m}_{l'}\right). \quad \text{(E.19)}$$

Additionally, we separate the temporal moments according to their polarization, i.e. vertical-vertical (vv), vertical-horizontal (vh), horizontal-vertical (hv), and horizontal-horizontal (hh), to compute the cross-polarization ratio, XPR, as

$$\text{XPR} = \frac{1}{2}\left[\frac{\bar{m}_0^{vv}}{\bar{m}_0^{vh}} + \frac{\bar{m}_0^{hh}}{\bar{m}_0^{hv}}\right]. \quad \text{(E.20)}$$

The summary statistics vector, $\mathbf{s}$, therefore has ten entries: the XPR, the three means $\bar{m}_l$, and, and six covariances $\widehat{\text{cov}}\left(m_l^{\text{ij}}, m_{l'}^{\text{ij}}\right)$ for $l, l' = 0, 1, 2$.

To verify that the chosen summary statistics are informative about the model parameters, we conduct a simulation experiment. One parameter at a time is sampled 100 times from its uniform prior distribution (given in Tab. E.1), while the other parameters are held fixed. Data is simulated from the polarimetric PG model for such a parameter vector and the statistics are computed. Each of the ten statistics are then plotted versus the five parameters in Fig. E.5.

We observe that XPR and the means of the temporal moments are informative about almost all the parameters. The covariances become informative for higher values of g and P_{vis} and lower values of γ. We see a clear functional relationship devoid of any jitter between XPR and γ, and $\bar{m}_2$ and σ_N^2. This indicates from the outset that γ and σ_N^2 should be estimated very accurately. In contrast, the summaries seem the least informative about g, suggesting that the estimate of g would be the most uncertain. In principle, a subset of these statistics could also be used in the calibration methods. However, we observed a degradation in performance on leaving out the covariances, and therefore include all ten statistics.

4 Performance Evaluation

We apply the ABC and DL methods to calibrate the polarimetric PG model using the summary statistics described in the previous section. First, we evaluate the performance of the two calibration methods via simulations, and later validate it using millimetre-wave NLOS measurements from [50]. The measurements were taken in a room of dimensions $3 \times 4 \times 3$ m^3 in the bandwidth range of 58 GHz to 62 GHz, sampled at $K = 801$ equidistant points. The frequency separation of Δf =5 MHz results in a signal observation interval of 200 ns in the time domain. A 5×5 virtual planar array of dual polarized

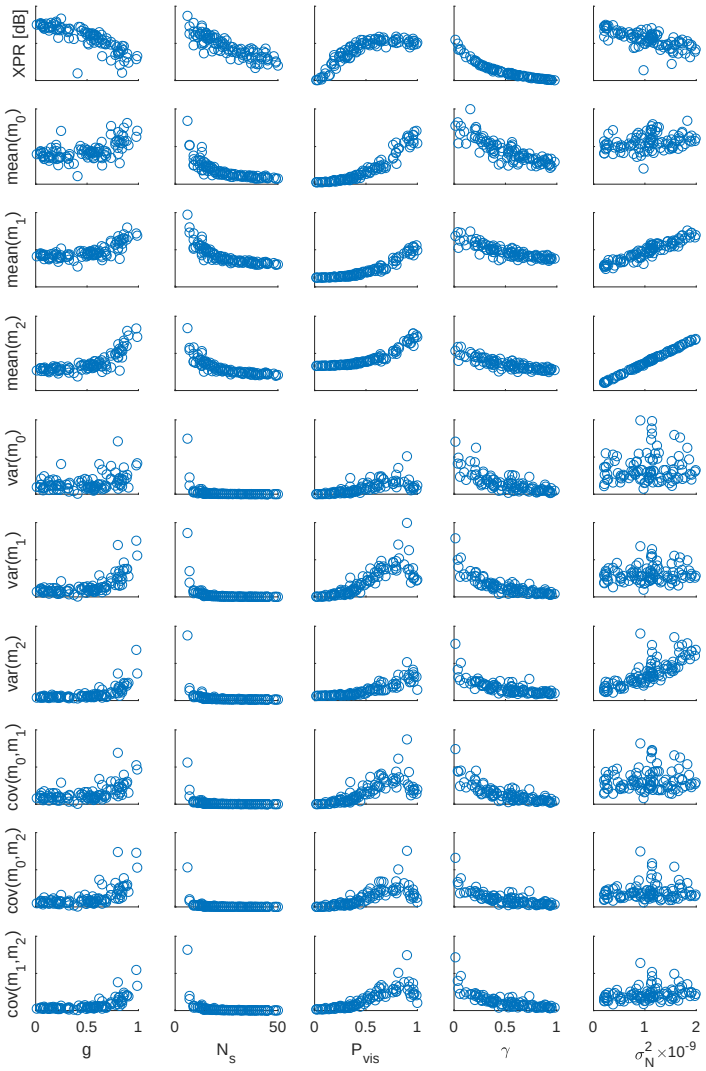


Fig. E.5: Summary statistics versus model parameters. Each plot is generated by varying one parameter while the others are held fixed to the values in Tab. E.1.

antennas with 5 mm inter-element spacing, was used at both the receiver and the transmitter. This gives $N_t = N_r = 50$.

For the simulation experiment, we set the parameters of the model to some "true" value, say $\boldsymbol{\theta}_{\text{true}}$, and generate data from the model that we consider as observed data. We then apply the proposed methods on this simulated data to estimate the parameters. We use the same settings for simulations as in the measurements [50]. The antennas in the PG model implementation are assumed to be omni-directional with perfect cross-polar isolation at both the receiver and the transmitter. The scatterers are distributed uniformly across the floor of the room.

For the simulation experiment, the observed statistics, $\mathbf{s}_{\text{obs}}$, that corresponds to $\boldsymbol{\theta}_{\text{true}}$ needs to be set. For a fixed $\boldsymbol{\theta}_{\text{true}}$, the stochastic model generates samples $\left(\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(Z)}\right)$ from $p\left(\mathbf{s}|\boldsymbol{\theta}_{\text{true}}\right)$. Running the estimator Z times with each realization of the statistics vector as $\mathbf{s}_{\text{obs}}$ results in Z parameter estimates, giving the error distribution around $\boldsymbol{\theta}_{\text{true}}$. However, as shown in Appendix 6, this is equivalent to taking $\mathbf{s}_{\text{obs}}$ as the sample mean of $\left(\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(Z)}\right)$ and running the estimator once. Here, we adopt the latter computationally convenient approach and compute $\mathbf{s}_{\text{obs}}$ from $\boldsymbol{\theta}_{\text{true}}$ using $Z = 200$. It should be noted that a similar approach is only possible in general for measured data if Z independent measurements can be obtained.

4.1 Approximate Bayesian Computation

The ABC method is applied with $M = 2000$ samples simulated for each iteration. For the first iteration, the samples are taken from a uniform prior distribution of the parameters with ranges given in Tab. E.1. Note that the prior for N_s is uniform integers in the specified range. A Gaussian kernel truncated to the prior range is used to generate populations of subsequent iterations. We set the tolerance threshold to $\epsilon = 5\%$, giving $M_\epsilon = 100$ accepted samples, and the total number of iterations to $T = 10$. The summary statistics are normalized by the estimate of their median absolute deviation before applying ABC. A logit transformation is applied to the parameters before regression adjustment to keep the adjusted samples within the prior boundary. The estimated marginal posterior distributions are shown in Fig. E.6 and Fig. E.7 for simulated and measured data, respectively. The obtained point estimates and the sample standard deviation of the accepted samples after $T = 10$ iterations are reported in Tab. E.1.

We observe in Fig. E.6 that the samples obtained from approximate posteriors lie around the true value for all the parameters, and that the MMSE estimates are fairly accurate even after the first iteration. As the iterations go on, the posteriors shrink and converge for each parameter, albeit some faster than others. For example, the posteriors for N_s, P_{vis}, γ and σ_N^2 barely change after the second iteration, while that of g seem to converge after around five iterations with the MMSE estimate getting better with further iterations. Similarly, the posterior for g is the widest, while that of the other parameters are

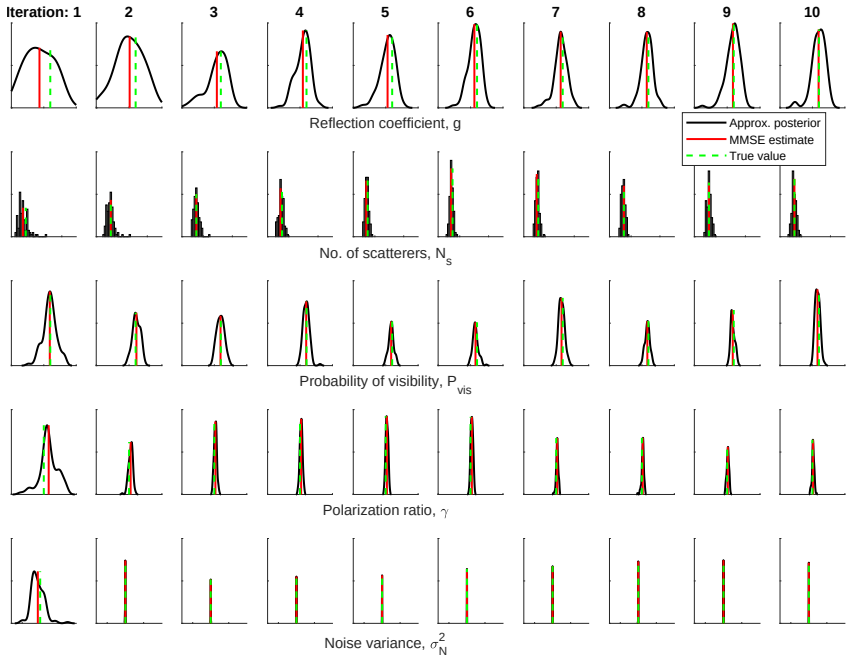


Fig. E.6: Kernel density estimates of the marginal approximate posteriors of the parameters obtained by applying ABC on simulated data, plotted in the prior range for each parameter. The posteriors are shown after each iteration of the algorithm, with the parameter estimate marked in red. The true value of the parameter is shown in green.

quite narrow. This uncertainty in the estimates of different parameters reflects their relationship with the summary statistics shown in Fig. E.5. Parameters that have a distinct functional relationship with, at least, a few statistics are easier to estimate than others.

Similar behavior is observed in Fig. E.7 for measurements as was seen for the simulation experiment. The approximate posteriors for P_{vis}, γ, and σ_N^2 are very narrow, and seem to have converged since the first iteration. However, the posteriors for g and N_s takes approximately four iterations to converge. The width of posterior for g and P_{vis} is narrower in comparison to those in simulation. This is attributed to the fact that for a high value of P_{vis}, as is the case in measurements, almost all the statistics become informative, see Fig. E.5. The averaged power delay profile (APDP) generated from the model using the point estimates obtained for the measurements is shown in Fig. E.11. The calibrated model fits both the co- and cross-polarized APDP of the measurements well.

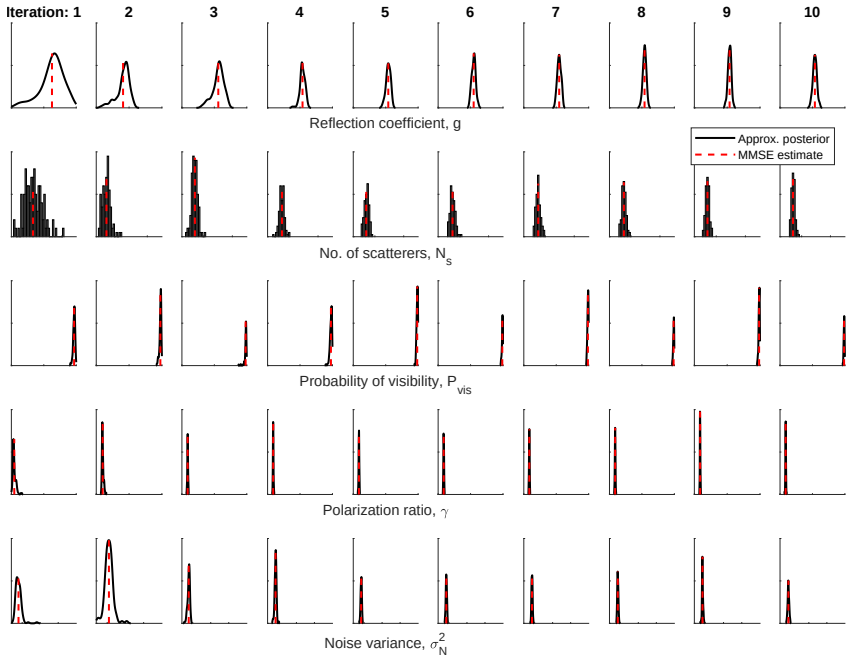


Fig. E.7: Kernel density estimates of the marginal approximate posteriors of the parameters obtained by applying ABC on measured data. The posteriors are shown after each iteration of the algorithm. The density is plotted in the prior range for each parameter. The sample mean is marked in red.

4.2 Deep Learning

We determine the structure of the NN for the calibration problem via a guided search procedure. First, we limit the number of hidden layers to $R = 2$ and assume that there are equal number of neurons in each layer, i.e. $E_r = E$. A single hidden layer architecture is excluded due to its poor performance during our preliminary experiments. The number of neurons in each layer is varied from 2 to 28. We then divide the data set into two equal subsets for training and cross-validation. The mean and standard deviation of squared error on the training and validation subsets are shown in Fig. E.8. We observe that the error stabilizes after around 12 neurons, and so we set $E = 20$ as this is sufficient. This results in a 10-20-20-5 network architecture which is used for evaluating the DL calibration method. We use the hyperbolic tangent sigmoid and linear activation functions [26] for the hidden and output layers, respectively.

The NN is trained using the same $M = 2000$ samples of $\mathbf{s}_i$ and corresponding $\boldsymbol{\theta}_i$ as used in the initialization of the ABC method. In principle, the training data could be as extensive as possible, thus leading to a better approximation of the summary-parameter function. In order to eliminate the sensitivity of

Table E.1: Summary of the parameter estimates obtained from ABC and DL for simulated and measured data. Note that the sample standard deviation reported in parenthesis is of the approximate posterior for ABC and of the estimate distribution for DL.

Parameter θ	Prior range $p(\theta)$	Estimate (standard deviation)					
		Simulated Data			Measured Data		
		True value	ABC	DL	ABC	DL	MoM [30]
Reflection gain, g	$[0, 1]$	0.6	0.59 (0.11)	0.58 (0.04)	0.54 (0.03)	0.56 (0.05)	0.64
Number of scatterers, N_s	$[5, 50]$	15	15 (1.56)	15 (0.92)	14 (1.25)	16 (6.5)	11
Probability of visibility, P_{vis}	$[0, 1]$	0.6	0.58 (0.04)	0.60 (0.014)	0.99 (0.006)	0.96 (0.03)	0.9[a]
Polarization ratio, γ	$[0, 1]$	0.5	0.51 (0.02)	0.50 (0.01)	0.09 (0.005)	0.09 (0.05)	0.06
Noise variance, σ_N^2	$[2 \times 10^{-10}, 2 \times 10^{-9}]$	10^{-9}	9.98×10^{-10} (2.86×10^{-12})	9.96×10^{-10} (8.8×10^{-12})	4.3×10^{-10} (2.11×10^{-11})	4.3×10^{-10} (2.8×10^{-11})	-

[a] Note that this value is not estimated but set in [30].

network to the range of values in the summary statistics, the entire data-set is normalized using the standard Z-score scaling prior to network training. The data-set is randomly partitioned into training, test, and validation subsets in the ratio 0.70, 0.15 and 0.15, respectively. We utilized the LM algorithm with damping parameters: $\lambda = 0.1$, $\lambda_{\text{inc}} = 10$ and $\lambda_{\text{dec}} = 0.1$ for training the NN. The training procedure is terminated when the number of epochs reaches 1000 or the gradient is below 10^{-7}.

Once the termination criterion is achieved, we apply the trained network on $\mathbf{s}_{\text{obs}}$ from simulated and measured data to get point estimate of the parameter vector. This process is repeated 200 times to estimate the distribution of the parameter estimates, which is shown in Fig. E.9 and Fig. E.10 for simulated and measured data, respectively. The sample mean of the estimates and their standard deviations are reported in Tab. E.1. We observe in Fig. E.9 that the DL method is able to estimate the model parameters accurately and with reasonable precision. The uncertainty in the estimates is fairly small. As was the case with the ABC method, the estimate of g has the largest standard deviation out of all the parameters, corroborating our conjecture based on Fig. E.5. The method performs similarly on measured data, as seen in Fig. E.10, although with slightly larger standard deviations. The APDP generated from the parameter estimates fits the measurements, see Fig. E.11, thus validating the methodology.

5 Discussion

In our example, we see that the calibration approach based on summaries is effective. However, as experienced during the development of the algorithms, the choice of summaries is important for obtaining a good calibration. Ideally, sufficient statistics should be considered, but such are rarely available or practical to extract in the context of radio channel models. Although the simulation method of checking summaries has proved useful for the example problem considered here, it only gives an indication of how informative the summaries are,

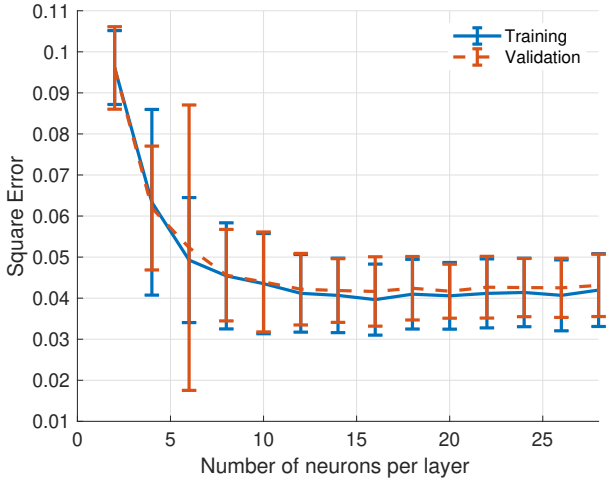


Fig. E.8: Learning curve for selecting NN size. The mean and standard deviation of the squared error at each number of neurons per layer is computed from 200 repeated network training.

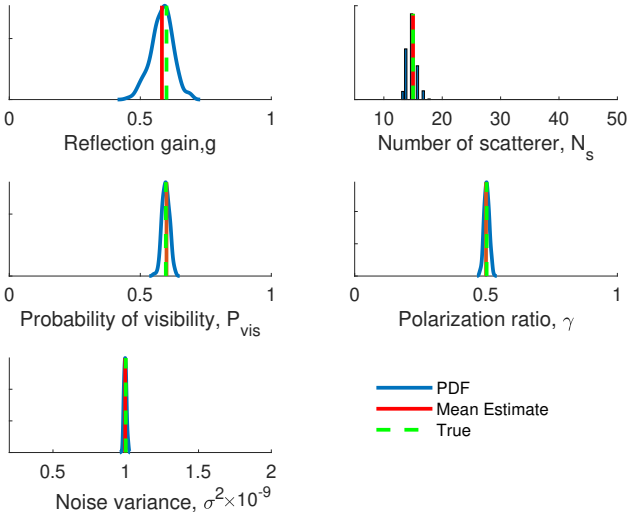


Fig. E.9: Parameter estimates obtained by applying the DL method on simulated data. Kernel density estimates of the distribution obtained after 200 estimator runs.

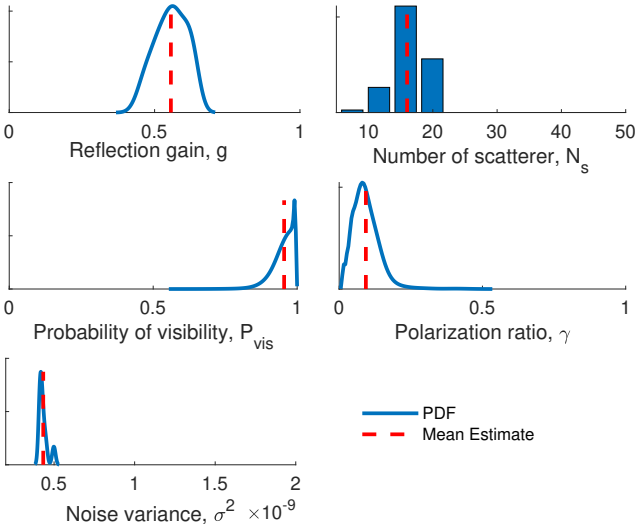


Fig. E.10: Parameter estimates obtained by applying the DL method on measured data. Kernel density estimates of the distribution obtained after 200 estimator runs.

and does not guarantee that the algorithms will work. The summary statistics used in this paper appear to be informative about the different aspects of the channel, and could be useful in calibrating other stochastic channel models as well. In the case of calibrating a directional model, the summaries possibly have to be chosen differently. The method for checking the summaries described here would be useful for this selection.

Although based on the summary statistics of the data, the two proposed methods are complementary to each other as they approach the same problem in distinct ways. To highlight this fact, we have intentionally avoided comparing the two methods with each other. ABC infers on the parameters in a Bayesian sense by learning the distribution of the parameters given the data. On the other hand, the DL method fits a function between the summaries and the parameters and provides point estimates of the parameters given the data in a frequentist manner without considering any prior information. Despite their differences, both methods are able to learn the parameters of the model such that, qualitatively, it is not apparent in Fig. E.11 which APDP fits the measurements better.

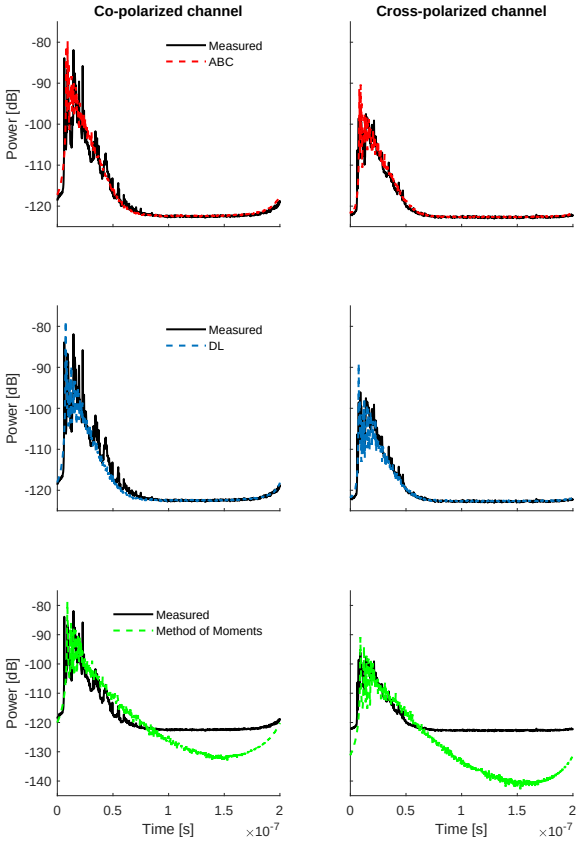


Fig. E.11: Averaged power delay profile from the measurements versus those obtained from the polarimetric PG model after calibration using ABC, DL and the method of moments approach from [30]. The parameter estimates for all three methods are reported in Tab. E.1.

The choice of prior naturally affects the posterior distribution of the ABC method. In this paper, we used uninformative (flat) priors in order to rely solely on the data to estimate the parameters. The ranges of these priors are chosen conservatively, i.e., to be very wide. Choosing more informative priors instead would lead to faster convergence of the approximate posterior in the ABC method, thus reducing run-time. In general, if the data is large enough, the effect of the prior distribution on the posterior becomes irrelevant.

Since the ABC method relies on simulations from the model, the computational complexity primarily depends on the complexity of the model. The computationally expensive step in the DL method is the training of the NN, which depends on the size of the training data and the chosen NN architecture. While the two layer architecture was found to be sufficient for the example

model in the paper, more complex models may require deeper networks. However, the network needs to be trained just once, and parameter estimates can be obtained for different measurement data instantly. In contrast, the sequential nature of the ABC algorithm requires running the iterations again for new observed statistics.

The computation time for the proposed methods depends on the particular implementation and the available hardware. The methods are lightweight enough to be run on standard laptops with reasonable run-time. As an indication of the required run-time of the proposed methods, our implementation of the proposed ABC method was able to complete ten iterations within a day on a Lenovo ThinkPad with Intel Core i7 processor having 24 GB RAM. Training the NN took less than 2 hours on a Lenovo ThinkPad with Intel Core i7 processor and 16 GB RAM. Due to the different choices of high-resolution and clustering algorithms that are available, their comparison with the proposed methods in terms of computational complexity is not feasible. We remark that in all the cases, the specific run-time depends not only on the choice of algorithms and hardware, but is significantly impacted by the particular implementation and choice of settings in the respective algorithms.

We experience that the ABC method is not very sensitive to the particular settings of the algorithm due to its iterative nature. Increasing the number of simulated samples per iteration, M, increases simulation time from the model, but leads to lower ϵ if M_ϵ is kept constant. Thus, the algorithm would converge in fewer iterations. Similarly, increasing ϵ would mean accepting samples that are further away from $\mathbf{s}_{\text{obs}}$, and therefore, would require more iterations to converge to a stable approximate posterior. Overall, changes in one setting is compensated by another, and the method performs similarly. This means that the performance primarily relies on the choice of statistics. In contrast, for the DL method the particular NN architecture should be chosen carefully since generalization accuracy is sensitive to the network size, particularly with small number of training samples.

In principle, the entire channel impulse response measurements could be used as input data instead of a set of summary statistics. However, the ABC method is then hit by the curse of dimensionality [35]. That is, the distances become very large due to the high-dimensional data, thus increasing the rejection rate significantly. For the DL method, this would increase the number of input layers, thereby increasing the computational complexity, along with complicating the training procedure.

An appealing feature of the summary-based approach is that it is specified exactly which criteria are used to fit the model to the data. In this sense, we obtain the best fitting model in the eyes of the summary statistics. This is very different from what is obtained by the multi-step methods relying on ad-hocery and possibly conflicting assumptions in the individual steps, e.g. assuming "well-separated" paths in multipath extraction, followed by application of clustering algorithms. Statistical techniques to construct a more informative subset of summaries [32] to be used in ABC can be explored. However, by do-

ing so we lose the transparency as to which summaries are being fitted. Both the proposed machine learning methods are also simpler to implement than the classical state-of-the-art approach of multipath extraction, with fewer settings of the algorithm. Moreover, since the proposed methods are integrated, their performance is easy to investigate by simulation studies. This is again a great advantage compared to multi-step approaches where the performance of each step is evaluated separately, thus making it difficult to judge the accuracy of the overall parameter estimates.

Additionally, the ABC and DL methods are able to estimate all model parameters, including the noise variance. Thus it is not necessary to provide side information to the algorithm (as is often done by separately estimating e.g. noise variance) or to post-process the data by setting the noise threshold. This advantage is clearly seen when comparing with the method of moments (MoM) approach of calibrating polarimetric PG models [30] where noise is not estimated (see Fig. E.11).

6 Conclusion

The proposed machine learning methods based on ABC and deep learning are able to accurately calibrate stochastic radio channel models. The model fit is obtained in light of the explicitly chosen summaries. The proposed methods demonstrate that stochastic channel models, in particular the PG model, can be calibrated without access to likelihoods. The methods also by-pass any intermediate step of extracting the multipath components. We observed that the choice of summaries is crucial in learning the parameters, and the uncertainty in parameter estimates decreases with informative summaries. The summaries used in this paper are general purpose, and we conjecture that they can work for other models as well. The methodology to qualify the summaries through simulation study is useful in the design of the algorithm, although it does not provide any guarantees. Availability of pseudocodes and libraries make the proposed methods easy to implement, compared to the state-of-the-art approach. The performance of the proposed methods is easy to evaluate, as opposed to the multi-step approach. Moreover, no additional information or post-processing is required to calibrate the model.

Appendix

We want to show that setting $\mathbf{s}_{\text{obs}}$ as one realization out of $p(\mathbf{s}|\boldsymbol{\theta}_{\text{true}})$ and running the estimator Z times gives the same estimate of $\boldsymbol{\theta}_{\text{true}}$ in mean as taking $\mathbf{s}_{\text{obs}}$ as the sample mean of Z such realizations out of $p(\mathbf{s}|\boldsymbol{\theta}_{\text{true}})$ and running the estimator once.

Let $\mathbf{s}_{\text{obs}}$ be the z^{th} sample, $\mathbf{s}^{(z)}$, from the distribution $p(\mathbf{s}|\boldsymbol{\theta}_{\text{true}})$, and the corresponding MMSE estimate be $\hat{\boldsymbol{\theta}}^{(z)}$. Then, the sample mean for Z such

estimates is

$$\hat{\boldsymbol{\theta}}_{\text{avg}} = \frac{1}{Z}\sum_{z=1}^{Z}\hat{\boldsymbol{\theta}}^{(z)} \approx \mathbb{E}[\boldsymbol{\theta}|\mathbf{s} = \mathbb{E}\left[\mathbf{s}|\boldsymbol{\theta}_{\text{true}}\right]]\,. \tag{E.21}$$

The mean of $\hat{\boldsymbol{\theta}}^{(z)}$ then reads

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}}^{(z)}\right] = \int \hat{\boldsymbol{\theta}}^{(z)} p\left(\mathbf{s}|\boldsymbol{\theta}_{\text{true}}\right) \mathrm{d}\mathbf{s} = \boldsymbol{\theta}_{\text{true}}. \tag{E.22}$$

Assuming $\left(\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(Z)}\right)$ are independent samples, the expected value of $\hat{\boldsymbol{\theta}}_{\text{avg}}$ can be computed as:

$$\begin{aligned}
&\mathbb{E}\left[\hat{\boldsymbol{\theta}}_{\text{avg}}\right] \\
&\quad= \int \cdots \int \frac{1}{Z}\sum_{z=1}^{Z}\hat{\boldsymbol{\theta}}^{(z)} \Pi_{z'=1}^{Z} p\left(\mathbf{s}^{(z')}|\boldsymbol{\theta}_{\text{true}}\right) \mathrm{d}\mathbf{s}^{(1)} \ldots \mathrm{d}\mathbf{s}^{(Z)} \\
&\quad= \frac{1}{Z}\sum_{z=1}^{Z}\int \hat{\boldsymbol{\theta}}^{(z)} p\left(\mathbf{s}^{(z)}|\boldsymbol{\theta}_{\text{true}}\right) \mathrm{d}\mathbf{s}^{(z)} \\
&\quad= \frac{1}{Z}\sum_{z=1}^{Z}\boldsymbol{\theta}_{\text{true}} = \boldsymbol{\theta}_{\text{true}}
\end{aligned}$$

Therefore, both $\hat{\boldsymbol{\theta}}^{(z)}$ and $\hat{\boldsymbol{\theta}}_{\text{avg}}$ converge to $\boldsymbol{\theta}_{\text{true}}$ in mean and thus, are unbiased estimates. The variance of $\hat{\boldsymbol{\theta}}_{\text{avg}}$ is, however, reduced by a factor of $1/Z$.

References

[1] L. Greenstein, S. Ghassemzadeh, S.-C. Hong, and V. Tarokh, "Comparison study of UWB indoor channel models," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 1, pp. 128–135, Jan 2007.

[2] 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. 21, no. 1, pp. 1–9, Feb 1972.

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

[4] 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.

[5] 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.

[6] 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 *Eur. Conf. on Antennas and Propag.*, 2011, pp. 3606–3610.

[7] P. Kyösti, "Winner II channel models, deliverables d1.1.2 v1.2, part I: Channel models," Tech. Rep. IST-4-027756 WINNER II Project, 2007.

[8] L. Raschkowski, P. Kyösti, K. Kusume, and E. T. Jämsä, "METIS channel models, deliverable d1.4 v3," Tech. Rep. ICT-317669 METIS Project, 2015.

[9] 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.

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

[11] X. Yin, T. Pedersen, N. Czink, and B. H. Fleury, "Parametric characterization and estimation of bi-azimuth and delay dispersion of individual path components," in *2006 First European Conference on Antennas and Propagation*. IEEE, nov 2006.

[12] X. Yin, G. Steinbock, G. E. Kirkelund, T. Pedersen, P. Blattnig, A. Jaquier, and B. H. Fleury, "Tracking of time-variant radio propagation paths using particle filtering," in *2008 IEEE International Conference on Communications*. IEEE, 2008.

[13] G. Steinbock, T. Pedersen, X. Yin, and B. H. Fleury, "Experimental characteristics of indoor wideband MIMO radio channels and their impact on stochastic modelling," in *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*. IEEE, jan 2009.

[14] A. Karttunen, C. Gustafson, A. F. Molisch, J. Jarvelainen, and K. Haneda, "Censored multipath component cross-polarization ratio modeling," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2016.

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

[16] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter Estimation for Stochastic Channel Models using Temporal Moments," in *Proc. 2019 IEEE Int. Symp. on Antennas and Propag. and USNC-URSI Radio Sci. Meeting*, 2019.

[17] ——, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

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

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

[20] M. A. Beaumont, "Approximate bayesian computation," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 379–403, 2019. [Online]. Available: https://doi.org/10.1146/annurev-statistics-030718-105212

[21] ——, "Approximate bayesian computation in evolution and ecology," *Annu. Rev. Ecol. Evol. Syst.*, vol. 41, no. 1, pp. 379–406, dec 2010.

[22] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, "Population growth of human y chromosomes: a study of y chromosome microsatellites," *Molecular Biology and Evolution*, vol. 16, no. 12, pp. 1791–1798, dec 1999.

[23] O. François, M. G. B. Blum, M. Jakobsson, and N. A. Rosenberg, "Demographic history of european populations of arabidopsis thaliana," *PLoS Genetics*, vol. 4, no. 5, May 2008.

[24] J. Akeret, A. Refregier, A. Amara, S. Seehars, and C. Hasner, "Approximate bayesian computation for forward modeling in cosmology," *Journal of Cosmology and Astroparticle Physics*, vol. 2015, no. 08, pp. 043–043, aug 2015. [Online]. Available: https://doi.org/10.1088%2F1475-7516%2F2015%2F08%2F043

[25] A. B. Abdessalem, N. Dervilis, D. Wagg, and K. Worden, "Model selection and parameter estimation in structural dynamics using approximate bayesian computation," *Mechanical Systems and Signal Processing*, vol. 99, pp. 306–325, jan 2018.

[26] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[27] C. Huang, G. C. Alexandropoulos, A. Zappone, C. Yuen, and M. Debbah, "Deep learning for ul/dl channel calibration in generic massive mimo systems," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.

[28] J. Lee, M. Y. Kang, and S. Kim, "Path loss exponent prediction for outdoor millimeter wave channels through deep learning," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2019, pp. 1–5.

[29] S. Sheehan and Y. S. Song, "Deep learning for population genetic inference," *PLOS Computational Biology*, vol. 12, no. 3, pp. 1–28, 03 2016. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1004845

[30] R. Adeogun, T. Pedersen, C. Gustafson, and F. Tufvesson, "Polarimetric Wireless Indoor Channel Modelling Based on Propagation Graph," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 10, pp. 6585–6595, 2019.

[31] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002. [Online]. Available: https://www.genetics.org/content/162/4/2025

[32] M. G. B. Blum, M. A. Nunes, D. Prangle, and S. A. Sisson, "A comparative review of dimension reduction methods in approximate bayesian computation," *Stat. Sci.*, vol. 28, no. 2, pp. 189–208, may 2013.

[33] D. Prangle, "Adapting the ABC distance function," *Bayesian Analysis*, vol. 12, no. 1, pp. 289–309, mar 2017.

[34] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate bayesian computation," *Biometrika*, vol. 96, no. 4, pp. 983–990, Oct 2009.

[35] M. G. B. Blum, "Approximate bayesian computation: A nonparametric perspective," *J. of the Amer. Stat. Assoc.*, vol. 105, no. 491, pp. 1178–1187, sep 2010.

[36] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *The Quarterly of Applied Mathematics*, no. 2, pp. 164–168, 1944.

[37] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, Nov 1994.

[38] R. Adeogun and T. Pedersen, "Propagation graph based model for multi-polarized wireless channels," in *IEEE WCNC*, April 2018.

[39] ——, "Modelling polarimetric power delay spectrum for indoor wireless channels via propagation graph formalism," in *2nd URSI Atlantic Radio Sci. Meeting*, May 2018.

[40] T. Pedersen and B. H. Fleury, "Radio channel modelling using stochastic propagation graphs," in *IEEE ICC*, June 2007, pp. 2733–2738.

[41] G. Steinböck, M. Gan, P. Meissner, E. Leitinger, K. Witrisal, T. Zemen, and T. Pedersen, "Hybrid model for reverberant indoor radio channels using rays and graphs," vol. 64, no. 9, pp. 4036–4048, Sept 2016.

[42] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of outdoor-to-indoor radio channels via propagation graphs," in *URSI General Assembly and Scientific Symposium*, Aug 2014, pp. 1–4.

[43] L. Tian, X. Yin, Q. Zuo, J. Zhou, Z. Zhong, and S. X. Lu, "Channel modeling based on random propagation graphs for high speed railway scenarios," in *IEEE PIMRC*, Sept 2012, pp. 1746–1750.

[44] T. Zhou, C. Tao, S. Salous, Z. Tan, L. Liu, and L. Tian, "Graph-based stochastic model for high-speed railway cutting scenarios," *IET Microwaves, Antennas Propagation*, vol. 9, no. 15, pp. 1691–1697, 2015.

[45] R. O. Adeogun, A. Bharti, and T. Pedersen, "An iterative transfer matrix computation method for propagation graphs in multi-room environments," *IEEE Antennas and Wireless Propag. Lett.*, vol. 18, no. 4, pp. 616–620, April 2019.

[46] R. Adeogun, T. Pedersen, and A. Bharti, "Transfer function computation for complex indoor channels using propagation graphs," *IEEE PIMRC*, Sept. 2018.

[47] J. Chen, X. Yin, L. Tian, and M. Kim, "Millimeter-wave channel modeling based on a unified propagation graph theory," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 246–249, Feb 2017.

[48] L. Tian, V. Degli-Esposti, E. M. Vitucci, and X. Yin, "Semi-deterministic radio channel modeling based on graph theory and ray-tracing," *IEEE Trans. on Antennas and Propag.*, vol. 64, no. 6, pp. 2475–2486, June 2016.

[49] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of reverberant radio channels using propagation graphs," vol. 60, no. 12, pp. 5978–5988, Dec 2012.

[50] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

Paper F

Auto-Generated Summaries for Stochastic Radio Channel Models

Ayush Bharti, Ramoni Adeogun, Troels Pedersen

The paper has been published in the
European Conference on Antennas and Propagation, 2021.

Abstract

Recently, a calibration method has been proposed for estimating the parameters of stochastic radio channel models using summaries of channel impulse response measurements without multipath extraction. In this paper, we attempt to automatically generate summaries using an autoencoder for calibration of channel models. This approach avoids the need for explicitly designing informative summaries about the model parameters, which can be tedious. We test the method by calibrating the stochastic polarized propagation graph model on simulated as well as measured data. The autoencoder is found to generate summaries that give reasonably accurate results while calibrating the considered model.

1 Introduction

Stochastic radio channel models are widely used for simulating the channel in order to design and test communication and localization systems. To ensure that such models yield accurate simulations, they must be calibrated. Most often, this is done by estimating the model parameters from measurement data. Unfortunately, the likelihood function for stochastic channel models are usually intractable. Thus, calibrating them from new measurements becomes challenging. Hence, it is standard practice to employ high-resolution path extraction algorithms to estimate the delays, gains, etc. of the multipath components. These estimates are further used to estimate the parameters of the channel model. This methodology has been followed to calibrate channel models from the early days of Turin [1] and Saleh-Valenzuela [2] to the more recent ones [3–5]. However, implementing such complicated algorithms is not trivial and requires a number of heuristic choices to be made. Such choices affect the accuracy of the results, and the overall performance of the estimator is difficult to assess.

Recently, calibration methods which circumvent the need for resolving the multipath components have been introduced [6–10]. In [6, 7], a method of moments approach is used for calibration. However, these methods rely on analytical expressions for the moments which may not be available for more complicated stochastic models. More general calibration methods based on summary statistics of the channel measurements have been proposed for the Saleh-Valenzuela model in [8, 9] and for the propagation graph model [11] in [10]. The methods proposed in [8, 10] are based on approximate Bayesian computation (ABC), which is a likelihood-free inference framework that relies on simulations from the model to infer on the model parameters [12]. ABC involves comparing summary statistics of the simulated and the measured data in some distance metric. The parameter samples that yield simulated summaries “close" to the measured summaries are accepted. These accepted parameters form a sample from the approximate posterior distribution. Thus, such methods rely on handcrafted summaries of the measurement data which should be

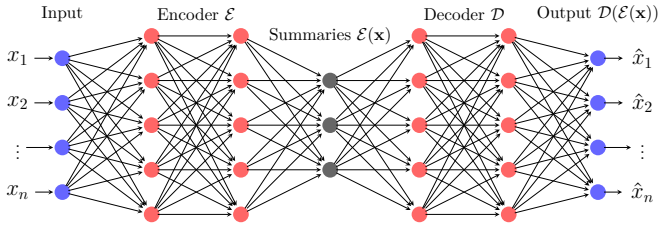


Fig. F.1: Illustration of a deep autoencoder with 2 hidden layers in both the encoder and decoder.

informative about the parameters of the stochastic channel models. However, designing such summaries is a time-consuming process which is not always straightforward for most models.

In this paper, we attempt to automatically generate the summaries by using an autoencoder [13]. An autoencoder is a combination of two neural networks; one encodes the input data-set into a low-dimensional set of features, and the other decodes those features in order to replicate the input data. We use the encoded feature vector as summary statistics to calibrate the polarized propagation graph model [11, 14] using the ABC algorithm [10]. This approach circumvents the need for manual design of summaries which can be a time-consuming process. We test the method by calibrating the stochastic polarized propagation graph model [11, 14] and comparing to previously obtained results for the algorithm [10] with handcrafted summaries. We find that the method requires much less effort than handcrafted summaries, while obtaining a comparable performance.

2 ABC using Autoencoder

We aim to fit a stochastic radio channel model, $\mathcal{M}(\boldsymbol{\theta})$, to a set of measurement data, $\mathbf{x} \in \mathbb{R}^n$. This amounts to estimating the $p-$dimensional parameter vector, $\boldsymbol{\theta}$ from the data. The ABC method in [10] allows us to do so provided simulations can be obtained from $\mathcal{M}(\boldsymbol{\theta})$. The ABC method relies on summarizing $\mathbf{x}$ into a low-dimensional vector of summaries, $\mathbf{s} = \mathcal{E}(\mathbf{x})$, informative about $\boldsymbol{\theta}$. In [8, 10], we relied on handcrafted summaries. Here, we circumvent designing the summarizing function, or encoder, $\mathcal{E}(\cdot)$ ourselves by using an autoencoder to automatically learn this function.

2.1 Generating Summaries using Autoencoder

As in Fig. F.1, a typical autoencoder comprises of two neural networks — an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$, and an intermediate layer often referred to as the latent space. The encoder converts high-dimensional input variables

into low-dimensional latent variables, $\mathbf{s} = \mathcal{E}(\mathbf{x}) \in \mathbb{R}^q; q \ll n$. The decoder $\mathcal{D}$ reconstructs the input from the latent variables. The goal is to obtain functions $\mathcal{E}$ and $\mathcal{D}$ such that the output, $\hat{\mathbf{x}}$, is close to the input in some metric, i.e.,

$$\hat{\mathbf{x}} = \mathcal{D}(\mathcal{E}(\mathbf{x})) \approx \mathbf{x}. \tag{F.1}$$

For the considered autoencoder architecture in Fig. F.1, the encoding function reads

$$\mathbf{s} = \mathcal{E}(\mathbf{x}) = \eta_2(\mathbf{W}_2(\eta_1(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2)), \tag{F.2}$$

where $\eta_i, \mathbf{W}_i$ and $\mathbf{b}_i$ denote the activation function, the weights, and the biases of the i^{th} hidden layer of the encoder, respectively. The decoding function, $\mathcal{D}$, is defined analogous to (F.2). The weights and biases are obtained by minimizing a reconstruction loss defined as the mean squared error between $\mathbf{x}$ and $\hat{\mathbf{x}}$. Standard packages for performing such optimization exist in languages such as MATLAB, R and Python.

2.2 Approximate Bayesian Computation method

The summarizing function $\mathcal{E}$ obtained by the autoencoder is now used in the ABC algorithm proposed in [10] to approximate the posterior distribution, $p(\boldsymbol{\theta}|\mathbf{s}_{\text{obs}})$, where $\mathbf{s}_{\text{obs}}$ is the summary vector of the measurements. The algorithm is stated in Alg. 10 and illustrated in Fig. F.2. Here, we give an overview of the ABC algorithm; see [10] for further details.

The ABC method proceeds by sampling $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$ independently from the prior distribution $p(\boldsymbol{\theta})$, and simulating the corresponding summaries $\mathbf{s}_1, \ldots, \mathbf{s}_M$ using the model and the summarizing function. The Euclidean distance between the simulated and the observed statistics, $\|\mathbf{s}_i - \mathbf{s}_{\text{obs}}\|$, are then computed. Note that the summaries are normalized using their mean absolute deviation before computing the distance. The first M_ϵ parameter samples that correspond to the smallest distance are accepted, along with their corresponding summary vectors. This results in an acceptance ratio of $\epsilon = M_\epsilon/M$. The accepted samples are then adjusted using local-linear regression [15] to improve the posterior approximation. Given the accepted set $\{(\mathbf{s}_i, \boldsymbol{\theta}_i)\}_{i=1}^{M_\epsilon}$, the i^{th} accepted parameter sample is adjusted as

$$\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i - (\mathbf{s}_i - \mathbf{s}_{\text{obs}})^T \hat{\boldsymbol{\beta}}, \quad i = 1, \ldots, M_\epsilon, \tag{F.3}$$

where $\hat{\boldsymbol{\beta}}$ is solution to the optimization problem

$$\arg\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{i=1}^{M_\epsilon} \left[\boldsymbol{\theta}_i - \boldsymbol{\alpha} - (\mathbf{s}_i - \mathbf{s}_{\text{obs}})^T \boldsymbol{\beta}\right]^2 K_\epsilon\left(\|\mathbf{s}_i - \mathbf{s}_{\text{obs}}\|\right). \tag{F.4}$$

Here, $K_\epsilon(\cdot)$ is the Epanechnikov kernel. The ABC method of [10] then draws a new set of M parameter samples from $(\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon})$ in a sequential Monte

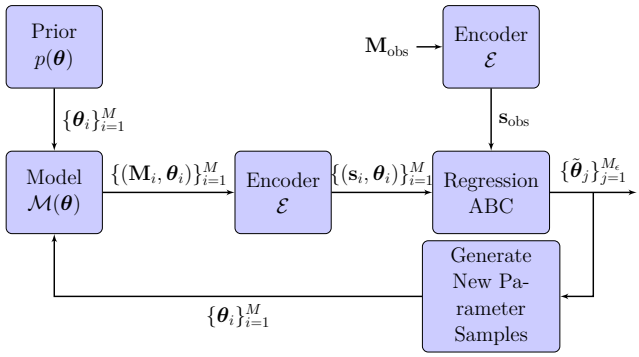


Fig. F.2: Block diagram of the data flow in the proposed PMC-ABC algorithm with regression adjustment.

Algorithm 9 Regression adjustment

Input: Parameter values $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ and corresponding simulated summaries $(\mathbf{s}_1, \ldots, \mathbf{s}_M)$, observed statistics $\mathbf{s}_{\text{obs}}$, number of accepted samples M_ϵ,

Accept $(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_{M_\epsilon}^*) \sim \{\boldsymbol{\theta}_i\}_{i=1}^{M}$ with the smallest $\|\mathbf{s}_i - \mathbf{s}_{\text{obs}}\|$

Solve optimisation problem (F.4) with $\left\{\theta_j^*\right\}_{j=1}^{M_\epsilon}$ and corresponding $\left\{\mathbf{s}_j^*\right\}_{j=1}^{M_\epsilon}$ to get $\hat{\boldsymbol{\beta}}$

Adjust accepted samples $\left\{\boldsymbol{\theta}_j^*\right\}_{j=1}^{M_\epsilon}$ using (F.3) to get $\left\{\tilde{\boldsymbol{\theta}}_j\right\}_{j=1}^{M_\epsilon}$

Output: Samples $(\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon})$ from approximate posterior

Carlo fashion [16]. These new samples form the prior distribution for the next iteration of the algorithm, where they are used to generate simulated data from the model again and perform regression adjustment. The idea is to iteratively converge towards the posterior distribution by sampling the parameter space efficiently. In iteration t, the parameter samples of the l^{th} parameter are drawn from the density kernel

$$\varphi_l^{(t)}(\cdot) = \sum_{j=1}^{M_\epsilon} w_{l,j}^{(t-1)} \mathcal{K}\left(\cdot | \tilde{\theta}_{l,j}^{(t-1)}; \sigma_{l,(t-1)}^2\right), \tag{F.5}$$

$l = 1, \ldots, p$, where $w_{l,j}^{(t-1)}$ and $\sigma_{l,(t-1)}^2$ are the importance sampling weight and the variance associated with $\tilde{\theta}_{l,j}^{(t-1)}$, respectively. Note that the univariate Gaussian kernel, $\mathcal{K}$, is truncated to be in the prior range. The adjusted parameter samples after N_{iter} iterations are then taken as samples from the approximate posterior distribution.

Algorithm 10 ABC method [10]

Input: Prior $p(\boldsymbol{\theta})$, model $\mathcal{M}(\boldsymbol{\theta})$, observed summaries $\mathbf{s}_{\text{obs}}$, M_ϵ, M, N_{iter}

Initialization: $t = 1$,

for $i = 1$ to M **do**
Sample $\boldsymbol{\theta}_i^{(1)} \sim p(\boldsymbol{\theta})$
Simulate $\mathbf{M}_i^{(1)} \sim \mathcal{M}(\boldsymbol{\theta}_i^{(1)})$ and compute $\mathbf{s}_i^{(1)} = \mathcal{E}(\mathbf{x}_i^{(1)})$
end for
Perform regression adjustment by applying **Algorithm** 9 on $\left\{\left(\mathbf{s}_i^{(1)}, \boldsymbol{\theta}_i^{(1)}\right)\right\}_{i=1}^{M}$ to obtain $\left\{\tilde{\boldsymbol{\theta}}_j^{(1)}\right\}_{j=1}^{M_\epsilon}$
Set

$$w_{l,j}^{(1)} = 1/M_\epsilon, \quad \text{and} \quad \sigma_{l,(1)}^2 = 2\widehat{\text{Var}}\left(\left\{\tilde{\theta}_{l,j}^{(1)}\right\}_{j=1}^{M_\epsilon}\right),$$
$$j = 1, \ldots, M_\epsilon, \quad l = 1, \ldots, p$$

for $t = 2$ to N_{iter} **do**

for $i = 1, \ldots, M$ **do**

for $l = 1, \ldots, p$ **do**
Sample $\theta_{l,i}^* \sim \left\{\tilde{\theta}_{l,j}^{(t-1)}\right\}_{j=1}^{M_\epsilon}$ with probabilities $w_{l,j}^{(t-1)}$
Generate $\theta_{l,i}^{(t)} \sim \mathcal{K}\left(\cdot|\theta_{l,i}^*; \sigma_{l,(t-1)}^2\right)$
end for
Simulate $\mathbf{M}_i^{(t)} \sim \mathcal{M}\left(\boldsymbol{\theta}_i^{(t)}\right)$ and compute $\mathbf{s}_i^{(t)} = \mathcal{E}\left(\mathbf{x}_i^{(t)}\right)$
end for
Perform regression adjustment by applying **Algorithm** 9 on $\left\{\left(\mathbf{s}_i^{(t)}, \boldsymbol{\theta}_i^{(t)}\right)\right\}_{i=1}^{M}$ to obtain $\left\{\tilde{\boldsymbol{\theta}}_j^{(t)}\right\}_{j=1}^{M_\epsilon}$
Set

$$w_{l,j}^{(t)} \propto \frac{p\left(\tilde{\theta}_{l,j}^{(t)}\right)}{\varphi_l^{(t)}(\tilde{\theta}_{l,j}^{(t)})}, \text{and} \quad \sigma_{l,(t)}^2 = 2\widehat{\text{Var}}\left(\left\{\tilde{\theta}_{l,j}^{(t)}\right\}_{j=1}^{M_\epsilon}\right),$$
$$j = 1, \ldots, M_\epsilon, \quad l = 1, \ldots, p$$

end for
Output: Samples $\left(\tilde{\boldsymbol{\theta}}_1^{(T)}, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon}^{(T)}\right)$ from the approximate posterior

3 Calibration of stochastic channel models

We apply the calibration method to calibrate the stochastic polarized propagation graph model (SPPGM) [11] in which the channel is represented as a

propagation graph [17] with the transmitters, the receivers, and the scatterers as vertices. Edges in the graph are defined randomly depending on the probability of visibility, P_{vis}. An edge transfer function accounting for depolarization effects, attenuation, delay and phase shifts is defined for each edge. To calculate the edge transfer functions, the SPPGM only requires three parameters viz: reflection coefficient g, number of scatterers N_{s} and the polarization ratio γ. The edge transfer functions are then used in a simple expression to compute the channel transfer function, H_k. Detailed description of the model and channel generation procedure can be found in [11].

We consider data from a linear, time-invariant radio channel, measured using a vector network analyzer (VNA) in the bandwidth B. The transfer function H_k is measured at K equidistant frequency points resulting in a frequency separation of $\Delta f = B/(K-1)$. The measured signal at each frequency point, Y_k, can be modeled as

$$Y_k = H_k + N_k, \quad k = 0, 1, \cdots, K-1, \tag{F.6}$$

where N_k denotes the measurement noise. The noise samples at each k are assumed independent and identically distributed (iid) as a circular complex Gaussian with variance σ_N^2. Discrete-frequency, continuous-time inverse Fourier transforming $(Y_0, \ldots, Y_{K-1})$ gives the measured signal in time-domain

$$y(t) = \frac{1}{K} \sum_{k=0}^{K-1} Y_k \exp(j2\pi k \Delta f t), \tag{F.7}$$

which is periodic with period $t_{\max} = 1/\Delta f$. Typically K is in order of hundreds or even thousands, and so we intend to summarize the high-dimensional measured signal into its first J temporal moments, defined as

$$m_j = \int_0^{t_{\max}} t^j |y(t)|^2, \quad j = 0, 1, 2, \ldots, (J-1). \tag{F.8}$$

The temporal moments are computed instantaneously per realization of $y(t)$. For the dual polarized channel from the SPPGM, this computation is performed for each of the four polarizations. Thus, the i^{th} realization yields a $4J$ dimensional vector $\mathbf{m}^{(i)}$. Consequently, a measurement with L independent polarimetric measurements yields an $L \times 4J$ matrix of temporal moments $\mathbf{M}_{\text{obs}} = [\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(L)}]^\top$. Including the noise variance as a parameter, calibration of the SPPGM therefore requires estimating the parameter vector $\boldsymbol{\theta} = [g, N_{\text{s}}, P_{\text{vis}}, \gamma, \sigma_N^2]^\top$.

3.1 Implementation and Training of the Autoencoder

We consider the first three temporal moments, $J = 3$, and $L = 625$ realizations. The training data is obtained from the SPPGM with 4000 parameter vectors generated uniformly over the prior ranges in Tab. F.1. For each parameter

vector, $\boldsymbol{\theta}_i$, we compute the temporal moments using (F.8) and convert the 625×12 matrix $\mathbf{M}_i$ into an input vector, $\mathbf{x}_i \in \mathbb{R}^{7500}$. To be consistent with [10], we set the number of summaries $q = 10$.

We adopt Python's popular machine learning libraries, *Keras* and *Tensorflow*, to design and implement the autoencoder. We use the Rectified Linear Unit activation function at the hidden layers of both the encoder and decoder due to its non-vanishing gradient property. A linear activation function is used at the output layer of the decoder. Based on our initial experiments, we observe that a network with two hidden layers in the encoder and decoder yields reasonably informative summaries. We adopt a mini-batch gradient descent procedure in which the training data is partitioned into equal batches of size 32. The training is performed using the Root Mean Square Propagation (RMSProp) algorithm with learning rate = 0.1 as optimizer.

3.2 Evaluation of summaries

We now use the trained encoder to test whether the auto-generated summary vector $\mathbf{s}$ is informative about the parameters of the SPPGM. We do that by varying one parameter at a time while setting others to a fixed value, and computing $\mathbf{s}$. If the summaries have a functional relationship with the parameters, then they are deemed informative. Each parameter is varied in its prior range, and the other parameters are fixed to true values given in Tab. F.1. The resulting plot for $q = 10$ summaries is shown in Fig. F.3. We see that the summaries are responsive to changes in each of the parameters, albeit more for some than others. The generated summaries are informative enough to be able to calibrate the SPPGM. However, this visual test gives only a picture of how well the summaries work separately. To evaluate how informative the joint summaries are about the parameters, we apply these in the ABC algorithm.

3.3 Application to simulated data

We apply Alg. 10 to calibrate the SPPGM from simulated data. We set the parameter vector to a true value and generate polarized channel measurements from the SPPGM which we then summarize using the encoder to get the summaries. We take an average over 200 realizations of such summaries to get $\mathbf{s}_{\text{obs}}$ in order to remove any bias in the estimate arising due to Monte Carlo approximation. The prior distributions were kept same as in [10] and are given in Tab. F.1, along with the approximate minimum mean squared error (MMSE) estimates. The approximate posterior distributions are shown in Fig. F.4. We observe that the marginal posteriors are concentrated around the true values, and that the algorithm seems to work. The width of the posteriors indicate how informative the summaries are about each parameter. For example, the fact that the posterior for g is the widest is due to its lack of any distinct relationship with statistics in Fig. F.3. Overall the method seems to work rea-

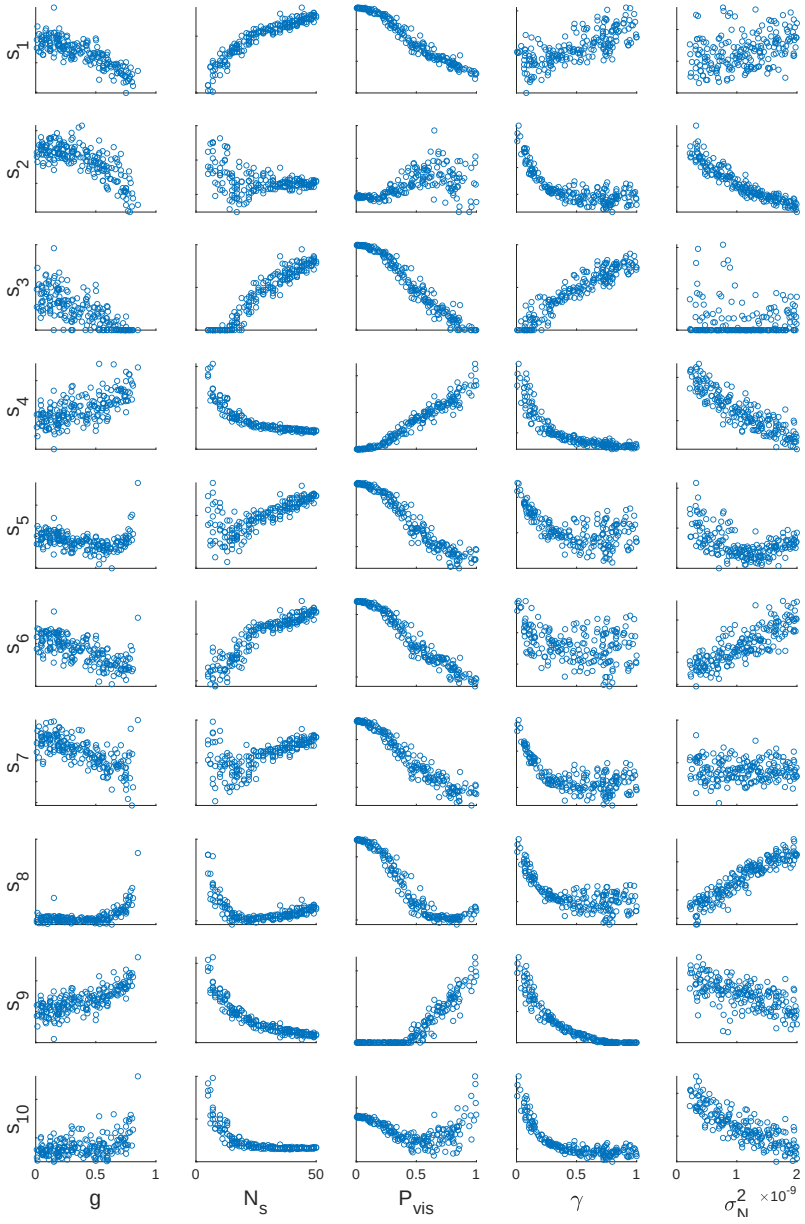


Fig. F.3: Auto-generated statistics versus the parameters of the SPPGM. Each plot is generated by varying one parameter while others are kept fixed to the true values in Tab. F.1.

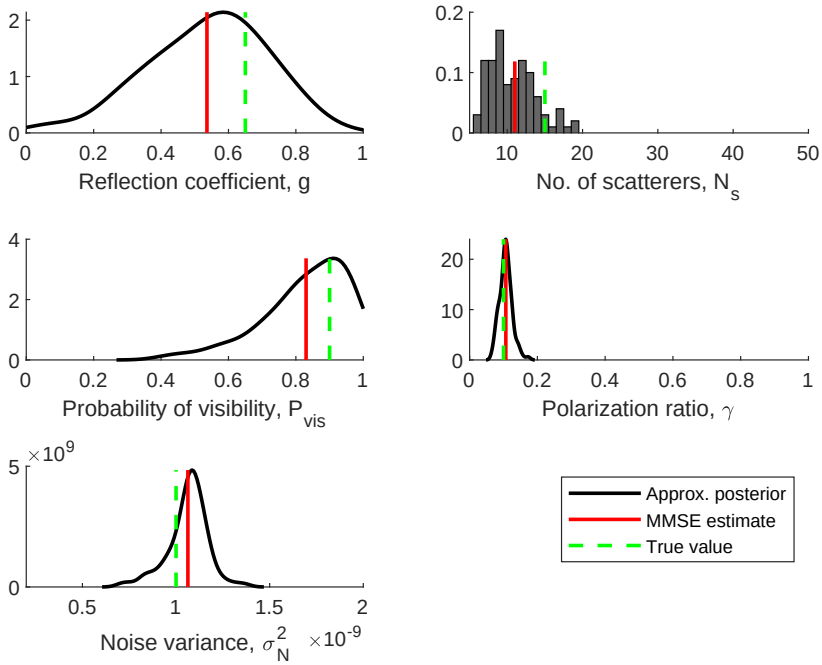


Fig. F.4: Kernel density estimates of the approximate marginal posteriors for simulated data after $N_{\text{iter}} = 10$ iteration. Settings: $L = 625, J = 3, M = 2000, M_\epsilon = 100$, $B = 4$ GHz, $K = 801, \Delta f = 5$ MHz, $t_{\max} = 200$ ns.

sonably well, considering that no manual effort went into designing the specific summaries.

4 Application to Measured Data

We now apply the method to calibrate the SPPGM using millimetre-wave polarized channel measurements from [18]. The measurements were conducted using dual-polarized antennas in a small conference room of dimensions $3 \times 4 \times 3$ m^3 in the frequency range 58 GHz to 62 GHz. A 5×5 virtual planar array was used at both the transmitter and receiver with 5 mm inter-element spacing, resulting in $L = 625$ realizations for each polarization.

The approximate marginal posteriors are given in Fig. F.5 while the approximate MMSE estimates, obtained by averaging the accepted samples after the last iteration, are reported in Tab. F.1. The width of the posterior estimates are similar to what we observed in simulation. The summaries therefore seem to be useful for calibration even in case of measurements. The averaged power delay profile (APDP) of the co- and cross-polarized channel from the measure-

Table F.1: Parameter estimates for simulated and measured data. The sample standard deviation of the approximate posterior is reported in the parenthesis.

Parameter $\boldsymbol{\theta}$	**Prior range** $p(\boldsymbol{\theta})$	**True value / Estimate (standard deviation)**	
		Simulated data	**Measured data**
Refl. coeff. g	[0,1]	0.65 / 0.54 (0.18)	— / 0.71 (0.15)
No. of scat. N_s	[5,50]	15 / 11 (3.08)	— / 16 (5.48)
Prob. of vis. P_{vis}	[0,1]	0.90 / 0.83 (0.13)	— / 0.70 (0.14)
Pol. ratio γ	[0,1]	0.10 / 0.11 (0.02)	— / 0.14 (0.02)
Noise variance $\sigma_N^2 \times 10^{-10}$	[2,20]	10.0 / 10.6 (1.07)	— / 3.75 (0.51)

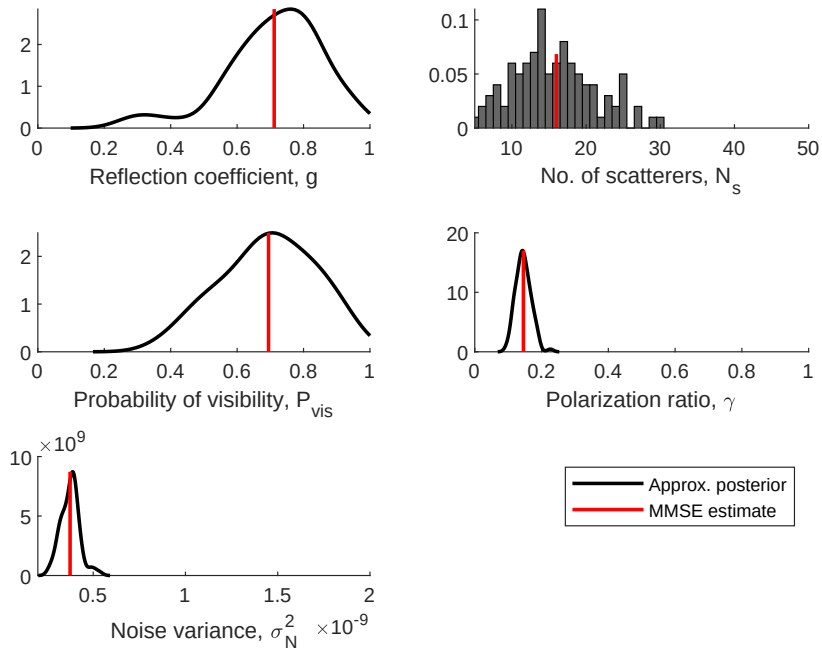


Fig. F.5: Kernel density estimates of the approximate marginal posteriors for measured data after $N_{\text{iter}} = 10$ iteration. The algorithm settings are same as given in Fig. F.4.

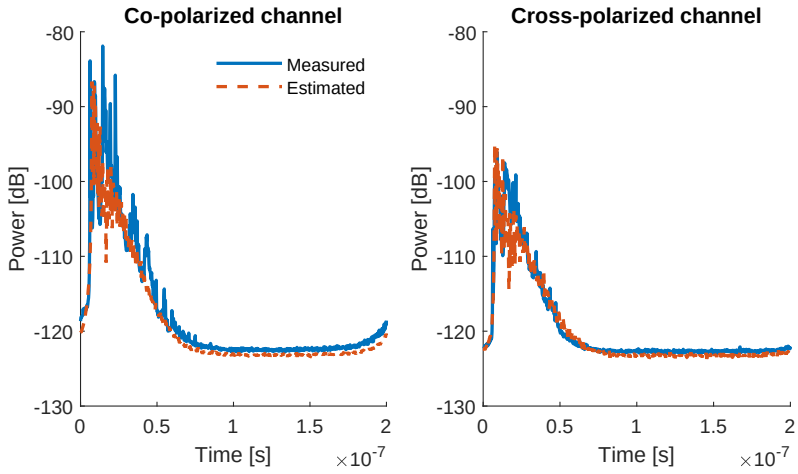


Fig. F.6: Averaged power delay profile from the measurements versus that obtained from the SPPGM after calibration. The parameter estimates are reported in Tab. F.1.

ments is compared to that from the SPPGM in Fig. F.6. The estimated APDP seems to fit the measurements well, thus validating the method.

5 Conclusion

The autoencoder is able to generate summaries that are informative enough to calibrate the parameters of the polarized propagation graph model. In this case, handcrafting of summaries is not necessary for implementing the calibration method of [10]. Avoiding the need for handcrafted summaries enables even non-propagation experts to easily apply the method. Even though the performance of the calibration method is better when using the handcrafted summaries in [10], we do get reasonably accurate results with fairly limited effort of training the autoencoder. However, since the training is done in an unsupervised manner, there is no guarantee that the generated summaries will be informative enough about the parameters to be able to estimate them. With auto-generated summaries, we are one step closer to fully automated model calibration.

Acknowledgment

The authors would like to thank Dr. Carl Gustafson and Prof. Fredrik Tufvesson (Lund University) for providing the measurement data. This work is supported by the Danish Council for Independent Research, grant no. DFF 7017-00265. Ramoni Adeogun is supported by grant no. DFF 9041- 00146B.

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. 21, no. 1, pp. 1–9, Feb 1972.

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

[3] P. Kyösti, "Winner II channel models, deliverables d1.1.2 v1.2, part I: Channel models," Tech. Rep. IST-4-027756 WINNER II Project, 2007.

[4] 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.

[5] 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. on Wireless Commun.*, vol. 18, no. 8, pp. 3856–3870, 2019.

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

[7] A. Bharti, R. Adeogun, and T. Pedersen, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

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

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

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

[11] R. Adeogun, T. Pedersen, C. Gustafson, and F. Tufvesson, "Polarimetric Wireless Indoor Channel Modelling Based on Propagation Graph," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 10, pp. 6585–6595, 2019.

[12] J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander, "Fundamentals and recent developments in approximate bayesian computation," *Syst. Biol.*, vol. 66, pp. 66–82, Jan 2017.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[14] R. Adeogun and T. Pedersen, "Propagation graph based model for multi-polarized wireless channels," in *IEEE WCNC*, April 2018.

[15] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002. [Online]. Available: https://www.genetics.org/content/162/4/2025

[16] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate bayesian computation," *Biometrika*, vol. 96, no. 4, pp. 983–990, Oct 2009.

[17] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of reverberant radio channels using propagation graphs," vol. 60, no. 12, pp. 5978–5988, Dec 2012.

[18] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

Paper G

Joint Statistical Modeling of Received Power, Mean Delay, and Delay Spread for Indoor Wideband Radio Channels

Ayush Bharti, Laurent Clavier, Troels Pedersen

The paper has been published in the
European Conference on Antennas and Propagation, 2020.

Abstract

We propose a joint statistical model for the received power, mean delay, and rms delay spread, which are derived from the temporal moments of the radio channel responses. Indoor wideband measurements from two different data sets show that the temporal moments are strongly correlated random variables with skewed marginals. Based on the observations, we propose a multivariate log-normal model for the temporal moments, and validate it using the experimental data sets. The proposed model is found to be flexible, as it fits different data sets well. The model can be used to jointly simulate the received power, mean delay, and rms delay spread. We conclude that independent fitting and simulation of these statistical properties is insufficient in capturing the dependencies we observe in the data.

1 Introduction

Characterization of radio channel properties such as the received power, mean delay, and rms delay spread are imperative for the design of communication systems. These statistics are computed from the moments of the instantaneous power of the received signal, known as *temporal moments*. Used since the 1970s [1], the temporal moments are ubiquitous in wireless communications literature, and are also used in simulations of communication systems. More recently, temporal moments have been used as summary statistics for parameter estimators for stochastic channel models [2–4]. In applications such as these, where more than one of the temporal moments are used simultaneously, knowledge of their statistical properties, including their dependencies, is beneficial.

Empirical averages and cdfs of received power, mean delay, and rms delay spread are reported frequently in the literature. In [5], Awad et al. survey the empirical data on the delay properties of the indoor radio channel, including the mean delay and rms delay spread, and fit marginal models to a large number of available data sets. They obtained normal, Weibull, or log-normal distributions as the best fit models, with Rayleigh, Rician, and Poisson being the other considered distributions. Although clearly handling a model selection problem, the selection was done by evaluating the best fit without adjusting for model complexity. Similar results were obtained in [6] where the rms delay spread was empirically modeled as a normally distributed random variable. The expected values of temporal moments are connected to parameters describing the model for the impulse responses for in-room scenarios [7]. This observation was further deepened in [8, 9] which show how the arrival rate of a process changes the variance of the mean delay and rms delay spread. It is wide-spread practice to report only empirical marginal distributions of rms delay spread and to disregard the dependencies between moments. Moreover, independent modeling of rms delay spread is prevalent in the literature, while

its dependency with received power and mean delay is not.

An exception is Greenstein et al. [10] who modeled the path gain[1] and delay spread jointly. They proposed a joint log-normal distribution for the path gain and rms delay spread based on intuitive arguments, and later validated it empirically using a wide range of outdoor measurements available from literature. However, Greenstein et al. did not consider mean delay. Moreover, they proposed a fixed correlation coefficient of –0.75 between path gain and rms delay spread, which might not be able to account for the variability observed in measurements.

In this contribution, we extend the Greenstein model to jointly characterize the mean delay, along with received power and rms delay spread. We propose a multivariate log-normal model for the temporal moments, from which we can obtain mean delay and rms delay spread using a simple transformation. We find that this easy-to-use model is flexible enough to capture the variability observed in data. We also provide a method to estimate parameters of the model so that it can be easily fitted to new measurements, something that the Greenstein model lacked. Finally, the model is validated using indoor channel impulse response measurements from two different campaigns.

2 Temporal Moments

Consider the case where measurements of the channel transfer function are recorded using a vector network analyzer (VNA) in a single-input, single-output (SISO) set-up. The transfer function is sampled in the measurement bandwidth B at N_s frequency points with separation $\Delta f = B/(N_s - 1)$. The model for the measured transfer function, Y_n, at frequency sample n reads

$$Y_n = H_n + W_n, \quad n = 0, 1, \ldots, (N_s - 1), \tag{G.1}$$

where H_n is the transfer function and W_n is measurement noise modeled as independent and identically distributed (iid) circularly symmetric Gaussian random variables. The time domain signal, $y(t)$, is obtained by the discrete-frequency, continuous-time inverse Fourier transform as

$$y(t) = \frac{1}{N_s} \sum_{n=0}^{N_s-1} Y_n \exp\left(j2\pi n \Delta f t\right), \tag{G.2}$$

where $y(t)$ has period $1/\Delta f$.

A particular realization of $y(t)$ can be summarized in terms of its temporal moments defined as

$$m_k = \int_0^{\frac{1}{\Delta f}} t^k |y(t)|^2 \mathrm{dt}, \quad k = 0, 1, \ldots, (K - 1). \tag{G.3}$$

[1] Greenstien et al. defined path gain as the ratio of received power to transmitted power.

In total K temporal moments are computed "instantaneously" per realization of the received signal, i.e. without "averaging" over multiple realizations. Thus, having N_{real} realizations of the channel results in the $N_{\text{real}} \times K$ dimensional matrix, $\mathbf{M} = \left[\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(N_{\text{real}})}\right]^T$, where $\mathbf{m}^{(i)} = \left[m_0^{(i)}, m_1^{(i)}, \ldots, m_{K-1}^{(i)}\right]$. The SI unit for the k^{th} temporal moment is s^k.

The instantaneous received power, P_0, equals m_0, while the instantaneous mean delay, $\bar{\tau}$, or the instantaneous rms delay spread, τ_{rms}, are obtained as transformations of the temporal moments as

$$\bar{\tau} = \frac{m_1}{m_0}, \quad \text{and} \quad \tau_{\text{rms}} = \sqrt{\frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2}. \tag{G.4}$$

Note that the unit of $\bar{\tau}$ and τ_{rms} is in seconds. For the purpose of our discussion, we will focus on the first three temporal moments, i.e. (m_0, m_1, m_2), as they suffice for the received power, mean delay, and rms delay spread. We refer to (m_0, m_1, m_2) as *temporal moments* and $(P_0, \bar{\tau}, \tau_{\text{rms}})$ as *standardized moments* of $|y(t)|^2$. Note that in (G.4), we make no attempt to compensate or remove the effect of a finite measurement bandwidth, i.e. the standardized moments are computed of the received signal, $y(t)$, and not of the channel impulse response.

3 Measurement Data and Observations

3.1 Dataset from Lund University

In [11], mm-wave measurements of the channel transfer function are recorded at 60 GHz using a VNA in a SISO set-up. The measurement is conducted in a small room of dimensions $3 \times 4 \times 3$ m^3 using a 25×25 virtual planar array, giving $N_{\text{real}} = 625$ realizations of the channel. Frequency bandwidth used is 4 GHz, with $N_s = 801$ frequency sample points. This gives a signal observation time of $1/\Delta f = 200$ ns in the time domain. Temporal moments are computed for this dataset for the non-line-of-sight (NLOS) case, and the density estimates and scatter plots are shown in Fig. G.1.

3.2 Dataset from Lille

The 60 GHz channel sounder developed in [12] measures the frequency transfer function using a VNA in bandwidth $B = 2$ GHz at $N_s = 1601$ sample points by steps of $\Delta f = 1.25$ MHz. This results in a signal observation time of $1/\Delta f = 800$ ns. Measurements were taken in a computer laboratory of floor area 7.15×5.2 m^2 at 26 sites, covering the whole room. At each site, 250 measurements were carried out. We use a subset of this data, specifically, line-of-sight (LOS) measurements with $N_{\text{real}} = 500$ realizations obtained from the first two sites having the same distance between transmitter and receiver.

Table G.1: Sample Pearson correlation coefficients between standardized and temporal moments of data.

	$\hat{\rho}_{P_0,\bar{\tau}}$	$\hat{\rho}_{P_0,\tau_{\text{rms}}}$	$\hat{\rho}_{\bar{\tau},\tau_{\text{rms}}}$
Lund	-0.28	-0.35	0.53
Lille	-0.55	-0.21	0.85
	$\hat{\rho}_{m_0,m_1}$	$\hat{\rho}_{m_0,m_2}$	$\hat{\rho}_{m_1,m_2}$
Lund	0.94	0.34	0.54
Lille	0.93	0.39	0.69

Density estimates and scatter plots of temporal moments for this dataset is shown in Fig. G.1.

3.3 Observations

The marginal distributions of the temporal moments appear to be skewed, more so for the Lille data. The scatter plots fan out towards the top-right of each plot, giving rise to the skewed marginals. It is evident from the scatter plots that the temporal moments are correlated random variables, suggesting that the standardized moments could be correlated as well. This conjecture is indeed found to be true from the correlation coefficients between the temporal and standardized moments for the two data sets reported in Table G.1. The Pearson correlation coefficient between random variables A and B is

$$\rho_{A,B} = \frac{\text{cov}(A,B)}{\sigma_A \sigma_B}, \tag{G.5}$$

where $\text{cov}(\cdot,\cdot)$ is the covariance operator and σ is the standard deviation. The sample Pearson correlation coefficients for both temporal and standardized moments are reported in Table G.1. We observe that the correlation between received power and rms delay spread is less than the value proposed by Greenstein et al. [10]. Moreover, the correlation of standardized moments across the two data sets vary significantly, while the correlation of temporal moments seems more stable.

4 Proposed model

The temporal moments are non-negative, correlated random variables with skewed marginals. Therefore, we propose to use a multivariate log-normal distribution to model the temporal moments. In principle, one could use a multivariate Gaussian distribution or copulas [13] to model the dependency structure. However, given the support for log-normality of standardized moments in the literature, coupled with the aim to have a general yet simple-to-use model, we propose a multivariate log-normal distribution to model the

vector, $\mathbf{m} = (m_0, m_1, m_2)$, of first three temporal moments, i.e. $K = 3$. The K-variate log-normal distribution, which is the exponential transform of a multivariate Gaussian, has the pdf

$$f(\mathbf{m}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\prod_{k=0}^{K-1} (m_k)^{-1}}{\sqrt{(2\pi)^K \det \boldsymbol{\Sigma}}} \times \exp(-\frac{1}{2}(\ln(\mathbf{m}) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\ln(\mathbf{m}) - \boldsymbol{\mu})), \quad \text{(G.6)}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix of the associated multivariate Gaussian pdf. The entries of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined as $\mu_k = \mathbb{E}[\ln m_k]$ and $\Sigma_{kk'} = \operatorname{cov}(\ln m_k, \ln m_{k'})$, for $k, k' = 0, 1, 2$, respectively. In contrast, the means and covariances of m_0, m_1, and m_2 are functions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as

$$\mathbb{E}[m_k] = \exp\left(\mu_k + \frac{1}{2}\Sigma_{kk}\right), \quad \text{and} \quad \text{(G.7)}$$

$$\operatorname{cov}(m_k, m_{k'}) = e^{\left(\mu_k + \mu_{k'} + \frac{1}{2}(\Sigma_{kk} + \Sigma_{k'k'})\right)} \left(e^{\Sigma_{kk'}} - 1\right). \quad \text{(G.8)}$$

The multivariate log-normal is a positive distribution which models the skewed marginals better than a Gaussian.

In principle, the received power, mean delay and rms delay spread could be the quantities modeled using the multivariate log-normal distribution. In practice, we do not observe any qualitative difference between one or the other. However, here we chose to model the temporal moments as their means and covariances are easier to compute analytically, given a channel model, as compared to the standardized moments due to the non-linear transformation. Using the model of the temporal moments, the standardized moments can be simulated via the one-to-one transformation given in (G.4).

4.1 Estimation of parameters

Fitting the matrix of temporal moments $\mathbf{M}$, obtained from N_{real} independent realizations of the channel impulse responses, to the proposed model requires the estimation of the mean vector, $\boldsymbol{\mu}$, and the covariance matrix, $\boldsymbol{\Sigma}$. This can be achieved by maximizing the likelihood of the data as

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \underset{\boldsymbol{\mu}, \boldsymbol{\Sigma}}{\operatorname{argmax}} \prod_{i=1}^{N_{\text{real}}} f\left(\mathbf{m}^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \quad \text{(G.9)}$$

Table G.2: Parameter estimates obtained after fitting.

	Lund data			Lille data		
$\hat{\mu}^T$	-38.8	-56.8	-74.4	-29.0	-47.2	-63.2
$\hat{\Sigma}$	2.8×10^{-3}	2.5×10^{-3}	1.4×10^{-3}	0.22	0.17	0.12
	2.5×10^{-3}	2.6×10^{-3}	2.1×10^{-3}	0.17	0.15	0.19
	1.4×10^{-3}	2.1×10^{-3}	5.3×10^{-3}	0.12	0.19	0.70

Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the parameters of the associated Gaussian, the maximum likelihood estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are

$$\hat{\boldsymbol{\mu}} = \frac{1}{N_{\text{real}}} \sum_{i=1}^{N_{\text{real}}} \ln \mathbf{m}^{(i)}, \quad \text{and} \tag{G.10}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N_{\text{real}}} \sum_{i=1}^{N_{\text{real}}} \left(\ln \mathbf{m}^{(i)} - \hat{\boldsymbol{\mu}} \right) \left(\ln \mathbf{m}^{(i)} - \hat{\boldsymbol{\mu}} \right)^T. \tag{G.11}$$

The estimates obtained after fitting the model to the two data sets are reported in Table G.2.

4.2 Simulation from the model

Simulation from the proposed model is straightforward. To generate one realization, three steps should be performed:

1. Draw $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
2. $\mathbf{m} = \exp(\mathbf{x})$ (entry-wise exponential)
3. Compute $\bar{\tau}$ and τ_{rms} from (G.4) (optional)

5 Model Validation

We check the validity of the model by fitting it to the two experimental datasets available, and then qualitatively investigating the simulated data against the measurements.

5.1 Simulation of temporal moments

We estimate the parameters of the proposed model for the two datasets, and then simulate temporal moments using the methodology described in the previous section. The results are shown in Fig. G.1 for both Lund and Lille data, and the parameter estimates are in Table G.2. The model appears to fit the

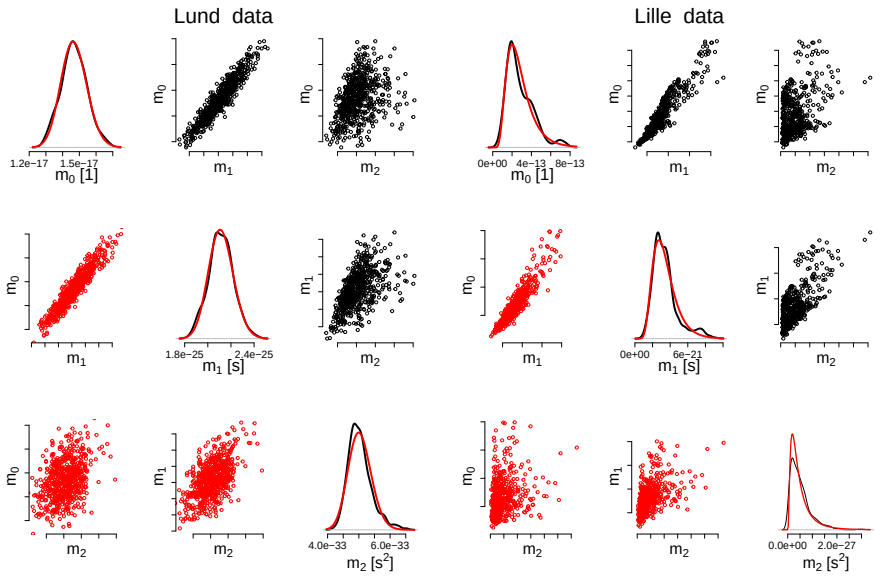


Fig. G.1: Density estimates and scatter plots of temporal moments obtained from (left) Lund data and (right) Lille data (shown in black) vs. those simulated from the fitted proposed model (shown in red). The scales of the corresponding scatter plots are the same. Number of points simulated is same as in the measurements, i.e. $N_{\text{real}} = 625$ for Lund data and $N_{\text{real}} = 500$ for Lille data. Correlation coefficients of temporal moments are reported in Table G.1 and the parameter estimates are in Table G.2.

marginals well, even for the very skewed case of Lille data. The high correlation between the temporal moments, especially between m_0 and m_1, is well captured by the model. For both Lund and Lille data, the fanning out of the scatter plots is well represented by the model.

5.2 Simulation of P_0, $\bar{\tau}$, and τ_{rms}

We now compare the scatter plots of received power, mean delay, and rms delay spread obtained from the proposed model with those from the measurements in Fig. G.2. Additionally, we also show the samples obtained by independently fitting log-normal pdfs to the standardized moments from the measurements. We observe a strong positive correlation between the mean delay and rms delay spread obtained from measurements. The received power, however, is negatively correlated with mean delay and rms delay spread. This dependency structure between the standardized moments is captured well by the proposed joint model. In contrast, any information on the correlation between the variables is lost when simulating them independently.

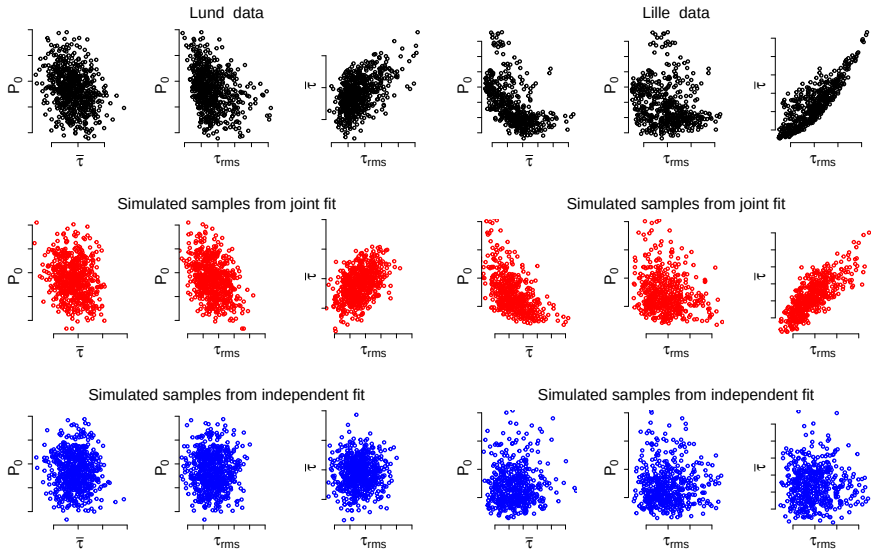


Fig. G.2: Scatter plots of received power, mean delay, and rms delay spread obtained from (left) Lund data and (right) Lille data (in black) vs. those simulated from the proposed model (in red). The samples simulated by independently fitting P_0, $\bar{\tau}$, and τ_{rms} to the data are shown in blue. The scales of the corresponding scatter plots are the same. Number of points simulated is same as in the measurements, i.e. $N_{\text{real}} = 625$ for Lund data and $N_{\text{real}} = 500$ for Lille data. Correlation coefficients of standardized moments are reported in Table G.1 and the parameter estimates are in Table G.2.

6 Conclusion

Observing that temporal moments of channel impulse responses, and hence their standardized moments, are dependent random variables, we propose to model them as jointly log-normal random variables. The proposed model is simple, easy to use, and analyze. We find that the model is flexible enough to fit measurements exhibiting contrasting behaviors. This model can be used to jointly simulate received power, mean delay, and rms delay spread. We validated the joint model using experimental data obtained from indoor environments.

Independent fitting and simulation of received power, mean delay, and rms delay spread leads to loss of correlation observed in the measurements, and these should be simulated jointly. Therefore, reporting only their marginal distributions, e.g. in the form of plots of their empirical cdfs, is insufficient. Instead, their means and covariances should be reported.

Acknowledgment

The authors would like to thank Dr. Carl Gustafson and Prof. Fredrik Tufvesson (Lund University) for providing the measurement data. This work is supported by the Danish Council for Independent Research, grant no. DFF 7017-00265 and performed within the framework of the COST Action CA15104 IRACON.

References

[1] D. Cox, "Delay doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment," *IEEE Trans. Antennas Propag.*, vol. 20, no. 5, pp. 625–635, sep 1972.

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

[3] A. Bharti, L. Clavier, and T. Pedersen, "Joint statistical modeling of received power, mean delay, and delay spread for indoor wideband radio channels," in *14th Eur. Conf. on Antennas and Propag.*, 2020, pp. 1–5.

[4] A. Bharti, R. Adeogun, and T. Pedersen, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

[5] M. K. Awad, K. T. Wong, and Z. Li, "An integrated overview of the open literature's empirical data on the indoor radiowave channel's delay properties," *IEEE Trans. on Antennas and Propag.*, vol. 56, no. 5, pp. 1451–1468, May 2008.

[6] Y. Yu, Y. Liu, W. Lu, and H. Zhu, "Measurement and empirical modelling of root mean square delay spread in indoor femtocells scenarios," *IET Commun.*, vol. 11, no. 13, pp. 2125–2131, 2017.

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

[8] T. Pedersen, "Modeling of path arrival rate for in-room radio channels with directive antennas," *IEEE Trans. on Antennas and Propag.*, vol. 66, no. 9, pp. 4791–4805, 2018.

[9] ——, "Stochastic Multipath Model for the In-Room Radio Channel Based on Room Electromagnetics," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 4, pp. 2591–2603, April 2019.

[10] L. Greenstein, V. Erceg, Y. Yeh, and M. Clark, "A new path-gain/delay-spread propagation model for digital cellular channels," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 477–485, May 1997.

[11] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

[12] M. Fryziel, C. Loyez, L. Clavier, N. Rolland, and P. A. Rolland, "Path-loss model of the 60-ghz indoor radio channel," *Microw. and Opt. Technol. Lett.*, vol. 34, no. 3, pp. 158–162, 2002. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mop.10402

[13] R. B. Nelsen, *An Introduction to Copulas.* Springer-Verlag GmbH, 2007. [Online]. Available: https://www.ebook.de/de/product/5270140/roger_b_nelsen_an_introduction_to_copulas.html

Paper H

Joint Modeling of Received Power, Mean Delay, and Delay Spread for Wideband Radio Channels

Ayush Bharti, Ramoni Adeogun, Xuesong Cai, Wei Fan,
François-Xavier Briol, Laurent Clavier, Troels Pedersen

The paper has been published in the
IEEE Transactions on Antennas and Propagation, 2021.

Abstract

We propose a multivariate log-normal distribution to jointly model received power, mean delay, and root mean square (rms) delay spread of wideband radio channels, referred to as the standardized temporal moments. The model is validated using experimental data collected from five different measurement campaigns (four indoor and one outdoor scenario). We observe that the received power, the mean delay, and the rms delay spread are correlated random variables, and therefore, should be simulated jointly. Joint models are able to capture the structure of the underlying process, unlike the independent models considered in the literature. The proposed model of the multivariate log-normal distribution is found to be a good fit for a large number of wideband data-sets.

1 Introduction

Standardized temporal moments such as received power, mean delay, and root mean square (rms) delay spread are widely used to summarize power-delay profiles (PDPs) of wideband radio channels. Characterization of these temporal moments is imperative for understanding the effects of multipath propagation on the received signal [1], and hence, for the design and analysis of communication and localization systems. The standardized temporal moments are derived from transformations of the raw temporal moments of the instantaneous power of the received signal. Therefore, the raw moments, and consequently the standardized moments, are dependent random variables. The raw temporal moments have recently been used to estimate parameters of stochastic radio channel models from measurements [2–7]. Mean delay and rms delay spread have also been used to fit an extension of the WINNER II model to measurements [8]. In applications where multiple temporal moments are used, it can be valuable to consider their dependencies to avoid biases which can occur due to false assumptions of independence.

Independent modeling of received power, mean delay, and rms delay spread is prevalent in the literature, with their empirical averages and cumulative distributions functions (CDFs) being reported frequently while disregarding their dependencies. A survey of the empirical data available for the delay properties of indoor radio channel is given in [9], where a variety of marginal models is fit to the mean delay and rms delay spread from the various data-sets. They obtained log-normal, Gaussian, and Weibull as the best fit models. Empirical distribution of delay spread has been modeled using a log-normal distribution in the 910 MHz channel [10, 11], the 30 MHz to 400 MHz frequency band [12], at 460 MHz [13], at 11 GHz [14], and at 39 GHz [15]. A Gaussian distribution for the rms delay spread has also been proposed based on empirical data in [16] and [17]. Recently, the rms delay spread has also been modeled using a bimodal Gaussian mixture distribution [18] and neural networks [19].

The shortcomings of independent modeling become clear by considering

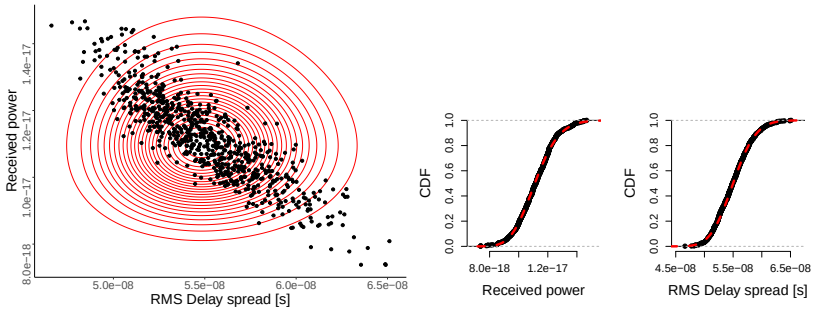


Fig. H.1: Scatter plot of received power and rms delay spread obtained from AAU-Hall measurements (see Sec. 4.4) is shown in black (above). The contour lines from independently fitting log-normal distribution to the measurements is shown in red. The empirical CDFs of the marginals is also shown with the fitted log-normal CDF in red (below). Note that the received power is unitless.

jointly the received power and rms delay spread as done in the example in Fig. H.1. It is apparent that by fitting independent log-normal models to the received power and the rms delay spread, the marginals of the data are modeled correctly. However, the correlation information in the data is lost on modeling them independently. Delay spread is previously found to be correlated to received power at 60 GHz [20] and to mean delay in the ultra-high frequency band [21]. One approach to mitigate this problem is to model the standardized moments jointly. An exception to the independent models is the one proposed by Greenstein et al. [11] where they accounted for the correlation between rms delay spread and shadow fading after analysing a wide range of outdoor measurements, mostly in the 900 MHz frequency band. They argued that rms delay spread is log-normally distributed at a given propagation distance, and proposed a joint log-normal model for path gain[1] and delay spread with a correlation coefficient of –0.75. However, they did not take mean delay into account. Moreover, the correlation coefficient was based on qualitative analysis of scatter plots and on a single measurement setting. The mutual relations between the means of the raw temporal moments have been modeled in [22–24] for the in-room case, while their joint distribution was not studied. To the best of our knowledge, joint characterization of the temporal moments in the millimetre-wave (mm-wave) band has not been done before.

Potentially, the temporal moments could be modeled jointly using a multivariate distribution such that the model could be fitted to new measurements. Joint modeling of multivariate random variables is considerably more involved than modeling of scalar random variables because the model is required to represent the marginals and the dependency structure in the data at the same

[1] Greenstein et al. [11] defined path gain as the ratio of received power to transmitted power.

time. Only a few univariate probability distribution functions (pdfs) exist that have unique multivariate extensions, such as the multivariate Gaussian, log-normal, and Gamma distributions [25]. Copulas [26] can also be used to model the dependency structure between the random variables, especially when the marginal distributions lead to a multivariate distribution that is difficult to handle due to the lack of analytical expression or difficulties to estimate the parameters.

After considering several of these methods, we conclude that the multivariate log-normal is a reasonable choice which provides a good balance between goodness-of-fit and ease of interpretation. Moreover, there is substantial support for log-normality of standardized temporal moments in the literature. In this paper, we propose and validate the multivariate log-normal model using a wide variety of measurements taken in different scenarios and frequency ranges, including both indoor and outdoor settings. Measurement campaigns were conducted at Lund University [27], University of Lille [28], and Aalborg University (AAU) [29]. We also present mm-wave measurements from one indoor and one outdoor campaign in the 28 GHz to 30 GHz band conducted recently at AAU. We compare the proposed model with the multivariate Gaussian and independent marginal models in terms of the Akaike Information Criterion (AIC). Finally, we investigate the model fits to the raw and standardized temporal moments from the measurements. Preliminary results have been published in the conference publication [30].

The paper is organized as follows: Section II describes the raw and standardized temporal moments, and Section III presents the model. In Section IV we compare the proposed model with other modeling choices. Section V and VI compare the model fits to the raw and standardized temporal moments of the measurements, respectively. Finally, the conclusions are outlined in Section VII.

2 Temporal Moments

Consider a measurement campaign where the channel transfer function between fixed transmit and receive antennas is recorded using a vector network analyzer (VNA). Sampling the transfer function, $H(f)$, at N_s frequency points in the measurement bandwidth B results in a separation of $\Delta f = B/(N_s - 1)$ between the points. We assume that the measurement noise at the n^{th} frequency point, W_n, is additive and independent of the transfer function, H_n. Then, the measured frequency-domain signal, Y_n, reads

$$Y_n = H_n + W_n, \quad n = 0, 1, \ldots, (N_s - 1). \tag{H.1}$$

Discrete-frequency, continuous-time inverse Fourier transform gives the $1/\Delta f$-periodic measured time-domain signal

$$y(t) = \frac{1}{N_s} \sum_{n=0}^{N_s-1} Y_n \exp(j2\pi n \Delta f t). \tag{H.2}$$

Note that $y(t)$ is often referred to as the impulse response despite suffering from limited bandwidth and noise. This terminology is somewhat misleading since strictly speaking the impulse response is the inverse Fourier transform of $H(f)$. For large bandwidth and high signal-to-noise ratio (SNR), $y(t)$ can be used as an approximation to the impulse response in the time interval $[0, 1/\Delta f]$, provided that the impulse response decays rapidly enough. To avoid this confusion, we refer to $y(t)$ as the measured signal.

The raw temporal moments are summary statistics of the measured signal $y(t)$, where the k^{th} temporal moment is defined as

$$m_k = \int_0^{\frac{1}{\Delta f}} t^k |y(t)|^2 \mathrm{d}t, \quad k = 0, 1, \ldots, (K-1). \tag{H.3}$$

Here, a total of K raw temporal moments are computed "instantaneously" per realization of $y(t)$, giving the K-dimensional vector $\mathbf{m} = [m_0, m_1, \ldots, m_{K-1}]^\top$. The raw temporal moments are correlated random variables as they are all derived from the received signal power, $|y(t)|^2$. The k^{th} temporal moment is measured in $[\text{second}]^k$.

The standardized temporal moments are obtained from the first three raw temporal moments. The received power, P_0, the mean delay, $\bar{\tau}$, and the rms delay spread, τ_{rms}, are given as

$$P_0 = m_0, \quad \bar{\tau} = \frac{m_1}{m_0}, \quad \text{and} \quad \tau_{\text{rms}} = \sqrt{\frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2}. \tag{H.4}$$

The unit of $\bar{\tau}$ and τ_{rms} is in seconds whereas P_0 is unitless. The deterministic relationship between the raw and the standardized temporal moments is depicted in Fig. H.2. The non-linearity of the above transformations and the dependency of the raw temporal moments complicates the joint characterization of mean delay and rms delay spread. Summarizing N_{real} realizations of the measured signal into K temporal moments therefore results in the $K \times N_{\text{real}}$ dimensional matrix, $\mathbf{M} = \left[\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(N_{\text{real}})}\right]$. We will focus our discussion on the first three temporal moments, (m_0, m_1, m_2), as they suffice for the received power, mean delay, and rms delay spread but it is straightforward to extend the framework to include more moments as long as the marginal distributions fit the same distribution.

Note that the standardized temporal moments in (H.4) are computed from the measured signal, $y(t)$, rather than the channel impulse response. The impulse response is unobservable due to the noise and bandwidth limitations.

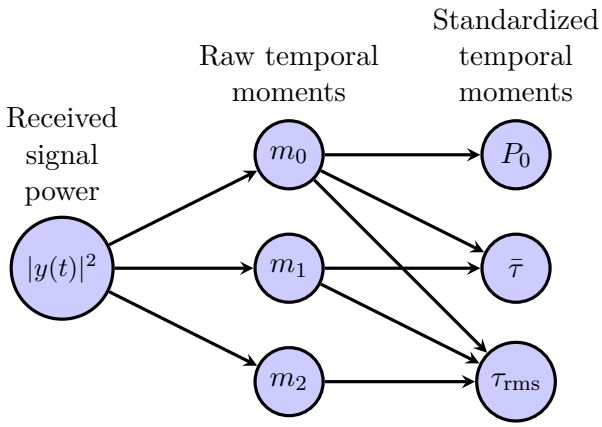


Fig. H.2: The connections between the magnitude square received signal and the summary statistics (raw- and standardized temporal moments).

It is widespread practice to employ a thresholding procedure to reduce the effect of the measurement noise on the estimation of temporal moments. However, such procedures require the setting of a threshold or dynamic range. The choice of the threshold affects the resulting estimates in a manner that makes comparison between measurements obtained with different equipment difficult. For this reason, we omit any thresholding procedure in the present work.

The finite measurement bandwidth also manifests itself in the rms delay spread as an approximately additive term equal to the delay spread of the transmitted signal. This effect can be partially removed by subtracting the delay spread of the frequency window. This is widespread practice in the literature and results in a good approximation if the bandwidth is large and the SNR is high. However, in case of low SNR and small signal bandwidth, this can lead to inaccurate and sometimes negative estimates of the delay spread. For the measurements considered in Section 4, where the bandwidth is very large, the effect of the transmitted signal can be ignored. Hence, we make no attempt to compensate for the effect of a finite measurement bandwidth.

3 Proposed Statistical Model

We intend to jointly model the first three raw temporal moments, (m_0, m_1, m_2), and use the transformation in (H.4) to simulate the mean delay and rms delay spread. In principle, the standardized temporal moments could be modeled instead of the raw moments. However, the distribution on the raw moments implies a distribution on the standardized moments from which sampling is straightforward. Modeling the raw moments has the added advantage that their means and covariances are easier to compute analytically for a given

Table H.1: Summary of different measurement data-sets.

Data set	Bandwidth (GHz)	No. of samples	No. of realizations	Antenna Tx/Rx	Dimensions (m^3)	Scenario	Environment
Lund Data [27]	58-62	801	625	Biconical/Open waveguide	$3 \times 4 \times 3$	NLOS	Small room
Lille Data [28]	59-61	1601	750	Microstrip/Microstrip	$5.20 \times 7.15 \times 2.90$	LOS	Large room
AAU-Industry [29]	3-8	5001	95	Biconical/Biconical	$33 \times 14 \times 6$	Both	Industry hall
AAU-Hall	28-30	1500	720	Biconical/Biconical	$44 \times 25 \times 10$	NLOS	Large hall
AAU-Outdoor	28-30	2000	360	Horn/Biconical	—	LOS	Outdoor

channel model than those of the standardized moments due to the non-linear transformation.

We model the vector $\mathbf{m} = [m_0, m_1, m_2]^\top$ as a multivariate log-normal variable. The exponential of a random vector following a multivariate Gaussian distribution is multivariate log-normal distributed. Let $\mathbf{x}$ be a K-variate normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then its entry-wise exponentiation, $\mathbf{m} = \exp(\mathbf{x})$, yields a log-normal vector with pdf

$$f(\mathbf{m}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\prod_{k=0}^{K-1} (m_k)^{-1}}{\sqrt{(2\pi)^K \det \boldsymbol{\Sigma}}} \times \exp\left(-\frac{1}{2}(\ln(\mathbf{m}) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\ln(\mathbf{m}) - \boldsymbol{\mu})\right). \quad \text{(H.5)}$$

Here the logarithm is taken entry-wise. By property of the marginals of the multivariate Gaussian, it is easy to see that this transform results in a distribution with log-normal marginals. Note that the parameters of a multivariate log-normal are the mean vector and the covariance matrix of the associated multivariate Gaussian distribution. The entries of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given as $\mu_k = \mathbb{E}[\ln m_k]$ and $\Sigma_{kk'} = \operatorname{cov}(\ln m_k, \ln m_{k'})$, for $k, k' = 0, 1, \ldots, K-1$, respectively. Given that raw temporal moments are log-normally distributed, their means and covariances can be related to μ and Σ as

$$\mathbb{E}[m_k] = \exp\left(\mu_k + \frac{1}{2}\Sigma_{kk}\right), \quad \text{and} \quad \text{(H.6)}$$

$$\operatorname{cov}(m_k, m_{k'}) = \exp\left(\mu_k + \mu_{k'} + \frac{1}{2}(\Sigma_{kk} + \Sigma_{k'k'})\right) \times (\exp(\Sigma_{kk'}) - 1). \quad \text{(H.7)}$$

Note that we model the raw temporal moments as opposed to Greenstein et al. [11] who model shadow fading and rms delay spread as jointly log-normal. With the proposed model, log-normality is preserved for the received power and mean delay due to the multiplicative transform applied on m_0 and m_1. However, the distribution of rms delay spread depends on a more complicated transformation (see (H.4)) and hence cannot easily be derived in closed form.

3.1 Estimation of parameters

The parameters of the proposed model need to be estimated from measured data in order to use the model for simulation purposes. Here, we refer to the matrix of raw temporal moments, $\mathbf{M}$, as the data. This data matrix is obtained by summarizing N_{real} realizations of the measured signal using (23). Assuming independent and identically distributed (iid) realizations, maximum likelihood estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is achieved by solving the optimization problem,

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \underset{\boldsymbol{\mu}, \boldsymbol{\Sigma}}{\operatorname{argmax}} \quad \prod_{i=1}^{N_{\text{real}}} f\left(\mathbf{m}^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \tag{H.8}$$

Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix, respectively, of the associated multivariate Gaussian distribution, their maximum likelihood estimates, $\hat{\boldsymbol{\mu}}$, and $\hat{\boldsymbol{\Sigma}}$, are

$$\hat{\boldsymbol{\mu}} = \frac{1}{N_{\text{real}}} \sum_{i=1}^{N_{\text{real}}} \ln \mathbf{m}^{(i)}, \quad \text{and} \tag{H.9}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N_{\text{real}}} \sum_{i=1}^{N_{\text{real}}} \left(\ln \mathbf{m}^{(i)} - \hat{\boldsymbol{\mu}}\right) \left(\ln \mathbf{m}^{(i)} - \hat{\boldsymbol{\mu}}\right)^{\top}. \tag{H.10}$$

3.2 Simulation from the model

Given a particular value of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, simulation from the proposed model is straightforward. To generate one sample of $\mathbf{m}$, or subsequently, one sample of $(P_0, \bar{\tau}, \tau_{\text{rms}})$, the following steps should be performed:

1. Draw $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
2. Compute entry-wise exponential, $\mathbf{m} = \exp(\mathbf{x})$
3. Compute $\bar{\tau}$ and τ_{rms} from (H.4)

4 Measurement Data Description

We now describe the different radio channel measurements used to validate the proposed model. An overview of the measurement data-sets is given in Tab. H.1. The parameter estimates obtained after fitting the proposed model to the measurements are reported in Tab. H.2.

4.1 Data-set from Lund University

Polarimetric radio channel measurements at 60 GHz was recorded in a small meeting room of dimensions $3 \times 4 \times 3$ m^3 using a VNA [27]. The room consisted

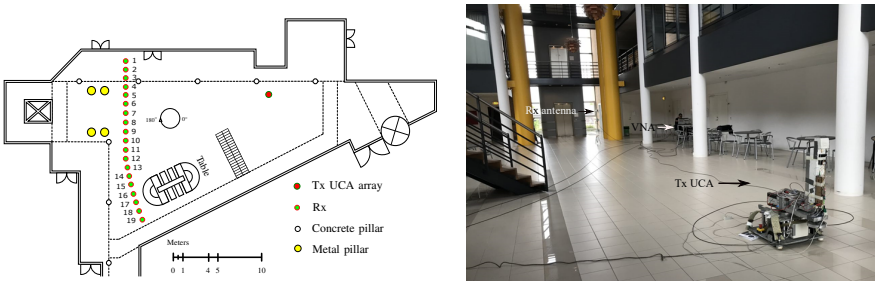
of a table, white-board, bookshelves, and a window on one of the walls. The receive antenna was placed at one corner of the room and the transmit antenna was placed on the table. A water-filled human phantom was used to block the line-of-sight (LOS) path to emulate non-line-of-sight (NLOS) scenario. A 5×5 virtual array of dual-polarized antennas was used with an inter-element spacing of 5 mm at both the transmitter and the receiver. This resulted in a 25×25 dual-polarized MIMO system, however, in this paper, we only use the vertical-vertical polarized channels. Measurements were performed in the bandwidth range of 58 GHz to 62 GHz using 801 equally spaced frequency points. For further details on the measurement campaign, see [27].

4.2 Data-set from Lille University

Measurements were taken in a computer laboratory of floor area 7.15×5.2 m^2 at 26 sites, covering the whole room. The 60 GHz channel sounder developed at IEMN [28] used two heterodyne emission and reception heads developed by monolithic integration with frequencies ranging from 57 GHz to 59 GHz and with intermediate frequencies of 1 GHz to 3 GHz. A dedicated network analyzer allows, after calibration, the vectored measure of the frequency transfer function by steps of 1.25 MHz. The resulting impulse response has a delay resolution of 0.5 ns and a maximum measurable delay of 800 ns. In this paper, we select a subset of the entire data-set, specifically, taking the measurements from the first three sites having the same distance between the transmit and receive antennas in LOS condition. Each site consists of 250 positions separated by 2 mm. The transmitter was fixed in a corner, close to the roof, pointed towards the opposite corner. The receiver was oriented towards the transmitter in the horizontal plane but not in the vertical one. Horizontal linear polarization patch antennas were employed.

4.3 AAU Data, Industry Scenario

Short-range ultra-wideband measurement campaigns were conducted in a $33 \times 14 \times 6$ m^3 industrial factory hall at the Smart Production Lab, AAU. The factory hall was a typical high clutter density environment with large metallic machinery such as welding machines, hydraulic press, and material processing machines. Measurements were collected over the frequency range 3 GHz to 8 GHz using a Rhode & Schwarz ZND 8.5 GHz VNA and omni-directional broadband bi-conical antennas at both the transmitter and the receiver. During the measurements, the transmitter was placed at a fixed location and the receiver location was varied to obtain horizontal distances between 1 m and 9 m. A total of 95 channel transfer functions were obtained with a frequency resolution of 1 MHz corresponding to 5001 samples over the 5 GHz bandwidth. Detailed description of the measurements can be found in [29].



(a) (b)

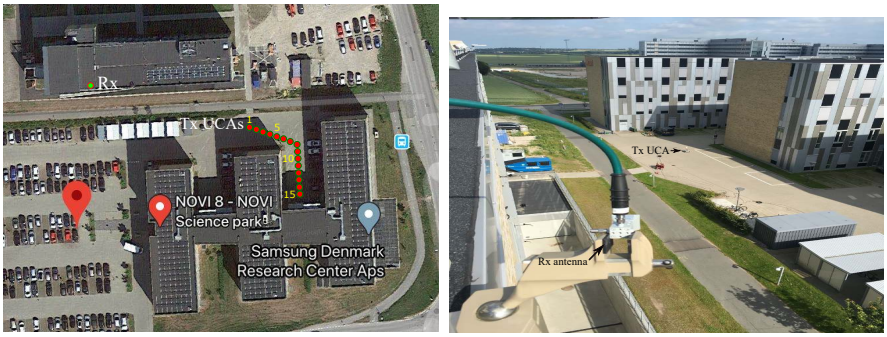
Fig. H.3: The layout (a) and a photo (b) of the indoor hall taken during the measurement campaign conducted at Aalborg University. The measurements corresponding to the 1st receive antenna array position are presented in this paper.

4.4 AAU Data, Hall Scenario

Measurements were conducted in a large hall scenario as illustrated in Fig. H.3(a). A photo taken during the measurement campaign is also shown in Fig. H.3(b). The hall had a floor area of 44×25 m^2 with a height of approximately 10 m. As shown in the picture and the layout sketch, tables, metallic pillars, concrete pillars, stairs, etc. were in the hall. The VNA measurements were taken with the ultra-wideband radio-over-fiber channel sounder developed at AAU [31]. Quasi-omnidirectional biconical antennas [32] were used. The receive antenna was fixed with a height of 1 m to the ground while transmit antenna was installed on a rotator and rotated with 720 uniform steps on a circle with a radius of 0.54 m. In each step, the channel transfer function from 28 GHz to 30 GHz was swept with 1500 samples in the frequency domain. In this paper we analyse the first of the 19 different locations recorded. For this location, the transmitter-receiver distance was around 15 m in NLOS condition.

4.5 AAU Data, Outdoor Scenario

Outdoor measurements were conducted in an open area to in-between the two buildings as shown in Fig. H.4. The same channel sounder is used as in the indoor hall scenario. In this case, the transmitter antenna was rotated with a radius of 0.25 m in 360 uniform steps. In each step, the channel transfer function from 28 GHz to 30 GHz was swept with 2000 samples. The receive antenna was fixed on a roof edge with a height of around 20 m. To increase the SNR, the receive antenna was replaced by a horn antenna with half-power-beamwidths around 30° in both azimuth and elevation. Moreover, its main beam was down tilted to appropriately cover the transmit antenna. Data was collected from 15 transmit transmit antenna locations as indicated in Fig. H.4. The data used in this paper is from the 7th location which was in LOS condition.



(a) (b)

Fig. H.4: Environment for the outdoor measurement campaign conducted at Aalborg University. Measurements from transmit antenna location number 7 are presented in this paper.

5 Model Comparison

To characterize the raw temporal moments jointly, their marginal distributions as well as their correlation structure needs to be well represented. We compare the proposed model against competing model choices for the available data-sets.

5.1 Model Comparison using AIC

We compare the proposed joint model with the model of a multivariate Gaussian distribution. We also include three independent models for the raw temporal moments based on log-normal, Gaussian, and Gamma distributions. We omit comparison with the multivariate Gamma distributions in [25] as they did not give useful results when fitted to the raw temporal moments. Model comparison is done by computing the Akaike Information Criterion (AIC) value [33] of the competing models. AIC is a common tool for model selection that estimates the quality of different models relative to each other. It compares models through their likelihoods, but penalises models with a larger number of parameters κ. One motivation for this penalty comes from Ockham's razor, which states that, when comparing models, one should prefer the simplest model which explains the data well. The criterion is computed as follows

$$\mathrm{AIC} = -2\mathcal{L} + 2\kappa, \tag{H.11}$$

where $\mathcal{L}$ is the maximized log-likelihood of the data. Given a set of models fitted by maximum likelihood to the same data, the preferred model is the one with the lowest AIC value. The reader is referred to [34, Ch. 2] for a detailed discussion. We also considered the Bayesian Information Criterion (BIC), which penalises more than AIC for a large number of parameters; see [35]

Table H.2: Parameter estimates obtained using maximum likelihood estimation. Each entry corresponds to the estimate for some scalar parameter θ, which corresponds to an element of either the 3-dimensional mean (column) vector $\boldsymbol{\mu}$ or the 3×3 dimensional covariance matrix $\boldsymbol{\Sigma}$. The number in bracket (δ) is the half-width of the 95% confidence interval for that parameter, so that the interval takes the form $(\theta - \delta, \theta + \delta)$.

Data set	Mean vector $\hat{\boldsymbol{\mu}}(\pm\delta)$	Upper triangle of Covariance matrix $\hat{\boldsymbol{\Sigma}}(\pm\delta)$		
Lund	–39 (4×10^{-3})	2.8 (0.3)$\times 10^{-3}$	2.5 (0.3)$\times 10^{-3}$	1.4 (0.3)$\times 10^{-3}$
	–57 (4×10^{-3})		2.6 (0.3)$\times 10^{-3}$	2.1 (0.3)$\times 10^{-3}$
	–74 (6×10^{-3})			5.3 (0.6)$\times 10^{-3}$
Lille	–29 (0.03)	0.19 (0.02)	0.15 (0.02)	0.11 (0.03)
	–47 (0.03)		0.14 (0.01)	0.19 (0.03)
	–63 (0.06)			0.70 (0.07)
AAU-Industry	–36 (0.31)	2.34 (0.67)	1.36 (0.40)	1.24 (0.38)
	–53 (0.18)		0.82 (0.23)	0.77 (0.23)
	–70 (0.18)			0.84 (0.24)
AAU-Hall	–39 (9×10^{-3})	1.4 (0.14)$\times 10^{-2}$	1.2 (0.12)$\times 10^{-2}$	6.6 (0.76)$\times 10^{-3}$
	–56 (7×10^{-3})		1.0 (0.11)$\times 10^{-2}$	6.2 (0.68)$\times 10^{-3}$
	–72 (5×10^{-3})			4.6 (0.48)$\times 10^{-3}$
AAU-Outdoor	–40 (1.2×10^{-2})	1.3 (0.20)$\times 10^{-2}$	9.9 (0.14)$\times 10^{-3}$	5.2 (0.82)$\times 10^{-3}$
	–56 (9×10^{-3})		7.6 (1.1)$\times 10^{-3}$	4.2 (0.64)$\times 10^{-3}$
	–71 (5×10^{-3})			2.7 (0.40)$\times 10^{-3}$

and [34, Ch. 3]. However, the ordering of the models was found to be the same for both the criteria, and therefore we omit the BIC values here.

The models are fitted to the five aforementioned data-sets by maximizing their likelihood. The parameter estimates obtained for the proposed model are reported in Tab. H.2. The AIC values of the joint fit of the raw temporal moments are reported in Tab. H.3, with $\kappa = 9$ for the multivariate distributions, and $\kappa = 6$ for the independent marginal models. The proposed model comes out as the better choice for the joint fit for three out of five data-sets, with the multivariate Gaussian performing better for Lille Data and AAU-Outdoor. However, the AIC values for both the joint models are close to each other. It is evident that modeling the random variables independently leads to a significantly poorer fit than either of the joint models, no matter which distribution is chosen. We remark that using more complicated models such as copulas [26] to model the dependency structure may lead to a better fit, but could be harder to interpret.

5.2 Log-normal vs. Gaussian Marginals

We now compare the marginal fits of the multivariate log-normal and Gaussian distributions for modeling the raw temporal moments. To assess model fit, the quantiles of the data are plotted against the theoretical quantiles of the model being assessed. If the model is a good fit, then the quantiles of the data and the theoretical quantiles should be close to one another, and the points will

Table H.3: AIC values for different model choices for the raw temporal moments. Best model is indicated in bold. Note that the joint AIC for the independent models is the sum of the AIC values of the three marginals.

Data set	Multivariate Log-normal	Multivariate Gaussian	Independent Log-normal marginals	Independent Gaussian marginals	Independent Gamma marginals
Lund	**−219636.0**	−219573.9	−217787.6	−217750.2	−217777.0
Lille	−208357.2	**−208665.6**	−205657.4	−204816.6	−205589.4
AAU-Industry	**−29815.61**	−28922.53	−29337.3	−28604.83	−29201.75
AAU-Hall	**−247225.8**	−247212.2	−243329.2	−243348.8	−243342.9
AAU-Outdoor	−125244.8	**−125286.7**	−122385.1	−122342.4	−122374.1

hence lie approximately on a straight line. On the other hand, any deviation from this line might indicate issues with the fit of the model. See e.g. [36, Sec. 10.2] for more details. We show the Q-Q plots for two of the five data-sets, namely the Lille and the AAU-Outdoor data, in Fig. H.5, as they highlight the difference between the fits obtained from both the distributions. The Q-Q plots of AAU-Outdoor data is representative of what we observed for the other data-sets, therefore we exclude reporting them.

We observe that for AAU-Outdoor data, the marginals are well-modeled by both the log-normal and the Gaussian distributions. The fit is similar for both distributions, and it is not apparent which model performs better. As can be seen in Fig. H.6, the marginals in AAU-Outdoor data are very close to being symmetric, which means that the Gaussian fits well. However, for the Lille data, it is evident that the log-normal distribution outperforms the Gaussian in terms of the marginals. The log-normal is able to model the left tail and the center of the distribution very well, but sometimes performs poorly for the right tail. On the other hand, the Gaussian is not able to model either of the tails. Moreover, the Gaussian assigns non-zero probabilities to quantiles below zero, which is not the case for the data as temporal moments are non-negative random variables. Hence, the multivariate log-normal is a better choice. Note that a good marginal fit does not imply good overall fit in terms of AIC and vice-versa, as is the case for Lille data. This is simply because the AIC measures a different property of the model which does not require the marginals to fit perfectly.

The deviation of the right tail of the data from the fitted marginals is to be expected due to the low number of such extreme points. Such points are in-frequent and could potentially arise due to a number of factors such as noise, interference, measurement conditions, etc. Therefore, we argue that the right tail is not as important to model perfectly, and thus make no adjustment for it. However, this should be scrutinized further in applications where this effect could be important.

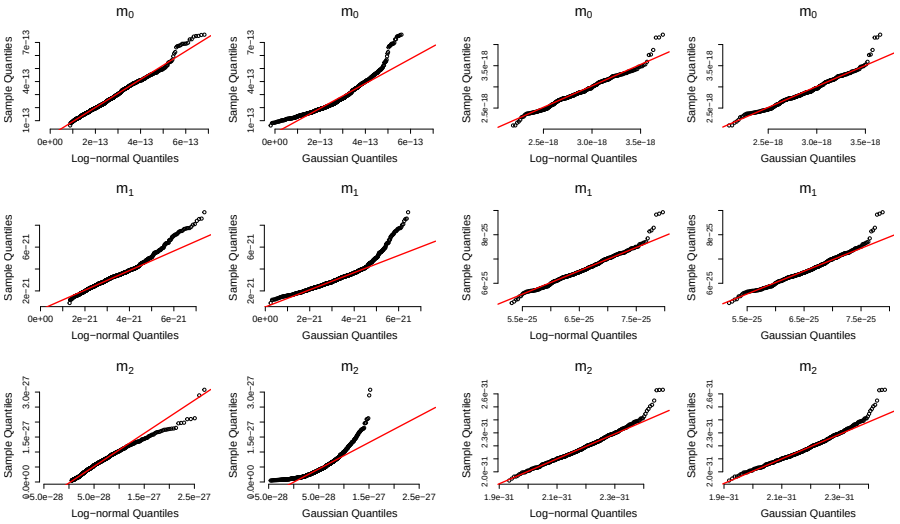


Fig. H.5: Quantiles of the measured raw temporal moments from Lille (left) and AAU-Outdoor (right) data versus the theoretical quantiles of fitted log-normal and Gaussian distributions. The theoretical quantile-quantile line passing through the first and third quantile is shown in red.

6 Model Fit to Raw Temporal Moments

The parameter estimates, obtained by fitting the proposed model to the data-sets using (H.9) and (H.10), are reported in Tab. H.2. We also compute and report the 95% confidence intervals for each of the estimates in Tab. H.2, see Appendix 8 for details. The confidence intervals are very small for the mean estimates, and an order of magnitude smaller for the covariance estimates. The fit of the proposed model to the various data-sets is shown in Fig. H.6 where each row corresponds to a particular data-set. The marginal distributions of the data and the fitted model is shown on the left, while 2D scatter plots for all pairs of temporal moments are shown on the right along with contour lines of the fitted distribution.

Firstly, we observe in Fig. H.6 that the distribution of the raw temporal moments varies across the different data-sets. This is attributed to the contrasting scenarios that the measurements were taken in, along with the use of different equipment, antennas, and measurement settings. We also observe that the raw temporal moments are highly correlated random variables. Marginal distributions for Lille and AAU-Industry data are skewed, while those from other data-sets are more symmetric. We notice a fanning out of the scatter plots on the top-right of all the indoor data-sets, which is not present in the outdoor data. Despite the variability in the data, the proposed model fits the data well, even the skewed ones. There is a high correlation between the raw

moments, in particular between m_0 and m_1, since the basis functions used to compute them in (23) are not orthogonal. This is captured well by the model.

7 Model Fit to Standardized Moments

We now compare the distribution of the standardized temporal moments obtained from the measurements with those from the proposed model. Mean delay and rms delay spread are computed from the raw temporal moments using (H.4), while the received power is equal to m_0. Pair-wise scatter plots of P_0, $\bar{\tau}$, and τ_{rms} from the data and the proposed joint model are shown in Fig. H.7. We also include the samples obtained from independently fitting a log-normal distribution to the standardized moments from the data-sets. The log-normal is chosen as it was the best in terms of AIC amongst the independent models as per Tab. H.3. Here we exclude the AAU-Industry data as the low number of sample points makes it difficult to make any useful conclusions on the correlation behavior. We observe in Fig. H.7 that the standardized temporal moments are also dependent random variables, and the proposed model is able to capture their dependency structure. In contrast, correlation information between the variables is lost when they are simulated independently.

Sample Pearson correlation coefficients between P_0, $\bar{\tau}$, and τ_{rms} from the data are given in Tab. H.4. For paired samples $\{(a_1, b_1), \ldots, (a_m, b_m)\}$, the sample Pearson correlation coefficient is defined as

$$\hat{\rho}_{a,b} = \frac{\sum_{j=1}^{m}(a_j - \bar{a})(b_j - \bar{b})}{\sqrt{\sum_{j=1}^{m}(a_j - \bar{a})^2}\sqrt{\sum_{j=1}^{m}(b_j - \bar{b})^2}}, \tag{H.12}$$

where $\bar{a}$ and $\bar{b}$ are the sample means. We also compute 95% confidence intervals for the correlation estimates using the bootstrap method [37, Chapter 6]. The correlation coefficients obtained from the fitted model, computed from 10,000 samples to get a robust estimate, are also reported in Tab. H.4. Mean delay and rms delay spread have a positive correlation that varies from 0.53 for the Lund data to as high as 0.97 for AAU-Outdoor. The received power is negatively correlated with both τ and τ_{rms}. In general, the correlation tends to increase with the size of the environment, with the outdoor case being highly correlated. The model is able to replicate the varying correlation between P_0 and τ_{rms} that is observed in the data, as opposed to having a fixed correlation coefficient suggested in [11]. Note that the correlation coefficient between $\bar{\tau}$ and τ_{rms} for the model fitted to the Lille data-set is not within the bootstrap interval. This is due to the banana-like shape of their scatter plot which is not replicated by the model, see Fig. H.7.

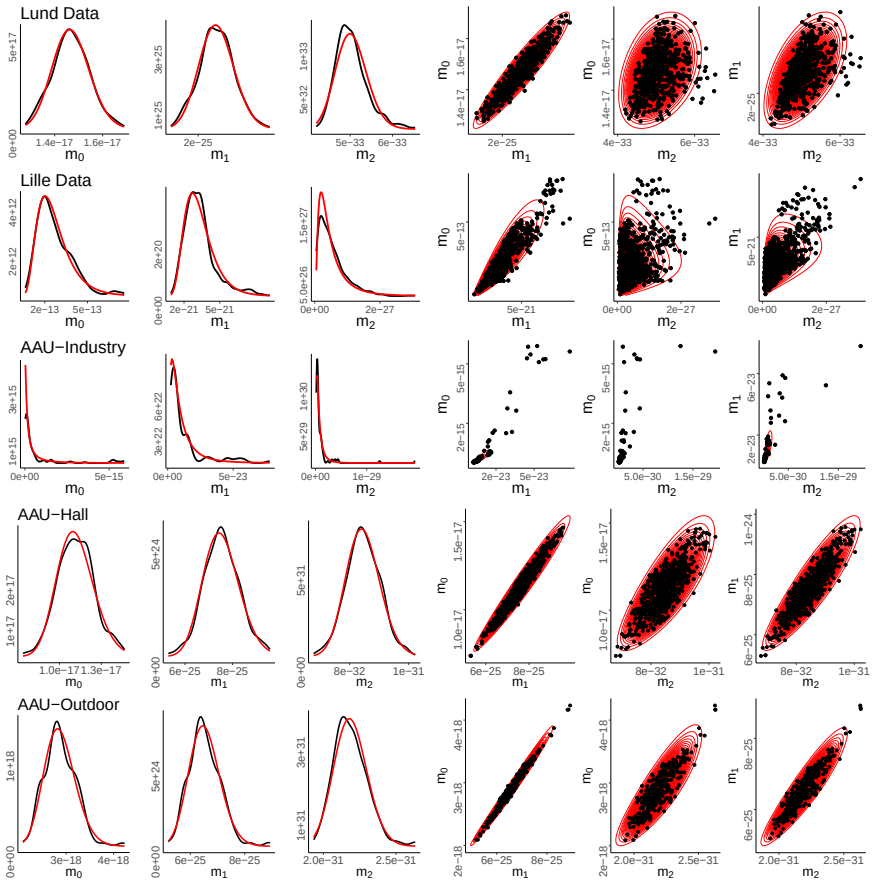


Fig. H.6: Density estimates and scatter plots of raw temporal moments obtained from the different measurements (shown in black) versus the density and contour plots of the fitted proposed model (shown in red). Each row corresponds to one of the data-sets. All the axes are in linear scale. The parameter estimates are in Tab. H.2.

Table H.4: Sample Pearson correlation coefficients between standardized temporal moments of measured data. The correlation coefficients for the model is computed using 10,000 samples of simulated data. The number in parenthesis (ϵ) is the 95% bootstrap confidence interval of the correlation estimates computed using 1000 re-samples, such that the interval is of the form $(\rho - \epsilon, \rho + \epsilon)$.

Data set	$\hat{\rho}_{P_0,\bar{\tau}}$		$\hat{\rho}_{P_0,\tau_{\text{rms}}}$		$\hat{\rho}_{\bar{\tau},\tau_{\text{rms}}}$	
	Data	**Model**	**Data**	**Model**	**Data**	**Model**
Lund	–0.28 (± 0.06)	–0.28	–0.35 (±0.05)	–0.36	0.53 (±0.05)	0.52
Lille	–0.48 (±0.03)	–0.51	–0.20 (±0.05)	–0.19	0.89 (±0.02)	0.83
AAU-Hall	–0.66 (±0.03)	–0.65	–0.87 (±0.02)	–0.87	0.70 (±0.03)	0.70
AAU-Outdoor	–0.91 (±0.01)	–0.92	–0.93 (±0.01)	–0.93	0.97 (±0.004)	0.97

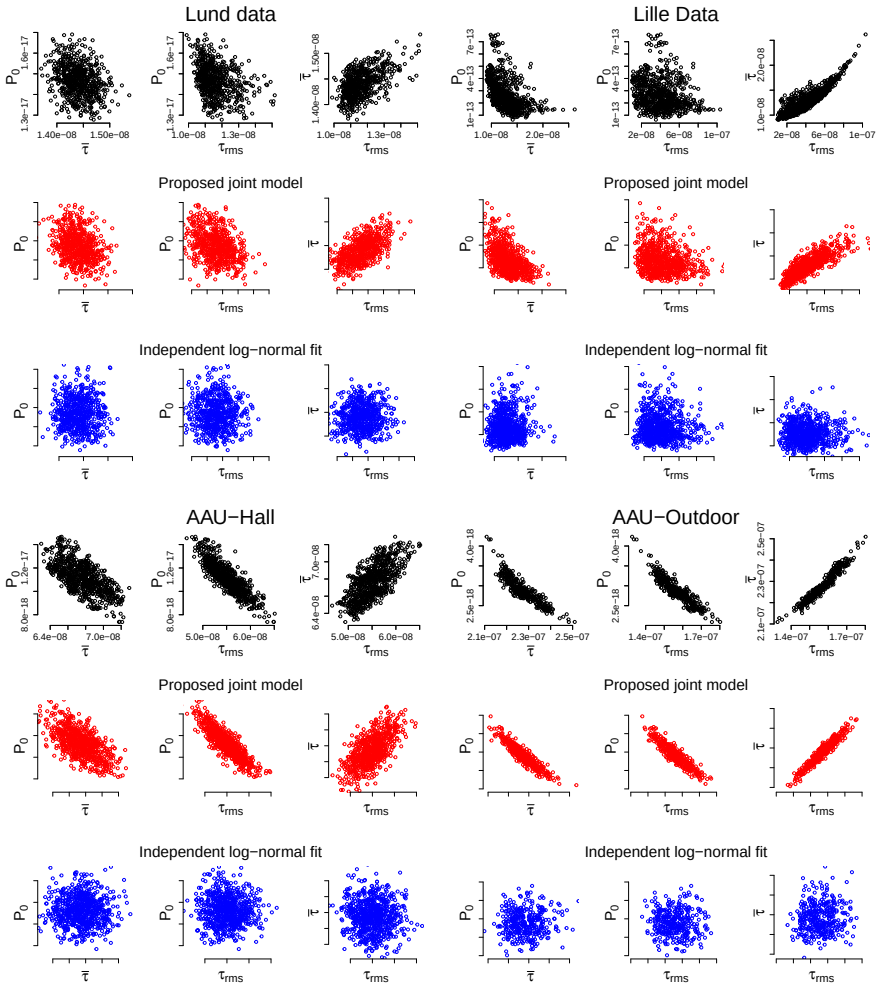


Fig. H.7: Scatter plots of received power, mean delay, and rms delay spread from data (in black), and from the proposed model (in red). The samples simulated by independently fitting log-normal marginals to P_0, $\bar{\tau}$, and τ_{rms} from the data are in blue. Number of points simulated is same as in the measurements. The scales of the corresponding plots are the same.

8 Conclusion

Joint modeling of received power, mean delay, and rms delay spread provides more accurate models in a range of scenarios as opposed to independent modeling. The proposed model of the multivariate log-normal distribution seems to be a reasonable choice for simulating these standardized moments, however the fit can be improved by using more complex models. The proposed model is simple, easy to simulate from, and easy to fit to new measurements in both indoor and outdoor settings using standard estimators. The raw temporal moments are dependent random variables which should be simulated jointly; as a result, the same is also true for the standardized temporal moments. The correlation of these moments changes from scenario to scenario, but can be inferred efficiently in each case.

In the light of the strong correlation observed in the measurements, assuming independence might lead to significant errors in some applications. Hence, reporting of the marginal distributions of the standardized moments is insufficient and a clearer picture can be obtained by considering both their means and covariances. The correlation between these standardized moments can be used to validate multipath models instead of just their marginal fits. The correlation should also be accounted for in the analysis and simulation of radio channels.

The means and covariances of the temporal moments potentially depend on a number of physical factors. The relation between the means and the transmitter-receiver distance has been studied for indoor scenarios. However, the effect of the distance on the covariance matrix is presently unclear. For multipath models, the covariance matrix is known to depend on the first- and second-order intensity functions which governs the arrival process. Since both intensity functions are affected by antenna directivity, the covariance matrix should also be. Nevertheless, these effects are not yet well-understood and should be the topic of further studies.

Acknowledgment

The authors would like to thank Dr. Carl Gustafson and Prof. Fredrik Tufvesson (Lund University) for providing the measurement data. This work is supported by the Danish Council for Independent Research, grant no. DFF 7017-00265 and performed within the framework of the COST Action CA15104 IRACON. Ramoni Adeogun was supported by grant no. DFF 9041- 00146B. François-Xavier Briol was supported by the Lloyds Register Foundation Programme on Data-Centric Engineering at The Alan Turing Institute under the EPSRC grant [EP/N510129/1].

Appendix: Parameter Inference for the Log-Normal

In this appendix, we recall how to derive the maximum likelihood estimates and related confidence intervals for a log-normal distribution. Let $Y = (Y_1, \ldots, Y_d)$ be a multivariate log-normal random variable. We will denote this distribution $\mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the parameters. Then, $X = (X_1, \ldots, X_d) = (\log(Y_1), \ldots, \log(Y_d))$ is a multivariate Gaussian random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Since the maximum likelihood estimator is invariant to one-to-one transformations of the data, we can simply take the logarithm of our data points and compute the maximum likelihood estimate corresponding to Gaussian data. Given N iid observations $\{\mathbf{y}_i\}_{i=1}^N$, we hence compute $\mathbf{x}_i = \log \mathbf{y}_i$ for $i = 1, \ldots, N$, and return the following estimates

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i, \quad \text{and} \tag{H.13}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}) (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\top . \tag{H.14}$$

Now, let the K free parameters be combined into a single vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$, where $\boldsymbol{\alpha} = (\mu_1, \ldots, \mu_d)$, and $\boldsymbol{\beta} = (\Sigma_{11}, \ldots, \Sigma_{dd})$. Note that $\Sigma_{ij} = \Sigma_{ji}$. The Fisher information matrix reads

$$I(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{bmatrix} I(\boldsymbol{\alpha}) & \mathbf{0} \\ \mathbf{0} & I(\boldsymbol{\beta}) \end{bmatrix} \tag{H.15}$$

where, for $1 \leq m, n \leq K$, the (m, n) entry of the matrix is

$$I(\boldsymbol{\alpha})_{m,n} = \frac{\partial \boldsymbol{\mu}^\top}{\partial \alpha_m} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \alpha_n}, \quad 1 \leq m, n \leq d \tag{H.16}$$

$$I(\boldsymbol{\beta})_{m,n} = \frac{1}{2} \mathrm{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \beta_m} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \beta_n} \right). \tag{H.17}$$

On further simplification, the entries of the Fisher information matrix become

$$I(\boldsymbol{\alpha})_{m,n} = \boldsymbol{\Sigma}^{-1}_{mn}, \tag{H.18}$$

$$I(\boldsymbol{\beta})_{m,n} = \frac{1}{2} \mathrm{tr} \left(\boldsymbol{\Sigma}^{-1} \mathbf{E}_m \boldsymbol{\Sigma}^{-1} \mathbf{E}_n \right), \tag{H.19}$$

where $\mathbf{E}_m$ is a $d \times d$ matrix of all zeros except the (i, i) entry corresponding to $\boldsymbol{\beta}_m = \Sigma_{ii}$ which is 1. Note that for $\boldsymbol{\beta}_m = \Sigma_{ij}, i \neq j$, both (i, j) and (j, i) entry of $\mathbf{E}_m$ will be 1. Same goes for $\mathbf{E}_n$. The 95% confidence interval for the m^{th} parameter of the Gaussian, $(\theta_m \pm \delta_m)$ is

$$\theta_m \pm \frac{1.96}{\sqrt{N}} \sqrt{\mathbf{I}^{-1}_{m,m}}$$

where $\mathbf{I}^{-1}_{m,m}$ is the (m, m) entry of $\mathbf{I}^{-1}$.

References

[1] A. Goldsmith, *Wireless Communications*. Cambridge University Press, Aug 2005.

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

[3] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter Estimation for Stochastic Channel Models using Temporal Moments," in *Proc. 2019 IEEE Int. Symp. on Antennas and Propag. and USNC-URSI Radio Sci. Meeting*, 2019.

[4] ——, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

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

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

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

[8] Z. Latinovic and H. Huang, "A channel model for indoor time-of-arrival ranging," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 2, pp. 1415–1428, feb 2020.

[9] M. K. Awad, K. T. Wong, and Z. Li, "An integrated overview of the open literature's empirical data on the indoor radiowave channel's delay properties," *IEEE Trans. on Antennas and Propag.*, vol. 56, no. 5, pp. 1451–1468, May 2008.

[10] D. Cox and R. Leck, "Distributions of multipath delay spread and average excess delay for 910 MHz urban mobile radio paths," *IEEE Trans. on Antennas and Propag.*, vol. 23, no. 2, pp. 206–213, March 1975.

[11] L. Greenstein, V. Erceg, Y. Yeh, and M. Clark, "A new path-gain/delay-spread propagation model for digital cellular channels," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 477–485, May 1997.

[12] J. Fischer, M. Grossmann, W. Felber, M. Landmann, and A. Heuberger, "A novel delay spread distribution model for VHF and UHF mobile-to-mobile channels," in *2013 7th Eur. Conf. on Antennas and Propag., EuCAP 2013*, Apr 2013, pp. 469–472.

[13] G. Wang, G. Zhu, S. Lin, J. Ding, D. Fei, and H. Zhang, "Channel measurement and modeling in highway scenario at 460 MHz," in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*. IEEE, jul 2019.

[14] 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. on Wireless Commun.*, vol. 18, no. 8, pp. 3856–3870, 2019.

[15] X. Zhang, G. Qiu, J. Zhang, L. Tian, P. Tang, and T. Jiang, "Analysis of millimeter-wave channel characteristics based on channel measurements in indoor environments at 39 GHz," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, oct 2019.

[16] Y. Yu, Y. Liu, W. Lu, and H. Zhu, "Measurement and empirical modelling of root mean square delay spread in indoor femtocells scenarios," *IET Commun.*, vol. 11, no. 13, pp. 2125–2131, 2017.

[17] M. Schmieder, T. Eichler, S. Wittig, M. Peter, and W. Keusgen, "Measurement and characterization of an indoor industrial environment at 3.7 and 28 GHz," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*. IEEE, mar 2020.

[18] J. Yu, W. Chen, F. Li, C. Li, K. Yang, Y. Liu, and F. Chang, "Channel measurement and modeling of the small-scale fading characteristics for urban inland river environment," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 5, pp. 3376–3389, may 2020.

[19] Y. Yu, W.-J. Lu, Y. Liu, and H.-B. Zhu, "Neural-network-based root mean delay spread model for ubiquitous indoor internet-of-things scenarios," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5580–5589, jun 2020.

[20] A. Prokes, T. Mikulasek, M. Waldecker, B. K. Engiz, and J. Blumenstein, "Multipath propagation analysis for static urban environment at 60 GHz," in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. IEEE, nov 2019.

[21] M. Varela and M. Sanchez, "RMS delay and coherence bandwidth measurements in indoor radio channels in the UHF band," *IEEE Trans on Veh. Technol.*, vol. 50, no. 2, pp. 515–525, mar 2001.

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

[23] T. Pedersen, "Modeling of path arrival rate for in-room radio channels with directive antennas," *IEEE Trans. on Antennas and Propag.*, vol. 66, no. 9, pp. 4791–4805, 2018.

[24] ——, "Stochastic Multipath Model for the In-Room Radio Channel Based on Room Electromagnetics," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 4, pp. 2591–2603, April 2019.

[25] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*. John Wiley & Sons, Inc., apr 2000.

[26] R. B. Nelsen, *An Introduction to Copulas*. Springer-Verlag GmbH, 2007. [Online]. Available: https://www.ebook.de/de/product/5270140/roger_b_nelsen_an_introduction_to_copulas.html

[27] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

[28] M. Fryziel, C. Loyez, L. Clavier, N. Rolland, and P. A. Rolland, "Path-loss model of the 60-ghz indoor radio channel," *Microw. and Opt. Technol. Lett.*, vol. 34, no. 3, pp. 158–162, 2002. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mop.10402

[29] M. Razzaghpour, R. Adeogun, I. Rodriguez, G. Berardinelli, R. S. Mogensen, T. Pedersen, P. Mogensen, and T. B. Sørensen, "Short-range UWB wireless channel measurement in industrial environments," in *2019 Int. Conf. on Wireless and Mobile Comput., Netw. and Commun. (WiMob)*, Oct 2019, pp. 1–6.

[30] A. Bharti, L. Clavier, and T. Pedersen, "Joint statistical modeling of received power, mean delay, and delay spread for indoor wideband radio channels," in *14th Eur. Conf. on Antennas and Propag.*, 2020, pp. 1–5.

[31] A. W. Mbugua, W. Fan, K. Olesen, X. Cai, and G. F. Pedersen, "Phase-compensated optical fiber-based ultrawideband channel sounder," *IEEE Trans. on Microw. Theory and Techn.*, vol. 68, no. 2, pp. 636–647, Feb 2020.

[32] X. Cai and W. Fan, "A complexity-efficient high resolution propagation parameter estimation algorithm for ultra-wideband large-scale uniform circular array," *IEEE Trans. on Commun.*, vol. 67, no. 8, pp. 5862–5874, Aug 2019.

[33] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. on Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec 1974.

[34] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging*. Cambridge University Press, 2008.

[35] J. Ding, V. Tarokh, and Y. Yang, "Model selection techniques: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 16–34, Nov 2018.

[36] J. A. Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. Duxbury Press, 1994.

[37] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, May 1994.

Paper I

A General Method for Calibrating Stochastic Radio Channel Models with Kernels

Ayush Bharti, François-Xavier Briol, Troels Pedersen

The paper has been submitted to the
IEEE Transactions on Antennas and Propagation, 2021.

Abstract

Calibrating stochastic radio channel models to new measurement data is challenging when the likelihood function is intractable. The standard approach to this problem involves sophisticated algorithms for extraction and clustering of multipath components, following which, point estimates of the model parameters can be obtained using specialized estimators. We propose a likelihood-free calibration method using approximate Bayesian computation. The method is based on the maximum mean discrepancy, which is a notion of distance between probability distributions. Our method not only by-passes the need to implement any high-resolution or clustering algorithm, but is also automatic in that it does not require any additional input or manual pre-processing from the user. It also has the advantage of returning an entire posterior distribution on the value of the parameters, rather than a simple point estimate. We evaluate the performance of the proposed method by fitting two different stochastic channel models, namely the Saleh-Valenzuela model and the propagation graph model, to both simulated and measured data. The proposed method is able to estimate the parameters of both the models accurately in simulations, as well as when applied to 60 GHz indoor measurement data.

1 Introduction

Stochastic channel models are used to simulate the behavior of the radio channel in order to test the performance of communication and localization systems. Often models are flexible enough to be applied to different scenarios, provided that their parameters can be adjusted accordingly. Adjustment of the model parameters based on data collected from measurement campaigns is called *calibration* (or inference). Calibration is usually challenging since most state-of-the-art stochastic radio channel models have intractable likelihood functions. This renders usual inference techniques such as maximum likelihood estimation or standard Bayesian inference inapplicable.

Instead of solving the whole calibration problem at once, it is wide-spread practice (e.g. [1–9]) to split the task into intermediate steps as outlined in Fig. I.1(a). The first step involves resolving the multipath components, i.e. estimating path parameters including delays, directions, and complex gains. This task can be carried out using high-resolution algorithms such as MUSIC, SAGE, and RiMAX, among others, see e.g [10, Ch. 5] for an overview. The second step is clustering of the extracted multipath components in the case of cluster-based models. Clustering is either performed manually, as in [2], or using automated algorithms such as [11–13]. In a final step, the model parameters are estimated from the extracted and clustered multipath components.

Despite being widely applied, the multi-step approach suffers from a range of issues, owing to the composite nature of the methodology. In particular, high-resolution and clustering methods, although very useful in analyzing and

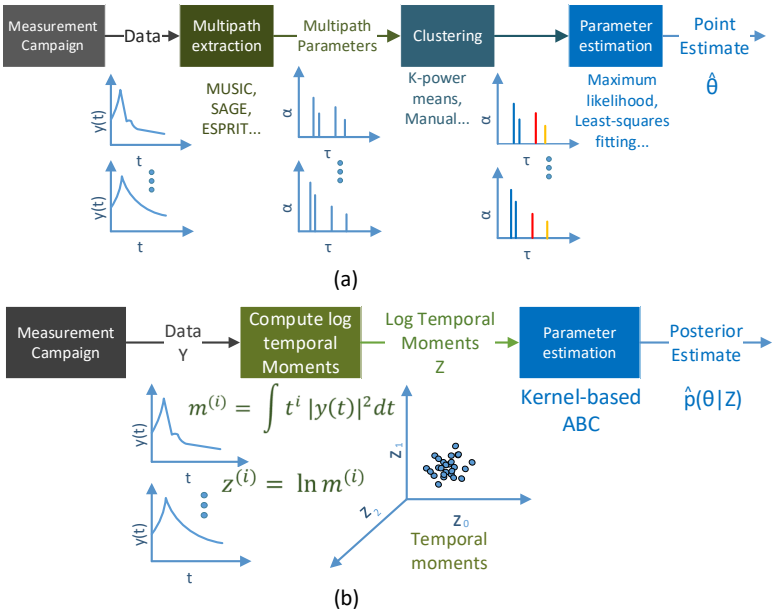


Fig. I.1: Methodologies for calibration of stochastic radio channel models: (a) State-of-the-art methodology based on multipath extraction and clustering; (b) proposed method based on generic summaries (here exemplified by log-temporal moments) and approximate Bayesian computation (ABC).

understanding the radio channel, are problematic when it comes to model calibration. These methods require implementation of sophisticated and specialized algorithms at each step, which involves a number of heuristic choices and settings which might be conflicting. An emblematic example is the assumption of "well separated" paths while extracting multipath components. The high-resolution methods are prone to estimation artifacts, especially if paths are not "well separated". However, this conflicts with the inherent assumption in the clustering step that multipaths arrive "close" to each other. Consequently, even though the performance of high-resolution and clustering algorithms are thoroughly investigated in isolation, the accuracy of the applied multi-step calibration techniques is unknown. Moreover, the calibration technique needs to be tailored to the particular model at hand. While attempting to calibrate and compare different ultra-wideband models using a large database, Greenstein et al. in [14] noted that *"the problem in doing so is that there is no simple, clear and established method for extracting cluster model parameters from measured data"*. As a result, they were unable to fit the cluster model to their calibration data.

Calibration methods that by-pass the need to resolve the multipath components have been recently proposed. They have been used to calibrate the Turin

model [1], the Saleh-Valenzuela (S-V) model [2] and the polarized propagation graph (PG) model [15]. These calibration methods rely either on a Monte Carlo approximation of the likelihood [16, 17], the method of moments [18, 19], or a summary-based likelihood-free inference framework [20–23] such as approximate Bayesian computation (ABC). First developed in the field of population genetics in 1997, ABC has since become a popular method for calibrating models with intractable likelihoods in various fields, see [24] for an overview. The main drawback of the calibration methods [17–19] is their reliance on equations that explicitly link the moments of the summaries with the model parameters, or in case of [16], on the model-specific point process. These methods should therefore be re-derived for each new model. We encounter this to be a non-trivial task, and it may not even be possible for the more elaborate channel models. Similar problems exist in [20–22] where a low-dimensional vector of statistics should be redesigned or trained using an autoencoder [23] for the channel model at hand, which is not always trivial and may not generalize to other models. Moreover, summarizing the data leads to information loss that can hamper the accuracy of the parameter estimates.

The aim of the present contribution is to propose a general method which can be applied to stochastic channel models of very different mathematical structure. This will be done without the need for specializing summaries, or extraction and clustering of multipaths. To achieve this, we follow the proposed calibration methodology depicted in Fig. I.1(b). First, we map the channel measurements into easily computable log temporal moments. These moments are then used for calibration in an ABC framework, where we use the maximum mean discrepancy (MMD) [25] to compare the distribution of simulated and measured data. The MMD has previously been used for frequentist inference in [26, 27], and in a Bayesian sense in [28]. Specific ABC methods using kernels include [29–32], and the MMD has also been used to train generative adversarial networks in [33–35]. These papers have shown MMD to be a powerful way to represent either data-sets or distributions, and as a result calibrate complex models. They have acted as inspiration for our work, but our algorithm specializes the approach to the problem of calibrating stochastic channel models. Our calibration method is automatic since it can be applied to different models without the need for further pre- or post- processing. Additionally, the method is able to account for model misspecification, which occurs when the model is not able to represent the data for any parameter setting.

The rest of the paper is organized as follows. Section 2 presents the model calibration problem. Section 3 gives an overview of the MMD, and Section 4 describes the proposed kernel-based ABC method. We demonstrate the method's generality by calibrating the seminal S-V model, which is a clustered multipath model, and the propagation graph model, which is based on a different principle, using exactly the same data and procedure. Indeed, no other method able to do this is available in the open literature. In Section 5, the performance is evaluated on simulated data and in Section 6 on data from a 60 GHz indoor measurement campaign. We find that the S-V model is misspecified for

the considered measurements, and hence fails to replicate its characteristics. Discussion and concluding remarks are given in Sections 7 and 8, respectively.

2 Stochastic Channel Model Calibration

Consider the transfer function measurement of a linear, time-invariant radio channel in a single-input, single-output (SISO) setup using a vector network analyzer (VNA). The transfer function is measured at N_s equidistant frequency points in the bandwidth B, resulting in a frequency separation of $\Delta f = B/(N_s - 1)$. The measured complex signal at the n^{th} frequency point, Y_n, is modeled as

$$Y_n = H_n + W_n, \quad n = 0, 1, \cdots, N_s - 1, \tag{I.1}$$

where H_n is the transfer function sampled at the n^{th} frequency and W_n is the complex measurement noise. The additive noise samples are assumed independent and identically distributed (iid) at each frequency point, and are usually modeled as zero-mean circular symmetric complex Gaussian variables with variance σ_W^2. The time-domain signal, $y(t)$, is obtained by taking the discrete-frequency, continuous-time inverse Fourier transform of Y_n as

$$y(t) = \frac{1}{N_s} \sum_{n=0}^{N_s-1} Y_n \exp(j2\pi n \Delta f t), \tag{I.2}$$

periodic with a period of $t_{\max} = 1/\Delta f$. Multiple realizations of the channel can be obtained by repeating the measurements N_{obs} times, yielding an $N_{\text{obs}} \times N_s$ complex data matrix $\mathbf{Y}$. The data can be thought of as iid realizations from some unknown distribution, $\mathbb{Y}$, which is the true state of nature.

A stochastic model can be seen as a parametric family of distributions $\{\mathbb{P}_{\boldsymbol{\theta}}\}$ with a p-dimensional parameter vector $\boldsymbol{\theta}$ defined on some Euclidean space[1]. In the case of generative models such as the stochastic channel models, it is straightforward to simulate realizations of $\mathbf{Y}$ from the model, even though the distribution $\mathbb{P}_{\boldsymbol{\theta}}$ is unknown. Calibration then amounts to finding the $\boldsymbol{\theta}$ for which the model output fits the observed data $\mathbf{Y}$ well, or in other words, to find the $\boldsymbol{\theta}$ such that $\mathbb{P}_{\boldsymbol{\theta}}$ is "closest" to $\mathbb{Y}$. Standard calibration techniques involve the likelihood function of the model given $\mathbf{Y}$. For iid realizations, the likelihood function, denoted as $p(\mathbf{Y}|\boldsymbol{\theta})$, is the product of the probability density or mass function of $\mathbb{P}_{\boldsymbol{\theta}}$ evaluated at each of the data points in $\mathbf{Y}$.

For most stochastic radio channel models, $p(\mathbf{Y}|\boldsymbol{\theta})$ is either intractable or cannot be approximated within reasonable computation time. Intractability

[1]The restriction to parameters in $\mathbb{R}^p$ is only needed in the adjustment method described in Section 4.2. The remaining part of the method can be used for more general parameter spaces, e.g. discrete, complex or subsets of $\mathbb{R}^p$. In this case, the adjustment algorithm should be modified to either accommodate or ignore such parameters.

here refers to the inability to numerically evaluate the likelihood function for a given value of $\boldsymbol{\theta}$. For intractable likelihood, the posterior, $p(\boldsymbol{\theta}|\mathbf{Y})$, also becomes intractable as it is proportional to $p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta})$ is the prior assumed on the parameters. An intractable likelihood prevents maximum likelihood estimation of $\boldsymbol{\theta}$ as well as Bayesian inference via sampling of the posterior. This is the case for stochastic multipath models, such as the Turin and the S-V model, which were constructed with the ease of simulation in mind.

Since stochastic channel models are easy to simulate from given an arbitrary $\boldsymbol{\theta}$ value, likelihood-free inference is possible by comparing simulated data-sets to the observed data. Therefore, we need a method to compute distances between the data-sets which is challenging as the data-sets are high-dimensional, and may have possibly different sizes. We tackle this problem using distance metrics based on kernels, in particular the maximum mean discrepancy (MMD).

3 The Maximum Mean Discrepancy between Probability Distributions

We now introduce the MMD which is a notion of distance between arbitrary probability distributions $\mathbb{P}$ and $\mathbb{Q}$ or data-sets. We aim to use MMD as a similarity measure within an ABC framework to compare simulated and observed data-sets. Note that we can identify any data-set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ to an empirical distribution $\frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i}$ where $\delta_{\mathbf{x}_i}$ denotes a distribution with mass one at $\mathbf{x}_i$ and 0 otherwise. We restrict our discussion to distributions defined on $\mathbb{R}^d$. This section will provide further details on constructing the MMD [25, 36].

3.1 Kernels and The Maximum Mean Discrepancy (MMD)

The MMD consists of first mapping the distributions to a function space $\mathcal{H}_k$, then using the distance in that space to compare the mapped distributions. See Fig. I.2 for an illustration. The mapping enables the use of distance defined on $\mathcal{H}_k$.

The spaces of functions to which we will map distributions are called reproducing kernel Hilbert space (RKHS). We denote the RKHS with $\mathcal{H}_k$, and $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$ and $\| \cdot \|_{\mathcal{H}_k}$ for its inner product and norm, respectively. Associated to each RKHS, there exists a symmetric and positive definite function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ called a reproducing kernel [37]. This function satisfies two properties: (i) for all $f \in \mathcal{H}_k$, $f(\mathbf{x}) = \langle f, k(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_k}$ (called the reproducing property), and (ii) $k(\mathbf{x}, \cdot) \in \mathcal{H}_k$ for all $\mathbf{x} \in \mathbb{R}^d$.

It is straightforward to map probability distribution $\mathbb{P}$ to $\mathcal{H}_k$ through what is called a *kernel mean embedding* defined as

$$\mu_{\mathbb{P}}(\cdot) = \mathbb{E}_{X \sim \mathbb{P}}[k(X, \cdot)] = \int_{\mathbb{R}^d} k(\mathbf{x}, \cdot) \mathbb{P}(\mathrm{d}\mathbf{x}), \tag{I.3}$$

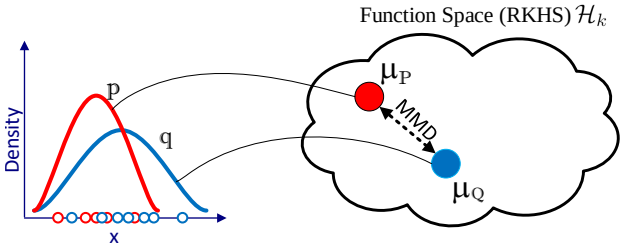


Fig. I.2: Given a kernel k, the distributions $\mathbb{P}$ and $\mathbb{Q}$ are mapped to their kernel mean embeddings $\mu_{\mathbb{P}}$ and $\mu_{\mathbb{Q}}$ using Equation I.3. The MMD is obtained by computing the distance between $\mu_{\mathbb{P}}$ and $\mu_{\mathbb{Q}}$ in the function space $\mathcal{H}_k$, as expressed in Equation I.5. This figure is inspired by [36].

under mild regularity conditions satisfied for all kernels in this paper, see [25, Lemma 3]. Here, $\mathbb{E}[\cdot]$ denotes the expectation with respect to the random variable and probability distribution given in subscript. Note that, $\mu_{\mathbb{P}} \in \mathcal{H}_k$. In the case where the probability distribution $\mathbb{P}$ has a probability density function p, the integral in (I.3) can be written in the more wide-spread form $\int_{\mathbb{R}^d} k(\mathbf{x}, \cdot)p(\mathbf{x})\mathrm{d}\mathbf{x}$. Alternatively, when $\mathbb{P}$ is an empirical distribution corresponding to a data-set, then the kernel mean embedding is given by $\frac{1}{N_X}\sum_{i=1}^{N_X} k(\mathbf{x}_i, \mathbf{x})$.

The MMD between probability distributions $\mathbb{P}$ and $\mathbb{Q}$ embedded in $\mathcal{H}_k$ is defined as the supremum taken over the mean of all functions in the unit ball in an RKHS, i.e. [36]

$$\mathrm{MMD}_k[\mathbb{P}, \mathbb{Q}] = \sup_{\|f\|_{\mathcal{H}_k} \leq 1} \left| \mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \mathbb{E}_{X\sim\mathbb{Q}}[f(X)] \right|. \tag{I.4}$$

As the name suggests, the MMD is the maximum distance between means of (unit norm) functions computed with respect to the distributions $\mathbb{P}$ and $\mathbb{Q}$. As shown in [25], the MMD in (I.4) can equivalently be expressed as

$$\begin{aligned} \mathrm{MMD}_k[\mathbb{P}, \mathbb{Q}] &= \left\| \mathbb{E}_{X\sim\mathbb{P}}[k(X, \cdot)] - \mathbb{E}_{Y\sim\mathbb{Q}}[k(Y, \cdot)] \right\|_{\mathcal{H}_k} \\ &= \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}_k} \end{aligned} \tag{I.5}$$

This gives an alternative interpretation of the MMD as the distance between mean embeddings in $\mathcal{H}_k$ as Fig. I.2 illustrates.

A third expression for the MMD appears upon expanding the squared norm in (I.5) and using the reproducing property of k which yields an expression in terms of k as

$$\begin{aligned} \mathrm{MMD}_k^2[\mathbb{P}, \mathbb{Q}] = \mathbb{E}_{X,Y\sim\mathbb{P}}[k(X, Y)] \\ - 2\mathbb{E}_{X\sim\mathbb{P},Y\sim\mathbb{Q}}[k(X, Y)] + \mathbb{E}_{X,Y\sim\mathbb{Q}}[k(X, Y)]. \end{aligned} \tag{I.6}$$

The latter expression is computationally more appealing than the two former as it only calls for computation of expectations of the kernel. Thus, computation

of the supremum in (I.4) is not required to compute the MMD. As discussed later in Section 3.3, the expression (I.6) forms the basis for estimation of the MMD from data.

3.2 Selecting a Kernel

The choice of kernel defines the RKHS and thus the properties of its distance, the MMD. In addition to being reproducing, it is a great advantage if the kernel is *characteristic* [38, 39]. This implies that the kernel mean embedding is an injective mapping, meaning that each distribution is mapped to a unique function. Thus, in the case of characteristic kernels, the kernel mean embedding captures all the information about the distribution. As a result, $\mathrm{MMD}_k[\mathbb{P}, \mathbb{Q}] = \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}_k} = 0$ if and only if $\mathbb{P} = \mathbb{Q}$. In this case, the MMD is capable of comparing infinitely many moments of two probability distributions without ever having to compute these moments explicitly. Consequently, the MMD is able to distinguish probability distributions even when these coincide in finite number of moments. This gives a great advantage over methods based on comparison of finitely many moments which are potentially blind to differences between distributions.

A very popular characteristic reproducing kernel is the squared-exponential (or Gaussian) kernel, defined as

$$k_{\mathrm{SE}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{l^2}\right), \tag{I.7}$$

for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$. Here, $\|\cdot\|_2$ is the Euclidean norm and $l > 0$ is a parameter called the lengthscale of the kernel. The norm inside the exponent can be chosen based on the specific data and application. For additional examples of characteristic kernels, see [38, 39].

We now give a simple example comparing Gaussian distributions, in which case the MMD can be derived analytically.

Example

Let $\mathbb{P} = \mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathbb{Q} = \mathcal{N}(\mu_2, \sigma_2^2)$ be two Gaussian distributions on $\mathbb{R}$. For the squared-exponential kernel in (I.7), the MMD takes the form (see [40, Appendix C]):

$$\begin{aligned}\mathrm{MMD}^2_{k_{\mathrm{SE}}}[\mathbb{P}, \mathbb{Q}] = \frac{l}{l + 2\sqrt{2}\sigma_1} + \frac{l}{l + 2\sqrt{2}\sigma_2} \\ - \frac{2l}{l + \sqrt{2}\sigma_1 + \sqrt{2}\sigma_2} \exp\left(-\frac{(\mu_1 - \mu_2)^2}{l^2 + 2\sigma_1^2 + 2\sigma_2^2}\right).\end{aligned} \tag{I.8}$$

It is apparent from (I.8) that the MMD is zero if and only if $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$ (as guaranteed by using a characteristic kernel). Fig. I.3 illustrates

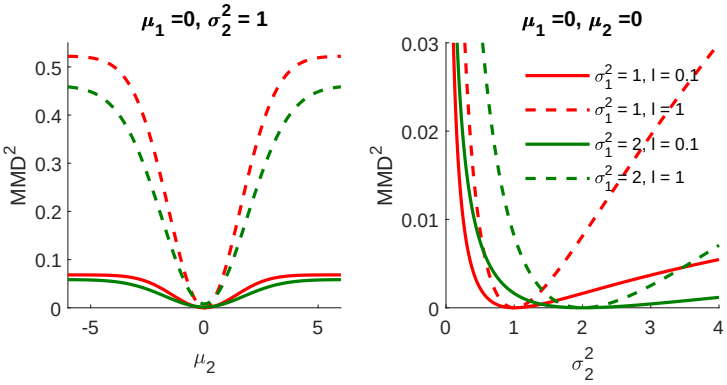


Fig. I.3: MMD between a $\mathcal{N}(\mu_1, \sigma_1^2)$ distribution with $\mu_1 = 0$ and $\sigma_1 = 1$ and a $\mathcal{N}(\mu_2, \sigma_2^2)$ varying parameters μ_2 and σ_2. Plots are shown for two different values of the lengthscale l.

how the MMD increases as the parameters of these distributions increasingly differ. Varying the lengthscale, l, of the kernel scales the overall MMD curve, but does not affect the point at which the MMD is minimised. The overall behaviour of the curves do not vary significantly on changing the lengthscale by an order of the magnitude.

3.3 Maximum Mean Discrepancy Between Data-sets

Unlike in the previous example, it is, in most realistic cases, not feasible to analytically calculate (I.6). Moreover, numerical integration is problematic, as the dimension of X and Y may be large and $\mathbb{P}$ or $\mathbb{Q}$ unavailable. Fortunately, it is straightforward to estimate the MMD if it is an empirical distribution, such as in the case of data-sets.

Imagine that we do not have access to $\mathbb{P}$ and $\mathbb{Q}$, but that we instead have two data-sets consisting of realisations from these distributions. More precisely, suppose we have access to $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_X}\} \overset{iid}{\sim} \mathbb{P}$ and $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{N_Y}\} \overset{iid}{\sim} \mathbb{Q}$. Then, an unbiased empirical estimate of $\mathrm{MMD}_k^2[\mathbb{P}, \mathbb{Q}]$ can be obtained as [25]

$$\widehat{\mathrm{MMD}}_k^2[\mathbf{X}, \mathbf{Y}] = \frac{\sum_{i \neq i'} k(\mathbf{x}_i, \mathbf{x}_{i'})}{N_X(N_X - 1)} - \frac{2\sum_{j=1}^{N_Y}\sum_{i=1}^{N_X} k(\mathbf{x}_i, \mathbf{y}_j)}{N_Y N_X} + \frac{\sum_{j \neq j'} k(\mathbf{y}_j, \mathbf{y}_{j'})}{N_Y(N_Y - 1)}. \quad \text{(I.9)}$$

Note that N_X and N_Y are permitted to differ, i.e. the two data-sets are not limited to be of the same size. To use this estimator with the kernel in (I.7), the lengthscale should be specified. Following [25], the lengthscale can be set

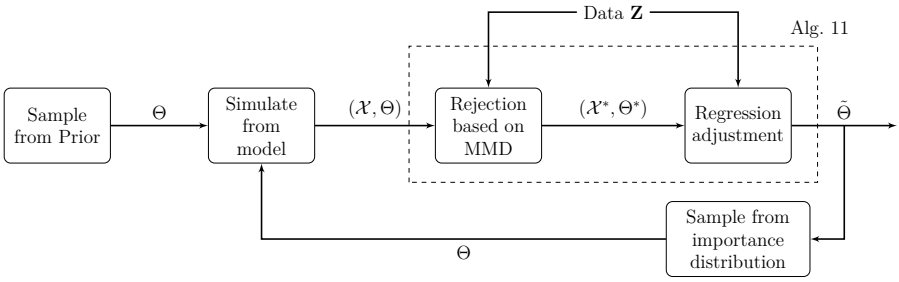


Fig. I.4: Diagram depicting steps in the proposed kernel-based ABC algorithm with regression adjustment described in Alg. 12. The block "Rejection based on MMD" corresponds to Sec. 4.1, "Regression adjustment" corresponds to Sec. 4.2, and "Sample from importance distribution" corresponds to Sec. 4.3. Here the term "Data" can be either obtained from physical measurements or as in Sec. 5 by simulation.

based on the data-set $\mathbf{X}$ using the median heuristic

$$l \;=\; \sqrt{\mathrm{med}/2}, \tag{I.10}$$

where med denotes the median of the set of squared two-norm distances $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$ for all pairs of distinct data points in $\mathbf{X}$. This setting of l scales the kernel with the spread of the data, and is robust to outliers.

Concentration bounds for MMD, such as [26, Lemma 1] or [36, Theorem 3.4], imply that with high probability,

$$\left|\widehat{\mathrm{MMD}}_k^2[\mathbf{X}, \mathbf{Y}] - \mathrm{MMD}_k^2[\mathbb{P}, \mathbb{Q}]\right| \leq C\left(\frac{1}{N_X} + \frac{1}{N_Y}\right), \tag{I.11}$$

for some $C > 0$. This tells us that the accuracy of the estimate converges linearly in both N_X and N_Y. The computational cost of computing this estimate is $\mathcal{O}(N_X^2 + N_Y^2)$ due to the need to compute double sums in both N_X and N_Y. In order to best balance computational cost and accuracy, N_X and N_Y should be chosen to be commensurate. These two results on accuracy and computational cost can be used to determine how to make default choices for the parameters of our ABC algorithm.

3.4 Kernels for Radio Channel Measurements

In order to use the MMD for calibrating stochastic radio channel models, we need a kernel defined on the space of transfer function measurements: $k_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$. Given such a kernel, we could then estimate the MMD between a measured data-set $\mathbf{Y}$ and a data-set $\mathbf{Y}_{\mathrm{sim}}$ simulated from the model.

A significant challenge with this approach is that, in the context of stochastic radio channel models, $\mathcal{Y}$ is usually a high-dimensional space. This is especially the case for large bandwidth measurements where N_s can be in the order

of thousands. Such high-dimensional problems are challenging for kernel methods based on default kernels such as the squared-exponential kernel [41]. These kernels indeed suffer from the curse-of-dimensionality, a phenomenon implying that the distance between points increases exponentially with the dimension of the space.

To tackle this issue, there exist kernels specialised to certain time-series or functional data models in the literature [42–47]. These use specific properties of the type of data in order to avoid the curse-of-dimensionality. In this paper, we contribute to this literature and construct a kernel specifically tailored to transfer function measurements. We base the kernel on the temporal moments of $y(t)$, defined as

$$m^{(i)} = \int_0^{t_{\max}} t^i |y(t)|^2 \mathrm{d}t, \quad i = 0, 1, 2, \ldots, I. \tag{I.12}$$

The integral in (I.12) is easy to compute numerically. The temporal moments can be seen as an expansion of $|y(t)|^2$ into the basis of monomials. Since the monomials form a complete basis for finite energy time-limited signals [48], no information is lost compared to $|y(t)|^2$ if $I \to \infty$. Referring to [49, 50], the first few moments are well modeled by a log-normal distribution. Thus, taking the entry-wise logarithm $z^{(i)} = \ln m^{(0)}$ brings the moments to the same scale and gives an approximately Gaussian vector $\mathbf{z} = [z^{(0)}, \ldots, z^{(I-1)}]$. Multiple channel realizations yield $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N_{\text{obs}}})$.

Define the mapping $A_I : \mathcal{Y} \to \mathbb{R}^I$ from $\mathcal{Y}$ to the I-dimensional space of log temporal moments. We propose to construct a kernel $k_{\mathcal{Y}}$ for transfer function data as

$$k_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}') := k_{\text{SE}}(A_I(\mathbf{y}), A_I(\mathbf{y}')), \quad \text{for all } \mathbf{y}, \mathbf{y}' \in \mathcal{Y}, \tag{I.13}$$

where k_{SE} denotes the squared-exponential kernel in dimension I. We note that this is the composition of a reproducing kernel and a map, and thus according to [51, Lemma 4.3] is a reproducing kernel on $\mathcal{Y}$. We also note that the MMD with kernel $k_{\mathcal{Y}}$ computed on the original data can be obtained through the MMD with kernel k_{SE} on the log temporal moments. Similarly, the empirical estimators of these quantities are also identical, i.e.

$$\widehat{\text{MMD}}^2_{k_{\mathcal{Y}}}[\mathbf{Y}, \mathbf{Y}_{\text{sim}}] = \widehat{\text{MMD}}^2_{k_{\text{SE}}}[\mathbf{Z}, \mathbf{X}], \tag{I.14}$$

where $\mathbf{X}$ is the simulated log temporal moments data-set.

In practice, we will have to limit ourselves to a finite I for computational reasons. This, however, is not a problem since we can expect the signal energy to be concentrated on the lowest moments. In fact, taking I to be small also allows us to by-pass issues with the curse-of-dimensionality.

From a theoretical viewpoint, since the squared-exponential kernel is characteristic, we should be able to recover any distribution on the space of log temporal moments. However, since the mapping A_I leads to loss of information when I is finite, $k_{\mathcal{Y}}$ will not be characteristic on $\mathcal{Y}$, and we may not be

able to uniquely identify the distribution on $|y(t)|^2$. This however is not an issue for the considered channel models, as will be shown in Section 5.

4 Proposed Kernel-based Approximate Bayesian Computation Method

ABC methods rely on simulation from the model to approximate the posterior, and can be used to estimate $\boldsymbol{\theta}$ such that the model fits to the observed data $\mathbf{Y}$. Let $\rho(\cdot,\cdot)$ be some notion of distance between data-sets. The basic form of ABC, called rejection ABC, proceeds by sampling M parameter values from $p(\boldsymbol{\theta})$ and generating the corresponding simulated data $\mathbf{Y}_{\text{sim}}$ from the model. The values of $\boldsymbol{\theta}$ for which $\rho(\mathbf{Y}, \mathbf{Y}_{\text{sim}})$ is less than some pre-defined threshold ϵ, form a sample from the approximate posterior distribution, $\tilde{p}(\boldsymbol{\theta}|\mathbf{Y}) = p(\boldsymbol{\theta}|\rho(\mathbf{Y}, \mathbf{Y}_{\text{sim}}) < \epsilon)$. The tolerance threshold impacts the degree of approximation in ABC methods. Setting $\epsilon = 0$ would lead to exact Bayesian inference, however, achieving equality for continuous-valued data is not possible. Hence, ϵ should be small but non-zero in order to be computationally feasible.

We now propose an ABC method based on the MMD as the distance metric to calibrate stochastic radio channel models. We employ the Population Monte Carlo (PMC) ABC method [52] to iteratively refine our approximation of the ABC posterior. At the end of each iteration, we perform local-linear regression adjustment [53] to further improve the posterior approximation. The complete algorithm is depicted in Fig. I.4 and outlined in Alg. 12. Individual steps of this PMC-ABC algorithm will be highlighted in Sec. 4.1 to 4.3. In Sec. 4.4, we describe how to detect and account for model misspecification in the algorithm.

4.1 Rejection based on MMD

The proposed ABC method uses the MMD between data-sets as a rejection criteria. Instead of setting the threshold ϵ in terms of the distance, we specify the proportion of accepted samples, i.e. $\epsilon = M_\epsilon/M$ where M_ϵ is the number of parameter samples accepted out of M. This is particularly convenient as it avoids the need to manually find a threshold, which may lead to unknown run-time of the algorithm.

The method computes $\widehat{\text{MMD}}^2_{k_{\text{SE}}}[\mathbf{X}, \mathbf{Z}]$, where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\text{sim}}})$ is the simulated log temporal moments data-set, as this is identical to estimating the MMD between $\mathbf{Y}$ and $\mathbf{Y}_{\text{sim}}$ (see Eq. I.14). First, M independent parameter samples $\Theta = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ are drawn from the prior $p(\boldsymbol{\theta})$. For each $\boldsymbol{\theta}_i$, the log temporal moments data-set, $\mathbf{X}_i \sim \mathbb{P}_{\boldsymbol{\theta}_i}$, is simulated. The simulated data-sets are gathered in $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_M)$. The $\widehat{\text{MMD}}^2_{k_{\text{SE}}}[\mathbf{X}_i, \mathbf{Z}]$ is computed for each i using (I.9), setting the lengthscale of k_{SE} as per (I.10). The parameter samples resulting in the M_ϵ smallest MMD values are then accepted.

In principle, the MMD could be computed between the samples of the temporal moments instead of their logarithm. However, the magnitudes of the different temporal moments may vary strongly and using a single lengthscale may lead to poor performance. Using a log transformation helps mitigate this issue. Alternatively, the lengthscale should be defined for each dimension of $\boldsymbol{\theta}$.

4.2 Regression Adjustment

As proposed in [53], it is possible improve the posterior approximation by adjusting the accepted samples using a model of the relationship between a low-dimensional vector of statistics and the parameter vector. Let $\mathbf{s}$ be a vector of summary statistics of $\mathbf{X}$ such that $\mathbf{s} = S(\mathbf{X})$ for a function $S(\cdot)$. Similarly, the observed summary statistics are denoted $\mathbf{s}_{\text{obs}} = S(\mathbf{Z})$. We begin by fitting a function, g, between the accepted parameters $\Theta^* = (\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_{M_\epsilon}^*)$ and the corresponding statistics $\mathcal{S}^* = (\mathbf{s}_1, \ldots, \mathbf{s}_{M_\epsilon})$ as [24, Ch. 3]

$$\boldsymbol{\theta}_i = g(\mathbf{s}_i) + \varepsilon, \quad i = 1, \ldots, M_\epsilon, \tag{I.15}$$

where $g(\mathbf{s})$ is the conditional expectation of $\boldsymbol{\theta}$ given $\mathbf{s}$, and ε is the residual. Here, $\boldsymbol{\theta}$ should belong to a subset of $\mathbb{R}^p$. Considering that the log of the temporal moments are well modeled by a Gaussian distribution, we take $\mathbf{s}_{\text{obs}}$ to be the vector consisting of the sample means and sample covariances of the elements of $\mathbf{z}$, similar to [22]. In total, $\mathbf{s}_{\text{obs}}$ consists of $(I^2 + 3I)/2$ elements for I temporal moments. The statistics $\mathbf{s}$ is computed in the same manner for $\mathbf{X}$. Note that $\mathbf{s}$ and $\mathbf{s}_{\text{obs}}$ are normalized by an estimate of their median absolute deviation to account for the difference in magnitude of the statistics. In case the prior distributions are bounded, a logit transformation is applied to the parameters before the adjustment.

For simplicity reasons, we assume g to be linear as in [53] and adjust the accepted parameters as

$$\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i^* - (\mathbf{s}_i - \mathbf{s}_{\text{obs}})^\top \hat{\boldsymbol{\beta}}, \quad i = 1, \ldots, M_\epsilon, \tag{I.16}$$

where $\hat{\boldsymbol{\beta}}$ is the solution to the weighted least-squares problem

$$\underset{\boldsymbol{\alpha}, \boldsymbol{\beta}}{\arg\min} \sum_{i=1}^{M_\epsilon} \left[\boldsymbol{\theta}_i^* - \boldsymbol{\alpha} - (\mathbf{s}_i - \mathbf{s}_{\text{obs}})^\top \boldsymbol{\beta} \right]^2 \mathcal{W}_{\left(\widehat{\text{MMD}}^2_{k_{\text{SE}}}[\mathbf{X}_i, \mathbf{Z}]\right)}. \tag{I.17}$$

The weighting function $\mathcal{W}$ applies weights to each $\boldsymbol{\theta}_i$ based on the estimated MMD value. This guarantees that parameters which yield simulated log moments "closer to" $\mathbf{Z}$ are weighted more heavily. We take $\mathcal{W}$ to be the Epanechnikov function, $\mathcal{W}_{(\delta)} = 1 - (\delta/\delta_{\max})^2$ for $|\delta| \leq \delta_{\max}$ and zero otherwise, as proposed in [53]. Here, $\delta_{\max}$ is the maximum estimated MMD associated to the accepted parameters. Note that choosing a constant regression function, i.e. $\boldsymbol{\beta} = 0$, and assigning equal weights to all $\boldsymbol{\theta}_i$'s results in the basic rejection ABC algorithm. The regression adjustment therefore gives the adjusted parameter values $\tilde{\Theta} = (\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon})$.

Algorithm 11 ABC with MMD and Regression Adjustment

Input: Parameter values Θ, corresponding simulated data $\mathcal{X}$, observed $\mathbf{Z}$ & number of accepted samples M_ϵ.

Compute $\widehat{\mathrm{MMD}}^2_{k_{\mathrm{SE}}}(\mathbf{X}_i, \mathbf{Z})$ for all data-sets $\mathbf{X}_i \in \mathcal{X}$ using I.9.
Accept the M_ϵ parameters with the smallest MMD distance and denote these $\Theta^* = (\boldsymbol{\theta}^*_1, \ldots, \boldsymbol{\theta}^*_{M_\epsilon})$.
Compute $\mathcal{S}^*$ and $\mathbf{s}_{\mathrm{obs}} = S(Z)$, then solve the optimisation problem in (I.17) with Θ^*, $\mathcal{S}^*$, and $\mathbf{s}_{\mathrm{obs}}$ to get $\hat{\boldsymbol{\beta}}$.
Adjust Θ^* using (I.16) to obtain $\tilde{\Theta}$.

Output: Adjusted samples $\tilde{\Theta}$ from the Rejection-ABC posterior.

4.3 Importance Sampling using PMC

As a means to explore the posterior distribution over the parameter space efficiently, we employ a sequential Monte Carlo technique called PMC [20, 22, 52]. In PMC, the current parameter values $\tilde{\Theta}$ are used to generate a new set of parameters for the next iteration of the algorithm through importance sampling. This is a two-step procedure: (1) sample from the current parameters based on their importance weights, and (2) perturb the sampled parameter values using a proposal density.

The set of parameters in the initial iteration, $\tilde{\Theta}^{(1)} = (\tilde{\boldsymbol{\theta}}^{(1)}_1, \ldots, \tilde{\boldsymbol{\theta}}^{(1)}_{M_\epsilon})$, are assigned equal weights. The next set of parameters is obtained by drawing M values from $\tilde{\Theta}^{(1)}$ and perturbing these according to a probability distribution, called proposal. For simplicity, we perturb independently in each dimension using a Gaussian distribution, and reject values outside the prior range. Thus, the proposal reads

$$\varphi(\boldsymbol{\theta}; \tilde{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) = \mathbb{1}(\boldsymbol{\theta} \in \mathcal{R}) e^{-\frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})} \tag{I.18}$$

where $\mathbb{1}$ is an indicator function, $\mathcal{R} \subset \mathbb{R}^p$ is the prior range, and $\boldsymbol{\Sigma}$ is a diagonal matrix with variances $\sigma_j^2 > 0$ corresponding to parameter θ_j along the diagonal. We set the diagonal elements of $\boldsymbol{\Sigma}$ to twice the empirical variance of the adjusted parameter samples. This is denoted as $\boldsymbol{\Sigma} = 2\widehat{\mathrm{Var}}(\tilde{\Theta})$.

The set of M parameter values at iteration t, $\Theta^{(t)}$, is then used to simulate $\mathcal{X}^{(t)}$ from the model for MMD computation and regression adjustment (i.e. Alg. 11). In subsequent iterations, weights are assigned as

$$w_j^{(t)} \propto p\left(\boldsymbol{\theta}_j^{(t)}\right) / \sum_{i=1}^{M_\epsilon} w_i^{(t-1)} \varphi\left(\boldsymbol{\theta}_j^{(t)}; \tilde{\boldsymbol{\theta}}_i^{(t-1)}, \boldsymbol{\Sigma}^{(t-1)}\right), \tag{I.19}$$

$j = 1, \ldots, M_\epsilon$. The adjusted parameter values after iteration T are taken as samples from the approximate posterior distribution. Point estimates of $\boldsymbol{\theta}$,

Algorithm 12 PMC-ABC with MMD

Input: Prior $p(\boldsymbol{\theta})$, model $\mathbb{P}_{\boldsymbol{\theta}}$, observed data $\mathbf{Z}$, M_ϵ, M and T.

Initialize $t = 1$, draw $\Theta^{(1)} \overset{\text{iid}}{\sim} p(\boldsymbol{\theta})$ and simulate $\mathcal{X}^{(1)}$ using the parameters in $\Theta^{(1)}$.
Apply **Algorithm** 11 on $\{\mathcal{X}^{(1)}, \Theta^{(1)}\}$ to obtain $\tilde{\Theta}^{(1)}$.
Set $w_j^{(1)} = 1$ for $j = 1, \ldots, M_\epsilon$, and set $\boldsymbol{\Sigma}^{(1)} = 2\widehat{\text{Var}}\big(\tilde{\Theta}^{(1)}\big)$.
for $t = 2, \ldots, T$ **do**
  Compute $q_j = w_j^{(t-1)} / \sum_{i=1}^{M_\epsilon} w_i^{(t-1)}$ for $j = 1, \ldots, M_\epsilon$.
  for $i = 1, \ldots, M$ **do**
    Sample $\boldsymbol{\theta}_i^*$ from $\tilde{\Theta}^{(t-1)}$ s.t. $\tilde{\boldsymbol{\theta}}_j^{(t-1)}$ is selected with prob q_j.
    Generate $\boldsymbol{\theta}_i^{(t)} \sim \varphi\big(\cdot; \boldsymbol{\theta}_i^*, \boldsymbol{\Sigma}^{(t-1)}\big)$.
    Simulate $\mathbf{X}_i^{(t)}$ from the model with parameter $\boldsymbol{\theta}_i^{(t)}$.
  end for
  Apply **Algorithm** 11 on $\{\mathcal{X}^{(t)}, \Theta^{(t)}\}$ to obtain $\tilde{\Theta}^{(t)}$.
  Set $w_j^{(t)}$ using (I.19) for $j = 1, \ldots, M_\epsilon$.
  Set $\boldsymbol{\Sigma}^{(t)} = 2\widehat{\text{Var}}\big(\tilde{\Theta}^{(t)}\big)$.
end for

Output: Samples $\big(\tilde{\boldsymbol{\theta}}_1^{(T)}, \ldots, \tilde{\boldsymbol{\theta}}_{M_\epsilon}^{(T)}\big)$ from the PMC-ABC posterior.

such as the approximate posterior mean,

$$\hat{\boldsymbol{\theta}}^{(T)} = \frac{1}{M_\epsilon} \sum_{i=1}^{M_\epsilon} \tilde{\boldsymbol{\theta}}_i^{(T)}, \tag{I.20}$$

are straightforward to compute from the samples.

4.4 Handling Model Misspecification

We have now completed the description of Alg. 12. However, the framework of ABC relies on the implicit assumption that there exist parameter values in the prior support that yield simulated data "close" to the measured data. This assumption may not always hold if the model parameters cannot be set in any way to reproduce the data well. In this case, we say that the model is misspecified for the data. Misspecification can be detected and accounted for in the algorithm as explained in this subsection.

Consider a univariate parameter θ in the range $[\theta_{\min}, \theta_{\max}]$ resulting in a univariate statistic s in $[s_{\min}, s_{\max}]$ simulated from the model. If the observed statistic $s_{\text{obs}} \notin [s_{\min}, s_{\max}]$, then the model is likely to be misspecified. This is a challenge since under model misspecification, the local-linear regression adjustment has been shown to concentrate posterior mass on a completely different value than the rejection ABC [54]. In fact for parameters with bounded support, the regression adjustment moves the parameter samples outside the

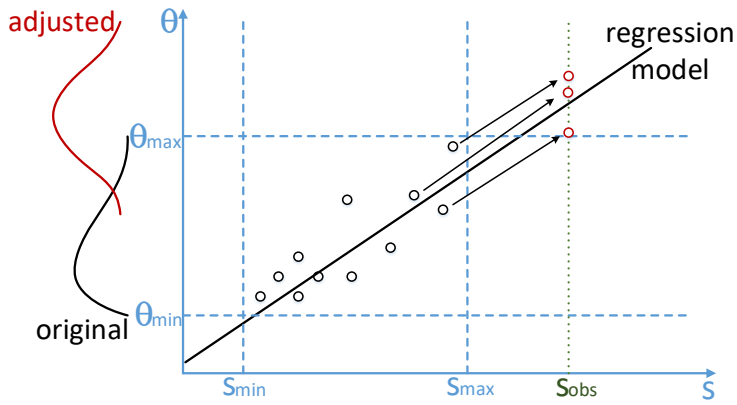


Fig. I.5: Local linear regression adjustment of parameter θ inspired from [55]. First, the regression model is fitted based on accepted parameter and statistic values. Then, the parameters are adjusted based on the fitted model, which can move them outside the prior range if $(s_i - s_{\text{obs}})$ is large.

prior range as illustrated in Fig. I.5. Hence, if s_{obs} lies outside the range of statistics that the model can simulate, then there is no guarantee that the adjusted samples of θ will lie inside the prior range.

We check for model misspecification by observing whether each element of $\mathbf{s}_{\text{obs}}$ lies within the range of corresponding statistics simulated from the model using $\Theta^{(1)}$. If any element of $\mathbf{s}_{\text{obs}}$ lies outside the range of values simulated from the model, then the model is deemed misspecified. In such a case, we replace $\mathbf{s}_{\text{obs}}$ by an alternative term, $\breve{\mathbf{s}}_{\text{obs}}$, computed from the model instead of the data using the parameter

$$\breve{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \quad f(\boldsymbol{\theta}; \Theta^*), \tag{I.21}$$

where $f(\boldsymbol{\theta}; \Theta^*)$ is the kernel density estimate computed from the samples Θ^*, and $\breve{\boldsymbol{\theta}}$ is the parameter corresponding to the mode of $f(\boldsymbol{\theta}; \Theta^*)$. Another choice for $\breve{\boldsymbol{\theta}}$ could be the posterior mean of rejection ABC [54]. However, we found the mean estimate to be unstable, especially in the initial iterations of the algorithm. Hence, in case of model misspecification, we set $\mathbf{s}_{\text{obs}} = \breve{\mathbf{s}}_{\text{obs}}$ in each iteration of the PMC-ABC algorithm, thus ensuring that the adjustment does not lead to parameter samples outside the prior range.

5 Simulation Experiments

We test the performance of the proposed calibration method on two different channel models, namely the Saleh-Valenzuela (S-V) and the propagation graph (PG) model. We chose models which differ significantly in their mathematical

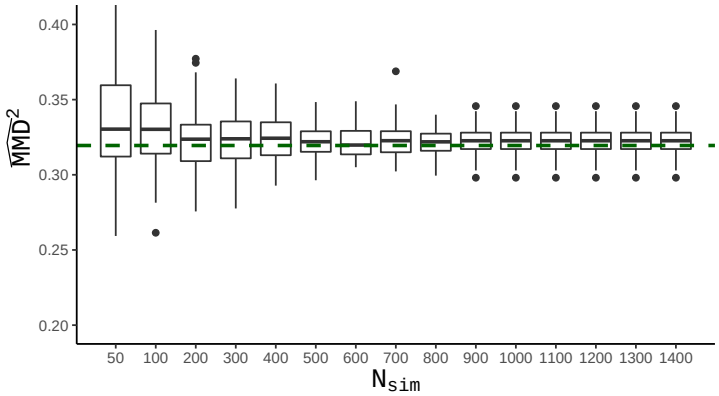


Fig. I.6: Boxplots of the estimated MMD^2 between $\mathbf{X}_{\text{true}}$ ($N_{\text{obs}} = 1000$) and $\mathbf{X}'$ as a function of N_{sim} computed by repeating the experiment 100 times for each value of N_{sim}. $\mathbf{X}'$ is generated from $\boldsymbol{\theta}' = [2 \times 10^{-8}, 6 \times 10^{7}, 10^{8}, 2 \times 10^{-8}, 10^{-9}, 5 \times 10^{-10}]^\top$ and $\mathbf{X}_{\text{true}}$ from $\boldsymbol{\theta}_{\text{true}} = [5 \times 10^{-8}, 2 \times 10^{7}, 10^{9}, 10^{-8}, 2 \times 10^{-9}, 10^{-9}]^\top$. The dashed green line is a corresponds to the value of the MMD^2 being approximated. Since this value is not available in closed-form, it is approximated by using $N_{\text{sim}} = 10^4$.

structure to highlight the generality of our approach. We first study in depth the advantages and drawbacks of our algorithm on simulated data. Then, in Section 6, we calibrate these models to data from an indoor measurement campaign [56].

For ease of comparison, we use the same measurement settings as in [56] for both simulations and measurements, i.e. $B = 4$ GHz, $N_s = 801$, and $t_{\max} = 200$ ns. We map the channel measurements to the first $I = 4$ temporal moments. In each iteration of the ABC algorithm, $M = 2000$ parameter samples are generated, out of which $M_\epsilon = 100$ are accepted to estimate the posterior distribution.

5.1 Application to the Saleh-Valenzuela model

The seminal S-V model [2] is widely used as it is easy to simulate from, but is notoriously difficult to calibrate due to its structure. Even though the model can be analyzed using the theory of spatial point processes [57, 58] and moments derived [59], its likelihood function is unavailable. Recent discussions of the physical interpretation of the S-V model, also outlining some difficulties with the model calibration, is given in [60–62]. These difficulties have inspired the use of many different heuristic calibration methods, as outlined in the introduction.

In the S-V model, the multipath components are assumed to arrive in clusters. The arrival time of the clusters and that of the rays within the clusters are modeled as one-dimensional homogeneous Poisson point processes with arrival rates Λ and λ, respectively. The gains of the multipath components are

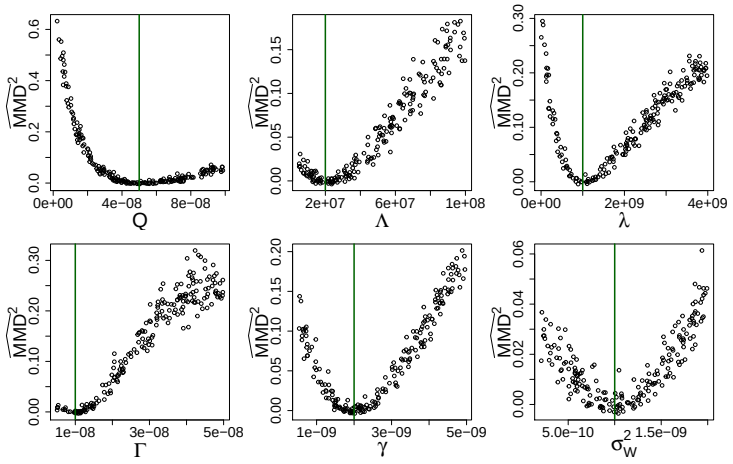


Fig. I.7: Estimated MMD^2 values plotted against parameters of the S-V model. The parameters are uniformly sampled 200 times from the prior range one at a time, keeping the others fixed to the true values denoted by the dark green lines. See Tab. I.1 for the prior ranges.

modeled as iid zero-mean complex Gaussian random variables with conditional variance that depends on three parameters; the average power of the first arriving multipath component Q, and the cluster and ray power decay constants Γ, γ, respectively. We refer the readers to [2] and [57] for a detailed description of the model. Including the noise variance, the parameter vector becomes $\boldsymbol{\theta} = [Q, \Lambda, \lambda, \Gamma, \gamma, \sigma_W^2]^\top$.

We begin by finding a reasonable value of N_{sim}. To that end, we generate pseudo-observed log moments, $\mathbf{X}_{\mathrm{true}}$, with $N_{\mathrm{obs}} = 1000$ realizations from the model by setting $\boldsymbol{\theta}$ to a "true" value. Using another value of the parameter vector, say $\boldsymbol{\theta}'$, we simulate $\mathbf{X}'$ from the model with varying N_{sim} and compute the estimated MMD between $\mathbf{X}'$ and $\mathbf{X}_{\mathrm{true}}$. This process is repeated 100 times to create error bars as shown in Fig. I.6. Although the MMD estimate gets more accurate as N_{sim} increases, the improvement however is small. Choosing a higher N_{sim} improves the MMD estimate, but increases the run-time of the of the algorithm significantly (since the computational cost is quadratic in N_{sim}, and simulating from the model can also be slow). Therefore, we set $N_{\mathrm{sim}} = 100$ as a reasonable compromise considering the trade-off between accuracy and computational cost.

We first verify that the MMD computed from the temporal moments reacts to changes in the S-V model parameters. To that end, we generate simulated data-sets by varying one parameter uniformly in the prior support while keeping the others fixed to their true value. As can be seen from Fig. I.7, the estimated MMD values increase as each of the parameters move away from their true value, and the minimum is (approximately) achieved when both the data-sets

Table I.1: Parameter estimates obtained for measured data. The standard deviation of the approximate posterior samples is given in parenthesis.

	$\boldsymbol{\theta}$	**Prior range**	**Estimate (std. deviation)**
S-V model	Q	$[10^{-9}, 10^{-7}]$	4.7×10^{-8} (4.6×10^{-9})
	Λ	$[5 \times 10^{6}, 10^{8}]$	8.6×10^{7} (9.8×10^{6})
	λ	$[5 \times 10^{-9}, 3 \times 10^{9}]$	1.5×10^{8} (4.2×10^{7})
	Γ	$[5 \times 10^{-9}, 5 \times 10^{-8}]$	8.2×10^{-9} (2.7×10^{-10})
	γ	$[5 \times 10^{-10}, 5 \times 10^{-9}]$	4.4×10^{-9} (4.7×10^{-10})
	σ_W^2	$[2 \times 10^{-10}, 2 \times 10^{-9}]$	3.5×10^{-10} (2.5×10^{-11})
PG model	g	[0,1]	0.50 (0.019)
	N_{scat}	[5,35]	18 (1.73)
	P_{vis}	[0,1]	0.99 (7.9×10^{-4})
	σ_W^2	$[2 \times 10^{-10}, 2 \times 10^{-9}]$	4.4×10^{-10} (4.3×10^{-12})

are generated from approximately the same parameters. The MMD reacts to changes in all the parameters, albeit more for some than others, as can be seen from the different scales of the y-axis. We therefore conclude that the distribution of the first four log temporal moments is informative about the S-V model parameters.

We now use the proposed method to calibrate the S-V model using $\mathbf{X}_{\text{true}}$. We assume uninformative (flat) priors in the range given in Tab. I.1 for all the parameters to ensure that their marginal posteriors are unaffected by any prior beliefs. The prior ranges were set according to the measurement settings as done in [20]. The plots indicating convergence of the algorithm and the marginal posterior distributions for $T = 10$ iterations are shown in Fig. I.8. The approximate posterior samples concentrate around the true value for all the parameters. The algorithm converges rather quickly and the posteriors taper as the iterations proceed. In principle, the iterations could be stopped after four or five iterations, but we let it run till $T = 10$ for clarity. The algorithm gives a reasonable estimate for the parameters even in the first iteration. The proposed method is able to estimate Λ accurately as well, unlike in [20] where some post-processing was required to estimate Λ.

5.2 Application to the Propagation Graph model

As our second example, we demonstrate the performance of our proposed method on the PG model. The PG model was first introduced in [63], and since then has been applied to a wide range of scenarios in [64–67]. Recently, it has been extended to account for polarization in [15, 68, 69]. Although the model is easy to simulate from, its likelihood function is unknown. A method of moments based estimator was applied to calibrate the model in [15], but the moments equations were based on approximation and it required manually fixing one of the parameters.

The PG model [63] represents the radio channel as a directed graph with

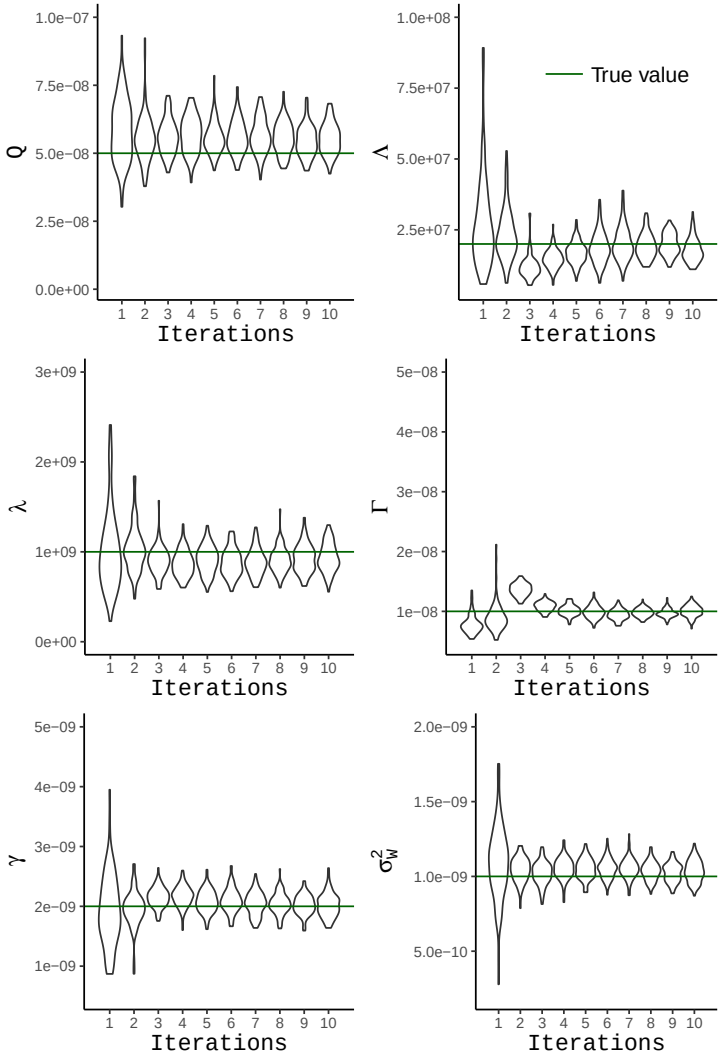


Fig. I.8: Violin plots of ABC posterior samples of S-V model parameters as a function of PMC iterations. Note that a violin plot is similar to a box plot with the addition of a rotated kernel density plot on each side. The dark green lines denote the true parameter values $\boldsymbol{\theta}_{\text{true}} = [5 \times 10^{-8}, 2 \times 10^{7}, 10^{9}, 10^{-8}, 2 \times 10^{-9}, 10^{-9}]^{\top}$.

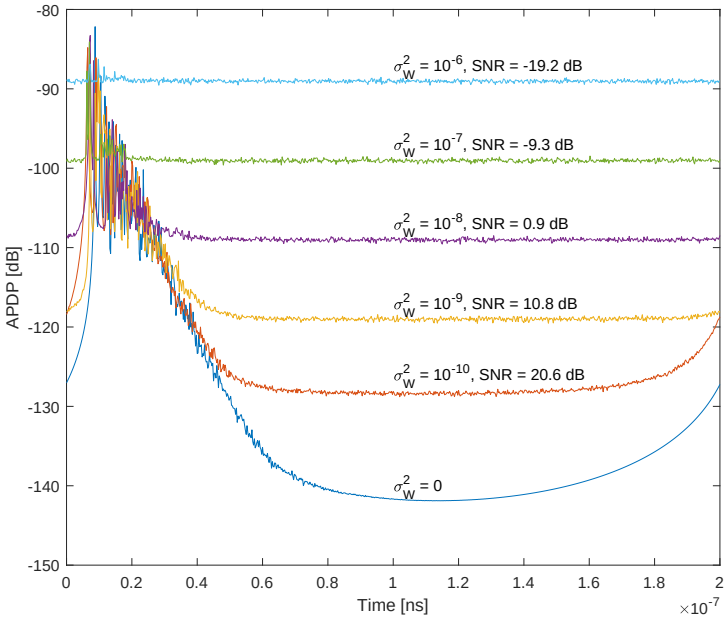


Fig. I.9: Averaged power delay profiles simulated from the PG model for different SNR levels.

the transmitters, receivers and scatterers as vertices. Edges model the wave propagation between the vertices. Edges are defined randomly depending on the probability of visibility, P_{vis}. Other parameters of the model include the number of scatterers, N_{scat}, and the reflection gain, g, resulting in the parameter vector $\boldsymbol{\theta} = [g, N_{\text{scat}}, P_{\text{vis}}, \sigma_W^2]^\top$. Note that N_{scat} is assumed to be real-valued during the regression adjustment, following which, its adjusted samples are rounded off to the nearest integer. We used the antenna positions and room geometry for the model according to the measurement conditions given in [56]. Hence, $N_{\text{obs}} = N_{\text{sim}} = 625$ for the PG model. For each call of the model, the scatterer positions are drawn uniformly across the room, and all 625 realizations are generated based on those positions.

We again use uniform priors for the parameters (see Tab. I.1) and apply $T = 10$ iterations of the proposed method to calibrate the PG model to the pseudo-observed data-set generated from $\boldsymbol{\theta}_{\text{true}}$. To prevent biased results due to a particular configuration of the scatterers, we generate the pseudo-observed data by combining data from four different calls of the model using $\boldsymbol{\theta}_{\text{true}}$. From Fig. I.11, we observe that the algorithm converges very quickly, and gives posteriors which are highly concentrated around the true value for P_{vis}, N_{scat}, and σ_W^2. The approximate posterior for g starts off very wide and then gets narrower as the iterations proceed. The method is therefore able to accurately calibrate the PG model.

To asses how the performance of the proposed algorithm is affected by the presence of noise, we now repeat this simulation experiment for different noise levels. We fix $g = 0.6$, $P_{\text{vis}} = 0.5$, $N_{\text{scat}} = 15$ and vary σ_W^2 from 10^{-10} to 10^{-6}. The signal-to-noise ratio (SNR), is defined as

$$\text{SNR} = 10 \log_{10}(\bar{m}_0 B/\sigma_W^2) \quad [\text{dB}], \tag{I.22}$$

where $\bar{m}_0$ is the sample mean of of the zeroth temporal moment computed by setting $\sigma_W^2 = 0$ in the PG model. The resulting averaged power delay profile (APDP) is shown in Fig. I.9. We run $T = 10$ iterations of the algorithm for each of the SNR values. The prior for σ_W^2 is adjusted according to the true value in each run of the algorithm. The violin plots of the approximate posterior after the tenth iteration in each case is shown in Fig. I.10.

We observe that the noise variance σ_W^2 is estimated extremely accurately at each SNR level. The estimation accuracy for P_{vis} and N_{scat} seems to suffer only at the lowest SNR level. Reducing the SNR impacts the estimation accuracy of g the most, with its approximate posterior converging to the prior as SNR decreases. This is expected as the higher the noise variance, the less visible the slope of the power delay profile which is determined by g. In conclusion, the algorithm performs well at SNR values encountered in measurements.

6 Application to Measured Data

We now attempt to fit both the S-V and the PG models to millimetre-wave radio channel measurements obtained from [56]. The measurements of the channel transfer function were performed in the bandwidth 58 GHz to 62 GHz with a VNA, using $N_s = 801$ equally spaced frequency points. The bandwidth of $B = 4$ GHz means the frequency separation was $\Delta f = 5$ MHz and $t_{\max} = 200$ ns. We use measurements taken in a small conference room of dimension $3{\times}4{\times}3$ m^3 in a non-line-of-sight scenario. At both transmitter and receiver sides, 5×5 antenna arrays were used. Although the antenna elements used in the measurement were dual polarized, we focus on the vertical-vertical polarization since both the models are uni-polarized. This gives $N_{\text{obs}} = 5 \times 5 \times 5 \times 5 = 625$. We keep the settings $M = 2000$ and $M_\epsilon = 100$ of the algorithm same as in the simulation experiments.

6.1 Calibrating the Saleh-Valenzuela model

Upon applying Alg. 12 to the measured data, regression adjustment yielded parameter samples outside the prior range. This indicated that the model is misspecified. That is indeed evident from Fig. I.12 where we plot elements of the vector $\mathbf{s}$, namely the mean and variance of $\mathbf{z}_0$ and $\mathbf{z}_1$, obtained from the measurements and the S-V model. The simulated summaries correspond to 2000 parameter values drawn from the prior. We observe that varying the

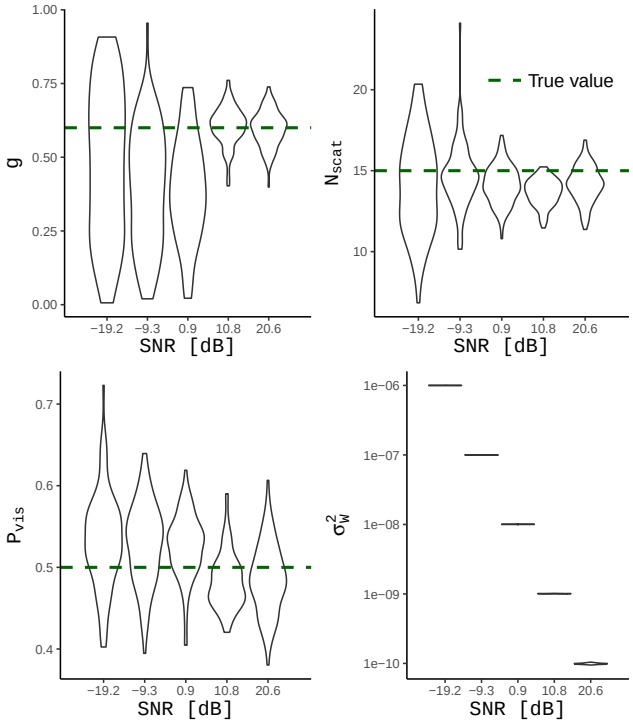


Fig. I.10: Violin plots of ABC posterior samples of PG model parameters after $T = 10$ iterations for different SNR levels. The APDP corresponding to each SNR is shown in Fig. I.9.

parameters of the S-V model in the prior range generated mean values that overlap the mean value from the measurements. However, the variance values from the S-V model does not capture the value observed in the measurements. That is, there exists no such $\boldsymbol{\theta}$ in the prior range that leads to $\mathbf{s}$ "close" to $\mathbf{s}_{\text{obs}}$ in terms of the variance of the temporal moments. Hence, the model is misspecified for this data and so we obtain $\mathbf{s}_{\text{obs}}$ from the model as per Sec. 4.4.

The posteriors obtained from the measured data are shown in Fig. I.13 for $T = 15$ iterations. The marginal approximate posteriors for λ, Γ, and σ_W^2 are highly concentrated. Posteriors for Γ and σ_W^2 appear to converge from the second iteration itself, indicating that these parameters affect the MMD the most. The posterior for λ becomes narrow and converges after the first few iterations. The posteriors for Q, Λ and γ take around eight or nine iterations to converge to a different location in the prior range than where they began from, unlike the simulation experiment. This is potentially due to the model being misspecified for the data, and so parameters that affect the distribution of the log temporal moments the most converge first. The approximate estimates after 15 iterations are reported in Tab. I.1. Considering that the regression

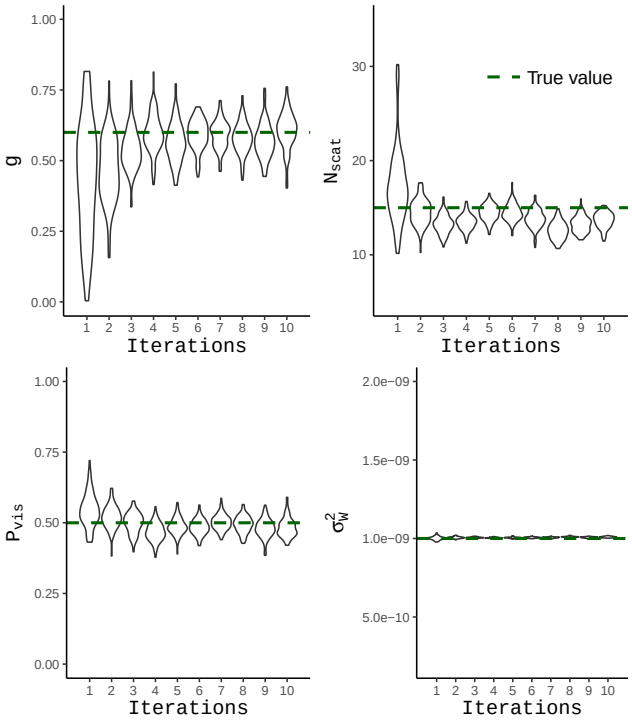


Fig. I.11: Violin plots of ABC posterior samples of PG model parameters as a function of PMC iterations. A violin plot is similar to a box plot with the addition of a rotated kernel density plot on each side. $\boldsymbol{\theta}_{\text{true}} = [0.6, 15, 0.5, 10^{-9}]^{\top}$ is denoted by the dark green dashed line.

adjustment in the first few iterations are done based on a coarse estimate of $\mathbf{s}_{\text{obs}}$ from the model, the algorithm seems to work very well. The estimate of Λ is high, indicating arrival of around 17 clusters on an average, while that of λ is quite low. The model is therefore forced to the case with many clusters having very few multipath components each, thus approaching the "unclustered" Turin model with constant rate.

The misspecification of the S-V model for the measured data is not surprising, as the measurement conditions are not replicated in the model. The virtual array measurements are from a single array position in the room, hence the same clusters are observed in each transmit-receive antenna pair. On the other hand, each realization out of the S-V model is an independent realization from the underlying point process. As a result, we hardly see any variance in the log temporal moments of the data, which is not achieved in the S-V model for any configuration of the parameters.

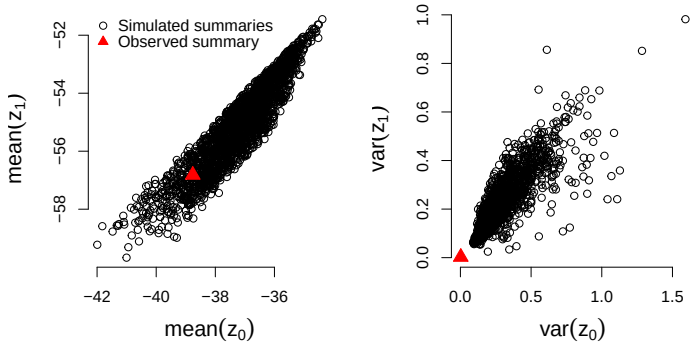


Fig. I.12: Mean (left) of $\mathbf{z}_0$ versus $\mathbf{z}_1$ simulated from the S-V model, along with the corresponding observed summary computed from the measured data (red). The mean and variance of z_0 is unitless, while that of z_1 is [log s] and [(log s)^2], respectively. The observed summary lies in the point cloud generated by the model. In contrast, the S-V model is not able to replicate the higher moments of the data, as seen from the variance plot (right), indicating model misspecification. Each of the 2000 simulated summaries correspond to one parameter drawn from the prior.

6.2 Calibrating the Propagation Graph model

The results obtained on calibration of the PG model on measured data after $T = 10$ iterations is shown in Fig. I.14. In this case, the model is not misspecified for the considered data. The approximate marginal posterior distributions for all the parameters start off wide and then seem to converge after around four or five iterations. The posteriors are also quite concentrated for all the four parameters, especially P_{vis} and σ_W^2. Overall, the results are similar to what is observed in the simulation experiment. See Tab. I.1 for approximate estimates of the parameters after $T = 10$ iterations. The estimates are very similar to the ones reported in [22] where the polarized PG model was calibrated on data from the same measurement campaign. The estimate of P_{vis} is almost one, indicating that nearly all scatterers are connected. The estimates of g and P_{vis} are consistent with the values reported from measurements [15] in other in-room scenarios for the PG model. Moreover, these values are close to those used in simulations with the PG model in [63, 70].

6.3 Model Validation

While the proposed method easily calibrates both the S-V and the PG models to measured data, there is no guarantee that the fitted models replicate the data well. This effect is of course not specific to the proposed method, but pertains to any calibration method. Thus, an extra step, termed model validation,

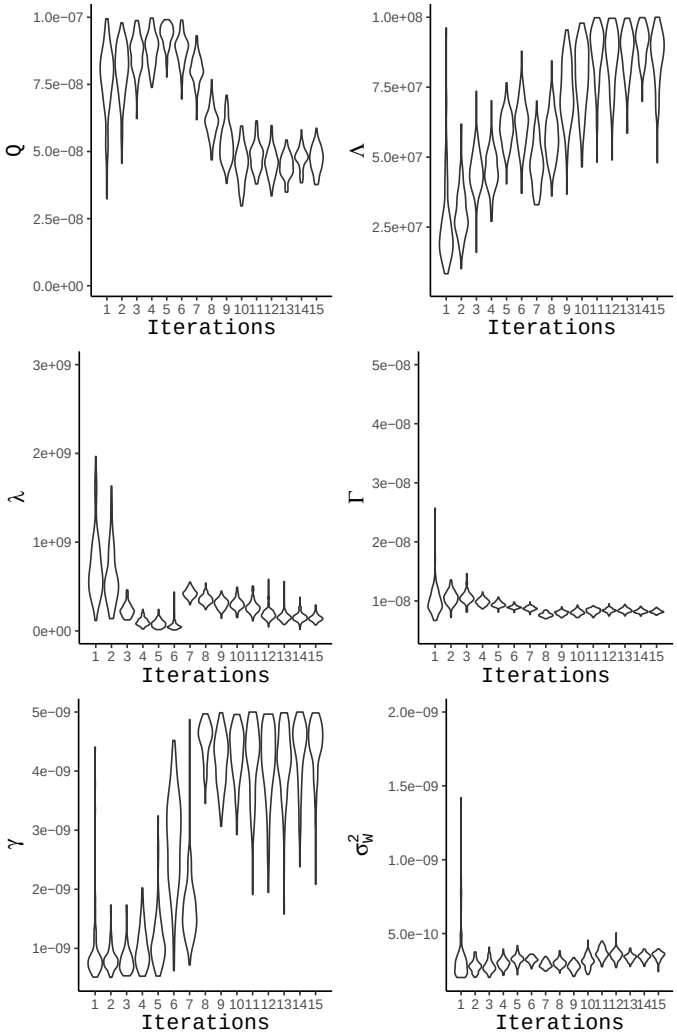


Fig. I.13: Violin plots of ABC posterior samples of S-V model parameters as a function of PMC iterations for measured data.

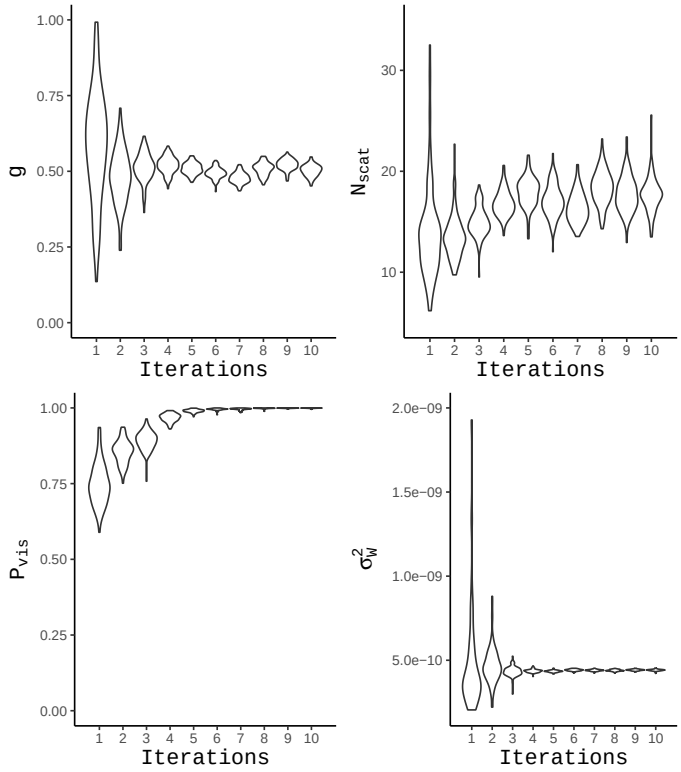


Fig. I.14: Violin plots of the ABC posterior samples of PG model parameters as a function of PMC iterations for measured data.

should be performed where predictions of the calibrated models are compared to the data, and possibly other data-sets not used in the calibration process. Performing a full model validation is out of scope of this paper, as our focus is on the calibration method itself. Instead, as a final step we check how well the two calibrated models fit the input data-set.

To this end, we simulate 625 channel realizations from both models with parameters set according to Tab. I.1. We compare the outputs from the models to the measured data in terms of the APDP and the empirical cumulative distribution function (cdf) of root mean square (rms) delay spread τ_{rms}, mean delay $\bar{\tau}$, and received power P_0 computed per channel realization, according to

$$P_0 = m_0, \quad \bar{\tau} = \frac{m_1}{m_0}, \quad \text{and} \quad \tau_{\mathrm{rms}} = \sqrt{\frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2}. \tag{I.23}$$

It appears from Fig. I.15 that both the models are able to fit the APDP of the measurements well. The slope is captured well by both the models, along with the noise floor, although the S-V model slightly underestimates it. The

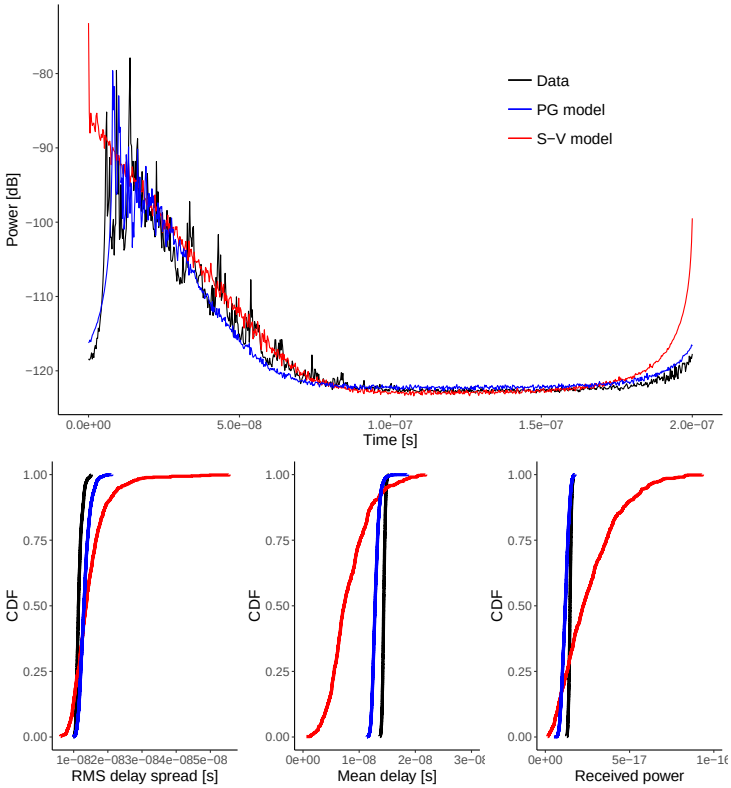


Fig. I.15: S-V and PG model fit to the measured data in terms of the APDP and empirical cdfs for rms delay spread, mean delay, and received power. Note that the first multipath component of the S-V model arrives at $t = 0$ as in [2].

S-V model, however, is not able to replicate the peaks in the APDP of the measurements, while the PG model represents the initial peaks better. This effect is to be expected for the particular settings of the S-V model with many clusters and very few within-cluster components. The peaks from the S-V model are averaged out since the channel realizations are independent. This is unlike the PG model where positions of the antennas in the virtual array are included, thus simulating correlated channel realizations.

Even though the APDPs are similar, the two models yields very different empirical cdfs of τ_{rms}, $\bar{\tau}$, and P_0 as reported in Fig. I.15. The PG model captures the behavior of the cdfs very well, while the S-V model clearly fails to do so, especially for the mean delay and the received power. The means of the rms delay spread from both the models are fairly close to the measured data, but the spread differs for the S-V model. As noticed theoretically in [61, 62], multipath models can yield temporal moments with similar means while differing vastly in variance. Indeed, for a stochastic multipath model,

the covariance structure of the temporal moments depends on both first- and second-order properties of the underlying point process [71].

The misspecification of the S-V model arises from disregarding the dependencies between the measurements obtained from different antennas in the array. This in turn leads to the discrepancy in the variance of the log temporal moments as observed in Fig I.12. Thus, to alleviate the misspecification, the array structure should be incorporated in the model, as is done inherently in the PG model. This could be a contributing reason why other authors [72] have found fully stochastic models inadequate and instead recommended using geometry-based and fully deterministic approaches for millimetre-wave data. Irrespective of the cause, such misspecifications can be detected by the proposed method, thereby assisting in the modeling process.

7 Discussion

The proposed method makes certain choices such as the number of temporal moments to use. We found that using the $I = 4$ temporal moments, (m_0, m_1, m_2, m_3), gives accurate estimates with narrow posteriors after the first couple of iterations itself, while slight degradation in the performance is observed with $I = 3$ moments m_0, m_1 and m_2. Although the method permits the use of arbitrarily many moments, we did not see significant improvements in performance when including more than four moments. Although the temporal moments seem adequate for calibration, the channel measurements could in principle be summarized into other statistics as long as they are informative about the model parameters.

Other choices for the method include the prior distribution and the settings of the ABC algorithm. We used uninformative priors to demonstrate the accuracy of the method based on data alone. However, including informative priors would speed up the convergence of the algorithm. For a reasonable approximation to the posterior distribution from samples, we suggest setting $M_\epsilon = 100$ or more. Depending on the computational budget, ϵ can be set around 5% or less. Our chosen settings seem to work well for both the models, and hence, they can be a good starting point for initial experiments. We do not provide a stopping criterion for the algorithm, but instead encourage monitoring the posterior distributions for convergence, as the number of iterations required may vary across different parameters and models. Potentially a stopping criterion could be implemented where the iterations are stopped if the MMSE estimate changes less than some tolerance over iterations.

To calibrate a new channel model using our method, we suggest the following sequence of steps. Start by setting up priors for the model parameters based on available knowledge. Taking $J = 4$ temporal moments as a starting point, perform the simulation study of computing the MMD^2 by varying one parameter at a time as done in Fig. I.7. This experiment is informative in qualifying the required number of temporal moments. If the MMD is clearly

impacted by varying the parameters, apply the method to calibrate the new model with the proposed settings of M and M_ϵ. If not, then adjust the number of temporal moments J and repeat the process. Finally, monitor the posterior distributions for convergence and terminate the algorithm accordingly.

As the MMD compares infinitely many summaries of the two data-sets, it works better than comparing only the low-order moments such as the means and covariances of the temporal moments as in [19, 20, 22]. When choosing a characteristic kernel, the MMD also guarantees that distributions are uniquely identified by these moments, unlike the case when comparing a finite number of moments. The MMD is a strong notion of distance in the sense that recovery of the true parameter value is guaranteed as the number of data points grows. The MMD also leads to robust estimators; i.e. estimators which will return reasonable estimates even in the presence of outliers in the data or mild model misspecification [26, 27]. The median heuristic is a reasonable choice for balancing robustness and efficiency as discussed in [26]. The choice of kernel is not as impactful as the choice of the lengthscale, and the proposed squared-exponential kernel seems to work well.

The proposed method is computationally lightweight and can be run on standard laptops with reasonable run-time. In the experiments, the algorithm ran on a Lenovo ThinkPad with Intel Core i7 processor having 24 GB RAM. This gave a run-time of 5.5 hours for the PG model and around 2 days for the S-V model for ten iterations of the algorithm. In our tests, the computation time is dominated by the particular model evaluation time, while computation of temporal moments and the MMD is negligible. Thus, the computational cost depends heavily on the specific model and its implementation. Furthermore, the run-time is impacted by specific settings of some parameters, e.g. Λ and λ in S-V model and N_{scat} in PG model. For higher "true" values of these parameters, the model, and in turn, the calibration algorithm, takes considerably longer time to run. An obvious way to reduce the run-time is to run the algorithm on hardware with more processing power or by making parallel calls to the model during each iteration.

The proposed method relies solely on the ability to simulate from the model being calibrated, and not on the tractability of the likelihood or moment functions. Moreover, the method does not depend on the particular mathematical construction of the model, which enables calibration of very different models using the exact same procedure. This presents the opportunity to compare and select the best fitting model for a given data-set. Additionally, the proposed method inherently estimates the uncertainty of the fitted parameters, which is lacking in the state-of-the-art calibration approaches. In contrast to the rather complex state-of-the-art calibration methods, the proposed method is simple to implement in R, MATLAB or Python and requires very few settings such as M and M_ϵ. This has the clear advantage that results obtained from the method are easy to reproduce. Moreover, the method can detect and calibrate misspecified models as well. This is usually ignored or treated heuristically in standard algorithms.

The proposed method can be used for a broad class of models where the likelihood is not known or difficult to compute. This is a great advantage in the model development as models can potentially be calibrated before their derivation is finalized. If the model is deemed worthy of further study, effort may be devoted to derive its likelihood function. The proposed method may also be used in cases where such a likelihood is in fact available, or available up to some intractable normalization constant. In such cases the ABC approach may, however, be less effective than methods based on the likelihood. In those cases, other distances could be used; see for example Stein discrepancies for cases where the likelihood is unnormalised [73]. Similarly, if factorization of the likelihood is possible and some factors can be evaluated, more efficient inference methods than ABC may be derived relying e.g. on message passing techniques. Such methods rely extensively on the particular models and the structure of their likelihoods. Thus, the gain in efficiency comes at a cost in the form of a loss in generality compared to the proposed ABC method. Finally, we remark that distance metrics such as the Wasserstein or the Hellinger distance could potentially be used instead of the MMD. However, future studies are required to assess their applicability for calibrating stochastic channel models.

8 Conclusion

The proposed ABC method based on MMD is able to accurately calibrate wideband radio models of very different mathematical structure. The proposed method relies on computing temporal moments of the received signal, and thereby circumvents the need for multipath extraction or clustering. As a result, the method is automatic as no pre- or post-processing of the data and estimates are required. We find that the method is able to fit models to both simulated and measured data. This work opens possibilities of developing similar methods for calibrating directional and time-dependent channel models. Potentially, maximum mean discrepancy could be used for other problems in propagation and communication studies that involve comparing data-sets.

Acknowledgment

The authors would like to thank Dr. Carl Gustafson and Prof. Fredrik Tufvesson (Lund University) for providing the measurement data.

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. 21, no. 1, pp. 1–9, Feb 1972.

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

[3] 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.

[4] L. Raschkowski, P. Kyösti, K. Kusume, and E. T. Jämsä, "METIS channel models, deliverable d1.4 v3," Tech. Rep. ICT-317669 METIS Project, 2015.

[5] P. Kyösti, "Winner II channel models, deliverables d1.1.2 v1.2, part I: Channel models," Tech. Rep. IST-4-027756 WINNER II Project, 2007.

[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, 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 *Eur. Conf. on Antennas and 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. on Wireless Commun.*, vol. 18, no. 8, pp. 3856–3870, 2019.

[9] M. Yang, B. Ai, R. He, G. Wang, L. Chen, X. Li, C. Huang, Z. Ma, Z. Zhong, J. Wang, Y. Li, and T. Juhana, "Measurements and cluster-based modeling of vehicle-to-vehicle channels with large vehicle obstructions," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 9, pp. 5860–5874, 2020.

[10] X. Yin and X. Cheng, *Propagation Channel Characterization, Parameter Estimation, and Modeling for Wireless Communications.* John Wiley & Sons Singapore Pte. Ltd, Feb 2018.

[11] N. Czink, P. Cera, J. Salo, E. Bonek, J. P. Nuutinen, and J. Ylitalo, "A framework for automatic clustering of parameteric MIMO channel data including path powers," in *Proc. IEEE 64th Veh. Technol. Conf.-Fall*, 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, 2013.

[13] 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.

[14] L. Greenstein, S. Ghassemzadeh, S.-C. Hong, and V. Tarokh, "Comparison study of UWB indoor channel models," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 1, pp. 128–135, Jan 2007.

[15] R. Adeogun, T. Pedersen, C. Gustafson, and F. Tufvesson, "Polarimetric Wireless Indoor Channel Modelling Based on Propagation Graph," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 10, pp. 6585–6595, 2019.

[16] C. Hirsch, A. Bharti, T. Pedersen, and R. Waagepetersen, "Maximum likelihood calibration of stochastic multipath radio channel models," *accepted IEEE Trans. on Antennas and Propag.*, 2020.

[17] A. Bharti, R. Adeogun, and T. Pedersen, "Parameter Estimation for Stochastic Channel Models using Temporal Moments," in *Proc. 2019 IEEE Int. Symp. on Antennas and Propag. and USNC-URSI Radio Sci. Meeting*, 2019.

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

[19] A. Bharti, R. Adeogun, and T. Pedersen, "Estimator for Stochastic Channel Model without Multipath Extraction using Temporal Moments," in *20th IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2019.

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

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

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

[23] ——, "Auto-generated summaries for stochastic radio channel models," in *15th Eur. Conf. on Antennas and Propag.*, 2021, pp. 1–5.

[24] S. A. Sisson, *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, Sep 2018.

[25] A. Gretton, K. Borgwardt, M. J. Rasch, and B. Scholkopf, "A kernel two-sample test," *J. of Mach. Learn. Res.*, vol. 13, pp. 723–773, 2012.

[26] F.-X. Briol, A. Barp, A. B. Duncan, and M. Girolami, "Statistical inference for generative models with maximum mean discrepancy," *arXiv:1906.05944*, 2019.

[27] B.-E. Chérief-Abdellatif and P. Alquier, "Finite sample properties of parametric MMD estimation: robustness to misspecification and dependence," *arXiv:1912.05737*, 2019.

[28] B.-E. Cherief-Abdellatif and P. Alquier, "Mmd-bayes: Robust bayesian estimation via maximum mean discrepancy," ser. Proc. of Mach. Learn. Res., C. Zhang, F. Ruiz, T. Bui, A. B. Dieng, and D. Liang, Eds., vol. 118. PMLR, 08 Dec 2020, pp. 1–21. [Online]. Available: http://proceedings.mlr.press/v118/cherief-abdellatif20a.html

[29] S. Nakagome, K. Fukumizu, and S. Mano, "Kernel approximate Bayesian computation in population genetic inferences," *Stat. Appl. in Genet. and Mol. Biol.*, vol. 12, no. 6, pp. 667–678, 2013.

[30] M. Park, W. Jitkrittum, and D. Sejdinovic, "K2-ABC: approximate Bayesian computation with kernel embeddings," *Proc. of the 19th Int. Conf. on Artif. Intell. and Statistics*, vol. 51, pp. 398–407, 2015. [Online]. Available: http://arxiv.org/abs/1502.02558

[31] J. Mitrovic, D. Sejdinovic, and Y. W. Teh, "DR-ABC: Approximate Bayesian computation with kernel-based distribution regression," *33rd Int. Conf. on Mach. Learn., ICML*, vol. 3, pp. 2209–2218, 2016.

[32] K. Kisamori, M. Kanagawa, and K. Yamazaki, "Simulator calibration under covariate shift with kernels," in *Proc.s of the 23rd Int. Conf. on Artif. Intell. and Statistics*, vol. 108. PMLR, Aug 2020, pp. 1244–1253. [Online]. Available: http://proceedings.mlr.press/v108/kisamori20a.html

[33] G. K. Dziugaite, D. M. Roy, and Z. Ghahramani, "Training generative neural networks via maximum mean discrepancy optimization," in *Proc. of 31st Conf. on Uncertain. in Artif. Intell.*, 2015, pp. 258–267.

[34] D. J. Sutherland, H.-Y. Tung, H. Strathmann, S. De, A. Ramdas, A. Smola, and A. Gretton, "Generative models and model criticism via optimized maximum mean discrepancy," in *Int. Conf. on Learn. Represent.*, 2017.

[35] Y. Li, K. Swersky, and R. Zemel, "Generative moment matching networks," in *Proc. of the 32nd Int. Conf. on Mach. Learn. - Vol. 37*, ser. ICML'15. JMLR.org, 2015, p. 1718–1727.

[36] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, "Kernel mean embedding of distributions: A review and beyond," *Found. and Trends® in Mach. Learn.*, vol. 10, no. 1-2, pp. 1–141, 2017.

[37] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. New York: Springer Science+Business Media, 2004.

[38] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. G. Lanckriet, "Hilbert space embeddings and metrics on probability measures," *J. of Mach. Learn. Res.*, vol. 11, 2010. [Online]. Available: http://arxiv.org/abs/0907.5309

[39] C.-J. Simon-Gabriel and B. Schölkopf, "Kernel Distribution Embeddings: Universal Kernels, Characteristic Kernels and Kernel Metrics on Distributions," *J. of Mach. Learn. Res.*, vol. 19, no. 44, pp. 1–29, 2018.

[40] F.-X. Briol, C. J. Oates, M. Girolami, and M. A. Osborne, "Frank-wolfe bayesian quadrature: Probabilistic integration with theoretical guarantees," in *Proc. of the 28th Int. Conf. on Neural Inf. Process. Syst. - Volume 1*, ser. NIPS'15, 2015, p. 1162–1170.

[41] S. Reddi, A. Ramdas, B. Poczos, A. Singh, and L. Wasserman, "On the High Dimensional Power of a Linear-Time Two Sample Test under Mean-shift Alternatives," in *Proc. of the 18th Int. Conf. on Artif. Intell. and Stat.*, vol. 38. PMLR, May 2015, pp. 772–780. [Online]. Available: http://proceedings.mlr.press/v38/reddi15.html

[42] S. Rüping, "SVM kernels for time series analysis," Tech. Rep., 2001.

[43] M. Cuturi, J.-P. Vert, Ø. Birkenes, and T. Matsui, "A kernel for time series based on global alignments," *IEEE Int. Conf. on Acoust„ Speech and Signal Process.*, vol. 2, pp. 413–416, 2007.

[44] M. Cuturi, "Fast global alignment kernels," *Proc. of the 28th Int. Conf. on Mach. Learn.*, pp. 929–936, 2011.

[45] I. Chevyrev and H. Oberhauser, "Signature moments to characterize laws of stochastic processes," *arXiv:1810.10971*, 2018.

[46] F. J. Kiŕaly and H. Oberhauser, "Kernels for sequentially ordered data," *J. of Mach. Learn. Res.*, vol. 20, pp. 1–45, 2019.

[47] G. Wynne and A. B. Duncan, "A kernel two-sample test for functional data," *arXiv:2008.11095*, 2020.

[48] L. E. Franks, *Signal Theory*. Englewood Cliffs, N. J., Prentice-Hall, 1969.

[49] A. Bharti, L. Clavier, and T. Pedersen, "Joint statistical modeling of received power, mean delay, and delay spread for indoor wideband radio channels," in *14th Eur. Conf. on Antennas and Propag.*, 2020, pp. 1–5.

[50] A. Bharti, R. Adeogun, X. Cai, W. Fan, F.-X. Briol, L. Clavier, and T. Pedersen, "Joint modeling of received power, mean delay, and delay spread for wideband radio channels," *IEEE Trans. on Antennas and Propag.*, pp. 1–1, 2021.

[51] I. Steinwart and A. Christmann, *Support Vector Machines*. Springer, 2008.

[52] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, "Adaptive approximate bayesian computation," *Biometrika*, vol. 96, no. 4, pp. 983–990, Oct 2009.

[53] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002. [Online]. Available: https://www.genetics.org/content/162/4/2025

[54] D. T. Frazier, C. P. Robert, and J. Rousseau, "Model misspecification in approximate bayesian computation: consequences and diagnostics," *J. of the R. Stat. Soc.: Ser. B (Stat. Methodol.)*, vol. 82, no. 2, pp. 421–444, Jan 2020.

[55] J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander, "Fundamentals and recent developments in approximate bayesian computation," *Syst. Biol.*, vol. 66, pp. 66–82, Jan 2017.

[56] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the polarimetric mm-wave propagation channel using censored measurements," in *2016 Global Commun. Conf.* IEEE, Dec 2016.

[57] M. L. Jakobsen, T. Pedersen, and B. H. Fleury, "Analysis of the stochastic channel model by saleh & valenzuela via the theory of point processes," *Int. Zurich Seminar on Commun.*, 2012.

[58] J. A. Gubner, B. N. Bhaskar, and K. Hao, "Multipath-cluster channel models," in *2012 IEEE Int. Conf. on Ultra-Wideband*, 2012, pp. 292–296.

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

[60] 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. on Antennas and Propag.*, vol. 62, no. 9, pp. 4780–4793, 2014.

[61] T. Pedersen, "Modeling of path arrival rate for in-room radio channels with directive antennas," *IEEE Trans. on Antennas and Propag.*, vol. 66, no. 9, pp. 4791–4805, 2018.

[62] ——, "Stochastic Multipath Model for the In-Room Radio Channel Based on Room Electromagnetics," *IEEE Trans. on Antennas and Propag.*, vol. 67, no. 4, pp. 2591–2603, April 2019.

[63] T. Pedersen and B. H. Fleury, "Radio channel modelling using stochastic propagation graphs," in *IEEE ICC*, June 2007, pp. 2733–2738.

[64] L. Tian, X. Yin, Q. Zuo, J. Zhou, Z. Zhong, and S. X. Lu, "Channel modeling based on random propagation graphs for high speed railway scenarios," in *IEEE PIMRC*, Sept 2012, pp. 1746–1750.

[65] L. Tian, V. Degli-Esposti, E. M. Vitucci, and X. Yin, "Semi-deterministic radio channel modeling based on graph theory and ray-tracing," *IEEE Trans. on Antennas and Propag.*, vol. 64, no. 6, pp. 2475–2486, June 2016.

[66] J. Chen, X. Yin, L. Tian, and M. Kim, "Millimeter-wave channel modeling based on a unified propagation graph theory," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 246–249, Feb 2017.

[67] R. O. Adeogun, A. Bharti, and T. Pedersen, "An iterative transfer matrix computation method for propagation graphs in multi-room environments," *IEEE Antennas and Wireless Propag. Lett.*, vol. 18, no. 4, pp. 616–620, April 2019.

[68] R. Adeogun and T. Pedersen, "Propagation graph based model for multi-polarized wireless channels," in *IEEE WCNC*, April 2018.

[69] ——, "Modelling polarimetric power delay spectrum for indoor wireless channels via propagation graph formalism," in *2nd URSI Atlantic Radio Sci. Meeting*, May 2018.

[70] T. Pedersen, G. Steinböck, and B. H. Fleury, "Modeling of reverberant radio channels using propagation graphs," vol. 60, no. 12, pp. 5978–5988, Dec 2012.

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

[72] S. Salous, V. D. Esposti, F. Fuschini, R. S. Thomae, R. Mueller, D. Dupleich, K. Haneda, J.-M. M. Garcia-Pardo, J. P. Garcia, D. P. Gaillot, S. Hur, and M. Nekovee, "Millimeter-wave propagation: Characterization and modeling toward fifth-generation systems. [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 58, no. 6, pp. 115–127, dec 2016.

[73] A. Barp, F.-X. Briol, A. Duncan, M. Girolami, and L. Mackey, "Minimum stein discrepancy estimators," in *Adv. in Neural Inform. Process. Syst.*, vol. 32, 2019, pp. 12 964–12 976.

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-932-9

AALBORG UNIVERSITY PRESS