ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

LOGNORMAL MIXTURE SHADOW FADING MODEL AND APPLICATIONS

M.Sc. THESIS

Metin VURAL

Department of Electronics and Communication Engineering

Telecommunication Engineering Programme

JUNE 2014

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

LOGNORMAL MIXTURE SHADOW FADING MODEL AND APPLICATIONS

M.Sc. THESIS

Metin VURAL
(504121323)

Department of Electronics and Communication Engineering

Telecommunication Engineering Programme

Thesis Advisor: Doç. Dr. Güneş KARABULUT KURT

JUNE 2014

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

LOGNORMAL KARIŞIMI GÖLGELEME MODELİ VE UYGULAMALARI

YÜKSEK LİSANS TEZİ

Metin VURAL
(504121323)

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Telekomünikasyon Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Güneş KARABULUT KURT

HAZİRAN 2014

Metin VURAL, a M.Sc. student of ITU Graduate School of Science Engineering and Technology 504121323 successfully defended the thesis entitled **"LOGNORMAL MIXTURE SHADOW FADING MODEL AND APPLICATIONS"**, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Doç. Dr. Güneş KARABULUT KURT** Istanbul Technical University	.....
Jury Members :	**Prof. Dr. Hakan Ali ÇIRPAN** Istanbul Technical University	.....
	Yrd. Doç. Dr. Serhat ERKÜÇÜK Kadir Has University	.....

Date of Submission : **04 May 2014**
Date of Defense : **03 June 2014**

To my lovely family,

FOREWORD

This thesis is dedicated to my lovely family, to whom I am completely thankful for their support during my all education life.

I would like to thank my advisor, Professor Güneş Karabulut Kurt, for her precious guidance and contributions throughout the years.

I am also thankful to Saliha Büyükçorak for her valuable contributions.

I would also like to thank to the Turkish Scientific and Technological Research Council (TÜBİTAK) for supporting me through BİDEB scholarship program. This thesis is supported by 2210-C Graduate Priority Subject Areas Scholarship Program.

June 2014 Metin VURAL

TABLE OF CONTENTS

Page

ABBREVIATIONS

ACF : Autocorrelation Function
App : Appendix
CCF : Crosscorrelation Function
DPM : Dirichlet Process Mixture
EM : Expectation Maximization Algorithm
GHz : Gigahertz
GSM : Global System for Mobile Communications
KL : Kullback-Leibler
K-S : Kolmogorov-Smirnov
LOS : Line of Sight
LTE : Long-Term Evolution
MGF : Moment Generating Function
MHz : Megahertz
MLE : Maximum Likelihood Estimate
MRD : Mean Relative Difference
NLOS : Non-Line of Sight
ITU : Istanbul Technical University
pdf : Probability Density Function
SNR : Signal to Noise Ratio
WMRD : Weighted Mean Relative Difference
3GPP : The 3rd Generation Partnership Project

LIST OF TABLES

Page

LIST OF FIGURES

Page

LOGNORMAL MIXTURE SHADOW FADING MODEL AND APPLICATIONS

SUMMARY

Modeling the variations in the local mean received power, the shadow fading is a relatively understudied effect in the literature. The inaccuracy of the universally accepted lognormal model is shown in many works. However, proposing other statistical distributions, such as Gamma and Weibull, that are not stemmed from the natural underlying physical process, can not provide sufficient insights. Conceding the physical process of multiple reflections generating the lognormal distribution, in this thesis a generalised mixture model is proposed that can address the inaccuracies observed with a single lognormal distribution.

In order to show that the lognormal mixture model can be used under any shadow fading condition, it is proven that an arbitrary probability density function can accurately be represented by a mixture of lognormal random variables.

One of the main issues associated with mixture models is the determination of the mixture components. Here, two solutions are investigated; a Dirichlet process mixture based estimation technique that can provide the optimum number of components, and an expectation maximization algorithm based technique for a more practical implementation.

The proposed lognormal mixture shadow fading model and existing different shadow fading models like lognormal, Gamma and Weibull are compared by using two real-life empirical datasets. The accuracy of the mixture model is verified with both confidence based and error vector norm based techniques.

Concluding the thesis, the analysis associated with outage and cellular coverage probabilities are provided. Gamma, Weibull, lognormal, and the proposed lognormal mixture distributions for shadow fading model are investigated to determine outage and cellular coverage probabilities both empirically and theoretically. With the help of both theoretical and empirical results, it is verified that better fitting shadow fading models yield more realistic estimates.

LOGNORMAL KARIŞIMI GÖLGELEME MODELİ VE UYGULAMALARI

ÖZET

Alınan gücün ortalama değerinin etrafındaki değişimlerin yani gölgeleme etkisinin modellenmesi literatürde görece olarak az araştırılmış bir konudur. Gölgeleme etkisinin evrensel kabul gören modeli, değişimlerin lognormal rastlantı değişkeni olarak modellendiği lognormal gölgeleme modelidir. Ancak bu modelin eksikliği birçok çalışmada gösterilmiştir. Örnek olarak, kentsel mikrohücre ve makrohücre ölçümleri lognormal rastlantı değişkeni oluşturmak için gerekli sönümlemeye sahip olmayabilirler. Ayrıca varolan lognormal gölgeleme modeline göre standart sapma değeri alıcı-verici mesafesi arttıkça artması gerekirken birçok çalışma gölgeleme etkisinin standart sapma değerinin mesafeden bağımsız olduğunu ifade etmiştir. Bu eksiklikler varolan lognormal gölgeleme etkisi modelini geçersiz kılmaz ama bu modelin her zaman mükemmel sonuç vermeyeceğini gösterir. Bu yüzden Gamma ve Weibull gibi değişik rastlantı değişkenine sahip çeşitli gölgeleme modelleri de önerilmiştir. Bu modellerin oluşturulmasında sistem performans metrikleri için kapalı denklem çözümleri sunmaları önemli rol oynasa da bu modellerin oluşumu fiziksel süreçlere bağlı değildir ve teorik olarak önerilen modellerin fiziksel süreçlere bağlı olması oldukça önemlidir.

Çoklu yansımaların lognormal dağılımı oluşturduğunun kabulü altında, bu tezde tekli lognormal dağılımın yetersizliğini çözmek amacıyla gölgeleme etkisi için genelleştirilmiş lognormal karışımı gölgeleme modeli önerilmiştir. Karışım modelleri makine öğrenmesinde sınıflandırma, olasılık yoğunluk fonksiyonu kestirimi ve imge bölütleme gibi birçok çalışma alanında kullanılmaktadır. Bu çalışmada önerilen lognormal karışımı gölgeleme modeli sayesinde gölgeleme etkisinin daha doğru modelleneceği iddia edilmiştir. Ayrıca gölgeleme etkisinin daha doğru modellemenin daha doğru (gerçeğe uygun) performans değerleri vereceği hizmet kesilme olasılığı ve hücresel kapsama alanı olasılığı analizleri kanıtlanmıştır.

Önerilen lognormal karışım modelinin her gölgeleme etkisi koşulunda kullanılabileceğini göstermek amacıyla, herhangi bir olasılık yoğunluk fonksiyonunun lognormal rastlantı değişkenleri karışımı şeklinde modellenebileceği gösterilmiştir.

Karışım modelleri ile ilgili önemli konulardan biri de karışım modelini oluşturan dağılımların parametrelerinin bulunmasıdır. Bu çalışmada parametrelerin kestirimi için Dirichlet karışım süreci (Dirichlet process mixture-DPM) ve beklenti maksimizasyonu (expectation maximization-EM) algoritmaları kullanılmıştır.

DPM algoritması, optimum bileşen sayısının bulunmasına dayanan bir tekniktir. DPM modelini bu çalışmada kullanabilmek için değişkenlerin aynı mesafe değerlerinde değişiklik göstermesi beklenmektedir. Bu çalışmada alınan güç değerleri aynı

mesafe değerlerinde farklılık göstermektedir ve birbirlerinden bağımsız oldukları ispatlanmıştır. Dolayısıyla DPM modelinin karışım bileşenlerinin parametrelerinin kestiriminde kullanılması uygundur. EM algoritması ise uygulanması DPM modeline göre daha kolay olan bir yöntemdir. EM algoritması beklenti adımı (E-adım) ve maksimize adımı (M-adım) olmak üzere iki adımdan oluşan iteratif bir yöntemdir. Her iterasyonda, öncelikle E-adım ile başlayarak her veri noktası için herhangi bir bileşenin üyesine ait olma olasılıkları bulunur, daha sonra M-adım ile her sınıfın olasılık dağılımı fonksiyonuna ait parametre değerleri üretilir. Kabul edilebilir hata oranına yaklaşınca veya maksimum iterasyon sayısına ulaşıncaya kadar iterasyon devam eder.

Önerilen lognormal karışımı gölgeleme modeli ile varolan lognormal, Gamma, Weibull gölgeleme modelleri iki ölçümsel veri seti kullanılarak karşılaştırılmıştır. Veri Seti I kentsel makrohücre ortamında oluşturulmuştur. GSM 900 bandında İstanbul Teknik Üniversite kampüsünde bulunan bir baz istasyonu kullanılarak alınan güç değerleri kaydedilmiştir. Veri setinin oluşturulmasında 940.51 Mhz frekansında 5 ayrı sabit ölçüm mesafesi kullanılarak toplamda 10000'in üzerinde alınan güç değeri kullanılmıştır. Ölçüm alınan baz istasyonunun boyu 6 metredir. Veri Seti II de kentsel makrohücre ortamında oluşturulmuş bir veri setidir. Almanya'da Ilmenau şehir merkezinde LTE bandında ölçüm yapılmıştır. 3 farklı baz istasyonundan 26250 görüş alanında olmayan ve 5391 görüş alanında olan alınan güç değerleri kullanılmıştır. Her iki veri seti için de karşılıklı ilinti fonksiyonları yardımıyla veri setindeki her alınan güç değerinin birbirinden bağımsız olduğu gösterilmiştir. Ayrıca mesafeye bağlı alınan güç saçılma grafiği de oluşturulmuştur. Bu çalışmada kullanılan iki veri setinin de gölgeleme etkisinin incelenmesi için uygun olduğu karşılıklı ilinti analizleriyle ve referanslarla belirtilmiştir.

Önerilen lognormal karışımı gölgeleme modeli histogramının ve diğer aday modellerin histogramının gerçek verinin histogramına uyumluğunu belirlemek için Kolmogorov-Smirnov (K-S) uyum kalitesi testi kullanılmıştır. Oluşturulan histogram ile eldeki verinin histogramını hata metrikleri ile karşılaştırabilmek için bu testin yapılması gerekmektedir. K-S uyum kalitesi testi sonuçlarına göre oluşturulan tüm lognormal karışımı modellerinin ve diğer aday modellerin (lognormal, Gamma, Weibull) histogramının gerçek histograma uyumlu olduğu gösterilmiştir. Önerilen lognormal karışımı gölgeleme modelinin gerçek veriye uygunluk derecesi ortalama bağıl fark (mean relative difference-MRD), ağırlıklandırılmış ortalama bağıl fark (weighted mean relative difference-WMRD) ve KL uyumsuzluk hata metrikleri kullanılarak var olan lognormal, Gamma ve Weibull gölgeleme modelleri ile karşılaştırılmıştır. Literatürde sıklıkla kullanılan bu hata metrikleri, kestirilen olasılık yoğunluk fonksiyonu ile eldeki verinin olasılık yoğunluk fonksiyonu arasındaki farkı ölçmek için kullanılır. Bahsedilen hata metrikleri yardımıyla önerilen lognormal karışımı gölgeleme modelinin eldeki veriye uygunluğunun var olan diğer lognormal, Gamma ve Weibull gölgeleme modellerinden daha fazla olduğu doğrulanmıştır.

Tez çalışmasının son bölümünde, hizmet kesilme olasılığı (p_{out}) ve hücresel kapsama alanı olasılığı ($C(\phi)$) analizleri incelenmiştir. p_{out} ve $C(\phi)$ performansları gölgeleme etkisinin dağılımının özelliğinden doğruca etkilenmektedir. Önerilen lognormal karışımı gölgeleme modeli için hizmet p_{out} ve ($C(\phi)$)'nın matematiksel ifadeleri türetilmiştir. Ayrıca lognormal, Gamma ve Weibull modelleri için p_{out} ve $C(\phi)$

matematiksel denklemleri ifade edilmiştir. Oluşturulan teorik ifadeler benzetim sonuçları ile karşılaştırılmıştır.

Veri Seti I ve Veri Seti II yardımıyla oluşturulan benzetim sonuçlarına göre önerilen lognormal karışımı gölgeleme modeli diğer modellere göre gerçeğe çok daha yakın p_{out} ve $C(\phi)$ değerleri vermektedir. Ayrıca teorik oluşturulan modeller benzetim sonuçları ile uyuşmaktadır. p_{out} ve $C(\phi)$ analizlerinin yapılmasının amacı gölgeleme etkisini daha doğru modellemenin daha gerçekçi performans sonuçları vereceğini göstermektir.

1. INTRODUCTION

In order to design and manage high data rate wireless communication networks, more accurate prediction of the received signal power is crucial. To determine this received signal power, we need to take three multiplicative factors about the wireless channel into account; the path loss, small scale (microscopic) fading and large scale (macroscopic or shadow) fading. The change in the average received power level related to the distance between the transmitter and the receiver is referred to as path loss. The widely accepted path loss model is the log-distance model, $10n\log(d)$, where n is the so called path loss exponent and d is the distance between the transmitter and the receiver [1, 2]. Range of n is typically shown to be around $2-12$ [1, 3]. The small scale fading effect results from the multi-path channel, and has been extensively studied in the literature with well established models such as Rayleigh, Ricean, Nakagami-q (Hoyt) and Nakagami-m channels [4]. As the third factor, the large scale fading has a fundamental role in wireless communication network design. Network's performance (such as outage probability and average data transmission rate) are planned according to the average received power levels, that are directly effected by the shadow fading characteristics. The lognormal like distributions are observed due to multiplication of large number of random attenuating factors. However, shadow fading, representing the variations in the local mean received power, is a surprisingly understudied phenomenon.

Shadow fading is first studied in [5], where Egli empirically demonstrated the compliance of the lognormal distribution over a small sector. Following this work, the lognormal distribution became the universally accepted shadow fading model in the literature and called lognormal shadowing. Lognormal shadowing has been extensively used in the literature, supported by empirical studies including for different environments such as outdoor and indoor mobile radio channels [6–12]. Although lognormal shadowing is a widely accepted shadow fading model, it has also some

deficiencies [6]. First of all, for urban microcell and macrocell measurements may have less than the necessary attenuation [13], [14] to generate Gaussian RV [15]. Secondly, according to existing shadow fading model, because of the number of random attenuations is expected to rise with increasing distance between the transmitter and the receiver, the standard deviation of shadow fading should increase. However, most studies assume that the standard deviation of shadow fading is distance independent [16], [17]. As mentioned in [6], such shortcomings does not make the existing model invalid, however show that the existing shadow fading model may not always present an impeccable solution. Because of the deficiencies mentioned above, other distributions recommended to model shadow fading like Gamma and Weibull [18], [19]. In [20], although not frequently utilized in channel modeling area, a mixture Gamma distribution is proposed to model the signal to noise ratio (SNR) of wireless channels, modeling the composite effects of both the small and large scale fading phenomenons. The main motivation in these works is the mathematical convenience of the Gamma and Weibull distributions in order to obtain closed form expressions for system performance metrics. However, since such distributions are not based on the actual underlying physical process, it is quite important to use theoretically justified models based on physical justifications [6, 21].

Even though the lognormal distribution is almost universally accepted to model shadow fading due to the solid physical basis, studies demonstrate that measurement histograms have skewed distribution curves [6–12]. Motivated by the clusters of [6] that may actually appear in real-life scattering conditions. With the help of this motivation, in this study different shadow fading models for lognormal, Gamma and Weibull are investigated. A lognormal mixture model as a more accurate and flexible shadow fading model is also presented. Mixture models are widely used in a variety of machine learning problems including classification, clustering, density estimation and image segmentation [22–26]. Presented mixture model makes use of the cluster concept of Salo and it is proposed a lognormal mixture model as a more accurate and flexible shadow fading model.

Modeling shadow fading more accurately should lead more realistic network performance results. Coverage area ($C(\phi)$) and outage probability (p_{out}) analysis

are important issues about network performance. Mathematical expressions for $C(\phi)$ and p_{out} are based on shadow fading characteristics. That is, determining the shadow fading characteristics properly leads to define more accurate $C(\phi)$ and p_{out}.

1.1 Outline and Contributions of The Thesis

In this thesis, first of all, a lognormal mixture model for shadow fading based on cluster concept is proposed. This model is a generalized form of lognormal shadowing, where we can consider the effects of distinct scattering clusters. Secondly, the flexibility and the applicability of proposed model is demonstrated by proving that an arbitrary probability density function (pdf) can be modeled as lognormal mixtures by using positive definite kernels. Then, two methodologies to determine the mixture model parameters are used. First approach is Dirichlet process mixture (DPM) based, where we determine the optimum number of mixture components according to maximum a posteriori (MAP) probability and the corresponding mixture parameters using classical Bayesian techniques. As the second methodology, a practical parameter estimation technique based on expectation maximization (EM) algorithm is considered which assumes that the number of mixture components are known a priori. The validity of the proposed model is verified with the empirical Dataset I is collected by me, in the Istanbul Technical University (ITU) campus from an operational network and the empirical Dataset II, is an urban macrocell channel dataset can be obtained by [27]. We quantify the accuracy of the mixture model by using both confidence based and error vector norm based techniques. Finally, as an application area, coverage area, ($C(\phi)$), and outage probability, (p_{out}), values corresponding to these mixture and considered shadow fading models is derived, and it is numerically validated that better fitting models can provide more accurate (realistic) estimates in terms of the coverage and outage probability.

This paper is organized as follows: In Chapter 2, wireless propagation model and lognormal shadow fading is explained. In Chapter 3, dataset properties are given along with the evaluation metrics to measure difference between obtained and the actual pdfs. In Chapter 4, several shadow fading distributions are investigated to determine $C(\phi)$ and p_{out} both with simulations and analytically. In Chapter 5, $C(\phi)$ and p_{out}

results for lognormal, lognormal mixture, Gamma and Weibull shadow fading models are demonstrated. Finally, in Chapter 6 conclusions of this thesis are presented.

2. WIRELESS PROPAGATION AND SHADOW FADING DISTRIBUTIONS

2.1 Wireless Propagation Model

In order to obtain more accurate radio propagation models, both analytical and empirical methods are used [1]. By considering empirical and theoretical approaches, the log-distance path loss model emerges to be the universally accepted propagation model. According to the log-distance model, $10n\log(d)$, received power decreases logarithmically with distance, where n is path loss exponent and d denotes the distance between the transmitter and the receiver. Additionally, to determine received power more accurately, we need to take large scale (macroscopic or shadow) fading into account in addition to the path loss. Considering shadow fading, the relation between the received power and the transmitted power of wireless communication system is defined in dB scale according to

$$P_{Rx}(d) = P_{Tx}(d) - 10n\log(d) + X, \tag{2.1}$$

where the received power $P_{Rx}(d)$ is calculated as the difference between the transmitted power $P_{Tx}(d)$ and the path loss. Here, X is the shadow fading term. In this thesis, the first goal is to investigate the distribution of X.

2.2 Lognormal Shadow Fading

In order to determine the shadow fading characteristics, distribution of X is a significant matter. Mostly, it is assumed that X is a normally distributed with zero mean and σ^2 variance at a particular location in dB [1, 3, 5, 7–9, 11, 12, 16]. Defining $X = \log(Y)$, the lognormal shadow fading implies $Y \sim f_Y(y) = \mathscr{LN}(0, \sigma^2)$. In lognormal shadow fading case, lognormal distribution describes the multiplication of large number of random attenuating factors. Despite the fact that the lognormal is the most studied and universally accepted distribution to model shadow fading, many studies show that measurement pdfs have close fit for lognormal pdf but do not have a perfect

fit [6–9, 11, 12, 16]. By using motivation of that the measurement histograms have skewed distribution curves, we target to determine the distribution of X considering also lognormal mixture, Gamma and Weibull distributions.

2.3 Lognormal Mixture Shadow Fading

In probability theory, lognormal distribution is a random variable whose logarithm is normally distributed. If we analyze P_{Rx} in dB, then analysis can be done with normal (Gaussian) random variables and mixture model. Gaussian mixture model is a parametric pdf that includes more than one Gaussian distribution. Let N denotes the number of Gaussian components in mixture, $\mathcal{N}_i(\mu_i, \sigma_i^2)$ represents the Gaussian pdf of components (i=*1,2,...,N*) with mean, μ_i, and variance, σ_i^2, w_i is the weight ($\sum_{i=1}^{N} w_i = 1$, for all $w_i \geq 0$), then pdf of Gaussian mixture model, p_M, can be written as

$$p_M = \sum_{i=1}^{N} w_i \mathcal{N}_i(\mu_i, \sigma_i^2) \tag{2.2}$$

In order to determine mixture components we resort to the well-known expectation maximization (EM) algorithm [28, 29].

2.3.1 Modeling an Arbitrary pdf with Lognormal Kernels

In this subsection, the flexibility and the applicability of proposed model will be demonstrated by proving that an arbitrary probability density function (pdf) can be modeled as lognormal mixtures by using positive definite kernels with kernel smoothing method.

Kernel smoothing method of nonparametric methods are frequently used in data-inference processes, the mathematical expression of which is unknown [30, 31]. An approximate expression can be obtained through kernel smoothing method by using a positive definite kernel defined on the related objects set Z to a function f, the closed form mathematical expression of which can not be determined. This approximation is expressed as

$$f : z \in Z \rightarrow f_Z(z) = \sum_{i \in I} w_i \kappa(z_i, z), \tag{2.3}$$

where $\kappa(\cdot)$ is the kernel, $(z_i)_{(i\in I)}$ and $(w_i)_{(i\in I)}$ represent the family of known points and real coefficients, respectively. Any kernel must provide $\int_{-\infty}^{\infty}\kappa(u)du = 1 < \infty,\ \kappa(u) \geq 0, \forall u$ to be positive definite [32].

Let Y denote an RV with arbitrary pdf, which is used to model shadow fading. It can be expressed as a linear combination of non-negative or positive kernels. In other words, it can be shown as the mixture of positive definite kernels such as lognormal or Gaussian. Here, the used kernel expressions satisfy the abovementioned properties due to the fact that they are pdfs, and hence due to the axioms of probability they are positive definite.

It can be observed from (**2.1**) that shadow fading and received power have the same distribution, by considering the fact that distance, d, path loss exponent, n terms are fixed. Hereby it is more suitable to carry out the processes through the received power, which can be measured accurately through various high resolution devices, instead of unmeasurable shadow fading; Y is defined above to model the shadow fading, will be used for received power as well. Hence, let the received power levels (P_{Rx}) be defined as a RV, Y, and let $y = y_1, y_2, \ldots, y_N$ denote our dataset of realizations of Y, where N represents the length of our dataset. Here, each y_i represents the value of the i^{th} P_{Rx} measurement, with $y_i \in \mathbb{R}$. When the literature about the large scale fading is examined, it can be seen that each observation is lognormally distributed. Consistent with the existing literature and providing more generalised model, here, we propose that the corresponding pdf of mixture of lognormal distributions can be written as

$$Y \sim f_Y(y) = \sum_{k=1}^{\infty} \omega_k \mathscr{LN}(\mu_k, \sigma_k^2), \tag{2.4}$$

where ω_k's are the mixing proportions which are positive, μ_k and σ_k^2 are the mean and the variance of the k^{th} mixture component for $k = 1, 2, \ldots, \infty$. Here, $\sum_{k=1}^{\infty}\omega_k = 1$. In equation (**2.4**) lognormal kernel (pdf) is

$$\kappa(y; \mu, \sigma^2) = \frac{1}{y} C_y e^{-\varepsilon_y(\log(y)-\mu)^2},\ \kappa : Y \times Y \to \mathbb{R}, \tag{2.5}$$

where $\log(\cdot)$ represents the natural logarithm, $C_y = \frac{1}{\sqrt{\pi\varepsilon_y}}$ and $\varepsilon_y = \frac{1}{2\sigma^2}$.

The lognormal kernel expression in (**2.5**) is not symmetric since $\kappa(u) \neq \kappa(-u)$. Nevertheless, due to their convenience in a number of nonparametric estimation

methods, the use of symmetrical kernels is preferred. Accordingly, it would be more suitable to model shadow fading (possibly an arbitrary pdf) by using symmetrical Gaussian kernels instead of non-symmetrical lognormal kernels. With this motivation, in Proposition 1 it is shown that Gaussian kernel can be used instead of lognormal kernel.

Proposition 1. *Noting that each P_{Rx} measurement (y_i) is lognormal distributed, logarithm of each observation can be modeled as a univariate Gaussian distribution with unknown mean and variance. Hence it can be stated that for $X = \log(Y)$*

$$X \sim f_X(x) = \sum_{k=1}^{\infty} \pi_k \, \mathcal{N}(\eta_k, \nu_k^2) \quad i = 1, \ldots, N \,, \tag{2.6}$$

where π_k's are the mixing proportions (which are positive and sum up one), $k = 1, \ldots, \infty$ and $\mathcal{N}(\cdot)$ represents Gaussian kernel (pdf)

$$\kappa(x; \eta, \nu^2) = C_x e^{-\varepsilon_x (x-\eta)^2}, \; \kappa : X \times X \to \mathbb{R}, \tag{2.7}$$

where $C_x = \frac{1}{\sqrt{\pi \varepsilon_x}}$ and $\varepsilon_x = \frac{1}{2\nu^2}$.

Proof of Proposition 1: Assuming that X is a Gaussian RV, Y has lognormal pdf when $Y = \exp(X)$ [26, 33]. Hence, logarithms of observations are composed of a mixture of Gaussian RVs.

Note that the proposed model with infinite number of components is not practically applicable a finite number of components with negligible modeling error is necessary. In the following proposition, it is shown the convergence of the posted model as the number of components increase by proving that this error is bounded.

Proposition 2. *The arbitrary shadow fading (also the received power) pdf $f_X(x)$ can be represented with a summation of infinitely many Gaussian pdfs. This model can be obtained with a measurable error by using Gaussian kernel in finite numbers for a practically applicable model.*

Proof of Proposition 2: An approximation is made to actual pdf $f_X(x)$, the mathematical expression of which is not completely known, by using $S_f(x)$ function defined as the linear combination of finite Gaussian kernels, provided that the symmetrical univariate is kernel function $\kappa(x)$ in equation (**2.7**)

$$S_f(x) = \sum_{k=1}^{K} \xi_k \ \kappa(x_k, x), \ x \in X \subseteq \mathbb{R}. \tag{2.8}$$

Standard pointwise error bound between actual pdf $f_X(x)$, including infinite components, and estimated pdf $S_f(x)$, including finite components, is defined as follows by using Cauchy-Schwarz inequality [34]

$$\|f_X(x) - S_f(x)\|_\infty \leq C_{d,x} N^{-K} \|f\|_{\mathscr{H}(K,X)}, \tag{2.9}$$

where $C_{d,x}$ is a coefficient based on dimension ($C_{d,x}$ increases with d but d is equal to 1 in this work as kernel expression is univariate), N is the length of our dataset, k is the number of mixture component, $\|\cdot\|$ represents the inner product in Hilbert space $\mathscr{H}(\cdot)$. When the aforementioned inequality is analysed, Gaussian kernels provide arbitrarily high approximation order to infinitely smooth functions $f_X(x)$, i.e., with $K = \infty$.

In Proposition 1, it was shown that lognormal and Gaussian expressions could be used interchangeably. Starting from this demonstration, the approximation made via Gaussian kernels in (**2.8**) can be expanded by modeling an arbitrary pdf $f_Y(y)$ through lognormal kernels with a finite number of kernels as

$$S_f(y) = \sum_{k=1}^{K} \xi_k \ \kappa(y_k, y), \ y \in Y \subseteq \mathbb{R}. \tag{2.10}$$

Based on (**2.9**), it can be observed that the approximation error through the utilization of finite number lognormal kernels modeling the mathematical expression of shadow fading is bounded by

$$\|f_Y(y) - S_f(y)\|_\infty \leq C_y N^{-K} \|f\|_{\mathscr{H}(K,Y)}. \tag{2.11}$$

The explanations made in (**2.9**) are valid for this inequality.

2.3.2 Statistical Expressions

Using the pdf expression in (**2.4**), the cumulative distribution function (cdf) of lognormal mixture can be evaluated as

$$F_Y(y) = \int_0^y f_Y(u) du, \tag{2.12}$$

to yield

$$F_Y(y) = \frac{1}{2}\sum_{k=1}^{K} \omega_k \left[1 + \text{erf}\left\{\frac{log(y) - \mu_k}{\sqrt{2\sigma_k^2}}\right\}\right], \tag{2.13}$$

where $\text{erf}(\cdot)$ denotes the error function defined as

$$\text{erf}(u) = \frac{2}{\pi}\int_0^u \exp(-t^2)dt, \tag{2.14}$$

and μ_k, σ_k^2 are the mean and the variance of k^{th} mixture component, respectively.

The moment generating function (MGF) of lognormal mixture can be defined as

$$M_Y(s) = \mathbb{E}(\exp(-sy)), \tag{2.15}$$

where $\mathbb{E}(\cdot)$ represents the expectation operator. Hence, it is expressed as

$$M_Y(s) = \int_0^\infty \exp(-sy) f_Y(y)dy, \tag{2.16}$$

but there is no generalized closed form expression for the real s. Nevertheless, it can be readily expressed by a series expansion based on Gauss-Hermite integration [35] and so the MGF of lognormal mixture for real s can be obtained as

$$\begin{aligned} M_Y(s) &= \int_0^\infty \exp(-sy)\sum_{k=1}^{K}\omega_k \frac{1}{y\sqrt{2\pi\sigma_k^2}}\exp\left\{-\frac{(\log(y)-\mu_k)^2}{2\sigma_k^2}\right\}dy \\ &= \frac{1}{\sqrt{\pi}}\sum_{k=1}^{K}\omega_k\int_{-\infty}^{\infty}\exp[-s\exp(\sqrt{2}\sigma_k\zeta+\mu_k)]\exp(-\zeta^2)d\zeta \\ &= \frac{1}{\sqrt{\pi}}\sum_{k=1}^{K}\omega_k\sum_{l=1}^{L}\varphi_l\exp[-s\exp(\sqrt{2}\sigma_k\vartheta_l+\mu_k)] + R_L\,, \end{aligned} \tag{2.17}$$

where φ_l is the Hermite series weights in the final expression denoting the Gauss-Hermite series expansion of the MGF function, L represents the Hermite integration order, and R_L is a remainder term that decreases as L increases. The weights and abscissas ϑ_l for L up to 20 are tabulated in [35]. From it, the Gauss-Hermite representation of lognormal mixture MGF can be defined by removing R_L as

$$\hat{M}_Y(s) = \sum_{k=1}^{K}\omega_k\sum_{l=1}^{L}\varphi_l\exp[-s\exp(\sqrt{2}\sigma_k\vartheta_l+\mu_k)]. \tag{2.18}$$

The r^{th} moment of the lognormal mixture RV Y, $m_Y(r)$ can be evaluated as $m_Y(r) = \mathbb{E}(y^r)$ to yield

$$m_Y(r) = \sum_{k=1}^{K} \frac{\omega_k}{\sqrt{2\pi\sigma_k^2}} \int_0^{\infty} y^{r-1} \exp\left\{ - \frac{(\log(y) - \mu_k)^2}{2\sigma_k^2} \right\} dy. \quad (\mathbf{2.19})$$

Although general closed form solution of the moments, belonging to the lognormal mixture distribution, can not be obtained easily by using (**2.19**), the first and second moments of the mixture model can be calculated as follows by using weighted superposition

$$m_Y(1) = \mu_Y = \sum_{k=1}^{K} \omega_k \mu_k, \quad (\mathbf{2.20})$$

$$m_Y(2) = \sigma_Y^2 = \sum_{k=1}^{K} \omega_k \sigma_k^2. \quad (\mathbf{2.21})$$

Consequently, mean and variance of the lognormal mixture, proposed for a more accurate shadow fading modeling, can be obtained easily through (**2.20**) and (**2.21**).

In this subection, it is shown that logarithm of each observation of P_{Rx} can be modeled as a Gaussian distribution with a mean and a variance by Proposition 1. Proposition 2 demonstrate that the arbitrary pdf, $f(x)$, can be represented with a summation of infinitely many Gaussian pdfs.

2.3.3 Obtaining Mixture Model Parameters

Mathematical descriptions of the used methods in parameter estimation for modeling shadow fading measurements by using finite positive definite kernels with an acceptable error rate are given in this subsection.

2.3.3.1 Dirichlet Process Mixture Model for Lognormal Mixtures

DPM models are useful tools for analysing mixture models with an unknown number of mixture components [36]. They can be applied for decomposing an estimated pdf into lognormal mixture components.

To be able to use DPM models, the parameters specifying the observation model are RVs and the observations need to be exchangeable. Although RVs are dependent among themselves, exchangeability demonstrates the assumption that RVs are independent of their observation orders [22, 37]. Noting that any P_{Rx} measurement set

is exchangeable for the same d, so we can use DPM to determine mixture components. Mathematical expressions and detailed statement of DPM are explained in App. A.1.

2.3.3.2 Expectation Maximization Algorithm

EM algorithm is a method that can be used to simplify the challenging maximum likelihood estimation (MLE) problems, which are typically experienced in mixture models and can not be analytically solved. Especially, because of the traditional MLE methods do not provide a closed form expression for Gaussian mixture models, the EM algorithm is originally developed to solve this problem.

The use of this algorithm can be expanded within the estimation of parameters belonging to mixture of lognormal distributions, by considering the Gauss-LN relationship provided in Proposition 1, although it was originally developed for Gaussian mixture models.

Within the scope of this thesis, the estimation of parameters related with the estimated pdf expression, which is the approximation carried out on actual pdf of lognormal shadowing by using a finite number of positive definite kernels, is suitable for the utilization of EM algorithm. Noting that the implementation of EM algorithm, which is one of the parametric methods, is much easier compared to the nonparametric DPM model, it is generally preferred for such problems despite the assumption that the number of mixture components are known a priori.

Let assume that X is a random variable that represents received power signal level, $x_n \in \{x_1, x_2, x_3, \ldots, x_N\}$ is the observation cluster that consist of N received power signals, $\theta = (\mu_k, \sigma_k^2)$ denotes the parameter values of k^{th} component of the mixture model, and $p(x)$ represents the probability density function of observations. In MLE method it is expected to find θ values that maximize the $p(x|\theta)$ conditional probability. However, MLE method does not provide a closed form expression for $p(x|\theta)$ [38]. EM algorithm is an iterative method that consists of iterations of the expectation step (E-step) and the maximization step (M-step). In every iteration, firstly in E-step the probability for every observation that belonging to any component is estimated, then by using these values in M-step the parameter values of components are produced. Iteration stops by approaching the maximum iteration number or an acceptable error rate. Let $\Theta^i =$

$\{w_1^i, w_2^i, \ldots, w_K^i, \theta_1^i, \theta_2^i, \ldots, \theta_K^i\}$ denotes the cluster that has initial parameter values and $y^i = \{y_1^i, y_2^i, y_3^i, \ldots, y_N^i\}$ is the cluster that consists of new estimated received power signal levels which are created by using x_i values. New mixture models parameters are obtained, where weights are

$$w_k^{new} = \frac{1}{N} \sum_{i=1}^{N} p(y_i|x_i, \theta^i), \tag{2.22}$$

mean values are calculated as

$$\mu_k^{new} = \frac{\sum_{i=1}^{N} x_i p(y_i|x_i, \theta^i)}{\sum_{i=1}^{N} p(y_i|x_i, \theta^i)}, \tag{2.23}$$

and variances become

$$\sigma_k^{2new} = \frac{\sum_{i=1}^{N} p(y_i|x_i, \theta^i)(x_i - \mu_i^{new})^2}{\sum_{i=1}^{N} p(y_i|x_i, \theta^i)}, \tag{2.24}$$

as given in equation (**2.4**) [28, 29, 39].

Flow chart of EM algorithm is shown in Fig. 2.1.

2.4 Other Candidate Distributions: Gamma and Weibull

As it is indicated in Section 2.2, since the lognormal distribution does not provide perfect solution for shadow fading, other distributions are taken into account. In order to investigate other distributions than lognormal, Gamma and Weibull distributions are also convenient to be considered for shadow fading in literature [18], [19]. Gamma distribution with a shape parameter and b scale parameter is

$$f_\phi(t|a,b) = \begin{cases} \frac{1}{b^a \Gamma(a)} t^{a-1} e^{\frac{-t}{b}} & 0 \leq t, \\ 0 & \text{otherwise}, \end{cases} \tag{2.25}$$

where $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} \, \mathrm{d}x$, and the Weibull distribution with k shape parameter and λ scale parameter is

$$f_\phi(\phi|\lambda,k) = \begin{cases} \frac{k}{\lambda} \left(\frac{\phi}{\lambda}\right)^{k-1} e^{\left(\frac{-\phi}{\lambda}\right)^k} & 0 \leq \phi, \\ 0 & \text{otherwise}. \end{cases} \tag{2.26}$$

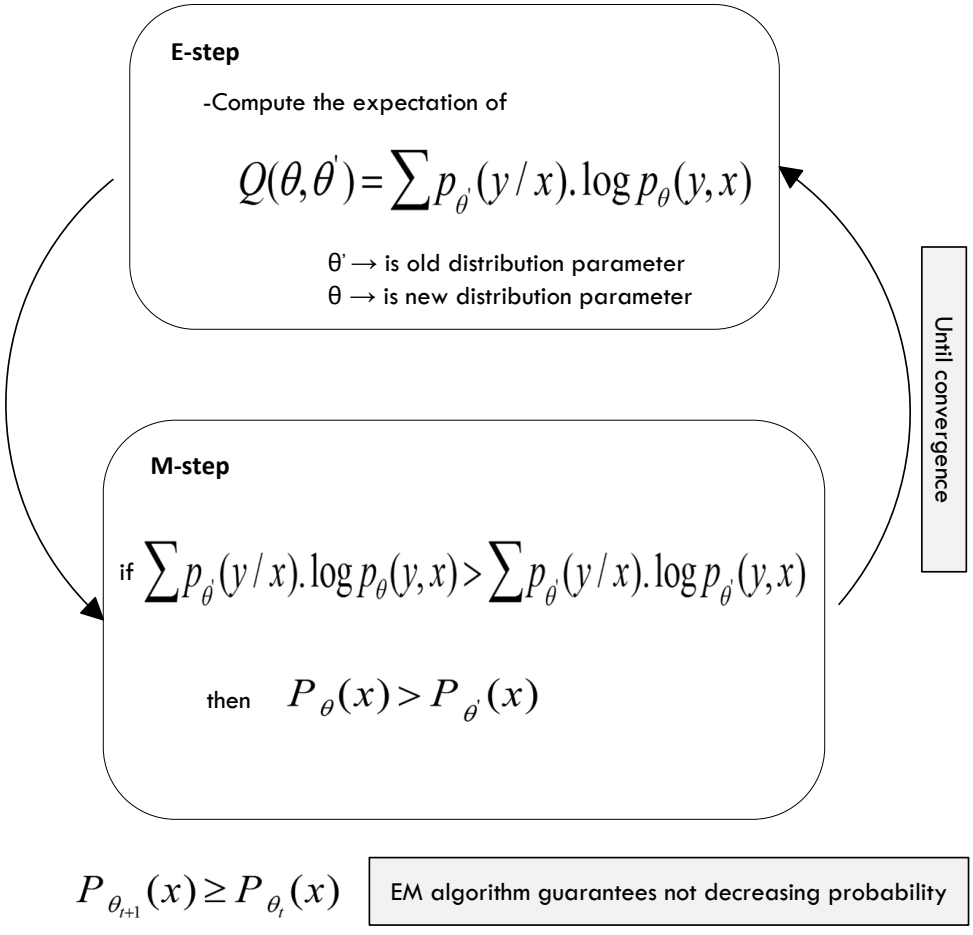

Figure 2.1: Flow chart of EM algorithm.

Gamma and Weibull distributions are mathematically practical to examine closed forms for performance metrics. Mathematical expressions of $C(\phi)$ and p_{out} for Gamma and Weibull random variables are described in Chapter 4.

Utilized pdfs and parameter expressions are shown in Table 2.1.

Table 2.1: Utilized pdfs for shadow fading analysis

	Probability Density Function	Mean
Gauss	$f_\phi(\phi\|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\phi-\mu)^2}{2\sigma^2}}$	μ
Gauss-k	$f_\phi(\phi\|\omega_k,\mu_k,\sigma_k) = \sum_{k=1}^{K} \omega_k f_{\phi k}(\phi\|\mu_k,\sigma_k)$	$\sum_{k=1}^{K} \omega_k \mu_k$
Weibull	$f_\phi(\phi\|\lambda,k) = \frac{\phi}{\lambda}^{k-1} . e^{\left(\frac{-\phi}{\lambda}\right)^k}$	$\lambda\Gamma\left(1+\frac{1}{k}\right)$
Gamma	$f_\phi(t\|a,b) = \frac{1}{b^a\Gamma(a)} t^{a-1} e^{\frac{-t}{b}}$	$a.b$
Lognormal	$f_\phi(\phi\|\mu,\sigma) = \frac{1}{\phi\sqrt{2\pi\sigma^2}} e^{\frac{-(log(\phi)-\mu)^2}{2\sigma^2}}$	$e^{\mu+\frac{\sigma^2}{2}}$
Lognormal-k	$f_\phi(\phi\|\omega_k,\mu_k,\sigma_k) = \sum_{k=1}^{K} \omega_k f_{\phi k}(\phi\|\mu_k,\sigma_k)$	$\sum_{k=1}^{K} \omega_k e^{\mu+\frac{\sigma^2}{2}}$

In this chapter, wireless propagation model and lognormal shadow fading are explained. Lognormal mixture shadow fading model is proposed. Additionally, in order to find the parameters of the lognormal mixture components, EM and DPM algorithms are explained. Then other candidate shadow fading distributions (Gamma and Weibull) are demonstrated.

In the next Chapter, by using two real-life datasets, proposed and candidate distributions for shadow fading will be compared with the evaluation metrics by measuring difference between obtained and the actual pdfs.

3. NUMERICAL RESULTS FOR DISTRIBUTIONS

3.1 Datasets Used for Shadow Fading Analysis

In order to investigate shadow fading distributions and analyze $C(\phi)$ and p_{out}, two datasets are used.

Dataset I is obtained in an urban macrocell environment in the GSM 900 downlink band by us based on real-life data. By using Anrtisu MS2711E spectrum analyzer, 5 individual tracks with over 10000 measurement data points are created. Channel sounding is carried out at 940.51 MHz frequency on a control channel of a GSM base station located at a height of 6 meters. Spectrum analyzer and base station is shown in Fig. 3.1. Note that the transmit power level, P_{Tx} of control channels are fixed. At every track point measurements are taken in a stationary fashion with 17ms sweep time on the ground level to obtain shadow fading effects. In Table 3.1, the test scenario and measurement setup for Dataset I are shown. The track point locations and the base station location are shown in Fig. 3.2.

Table 3.1: Measurement setup

GENERAL MEASUREMENT PROPERTIES	
Scenario	Urban macrocell
Location	Istanbul Technical University Maslak Campus
Measurement setup	1 Base Station, 5 tracks, 10041 data points
Track distances from base station	5m, 10m, 15m, 24m, 40m
CHANNEL SOUNDER PROPERTIES	
Make/Model	Anritsu MS2711E
Center frequency	940.51MHz
Sweep time	17ms

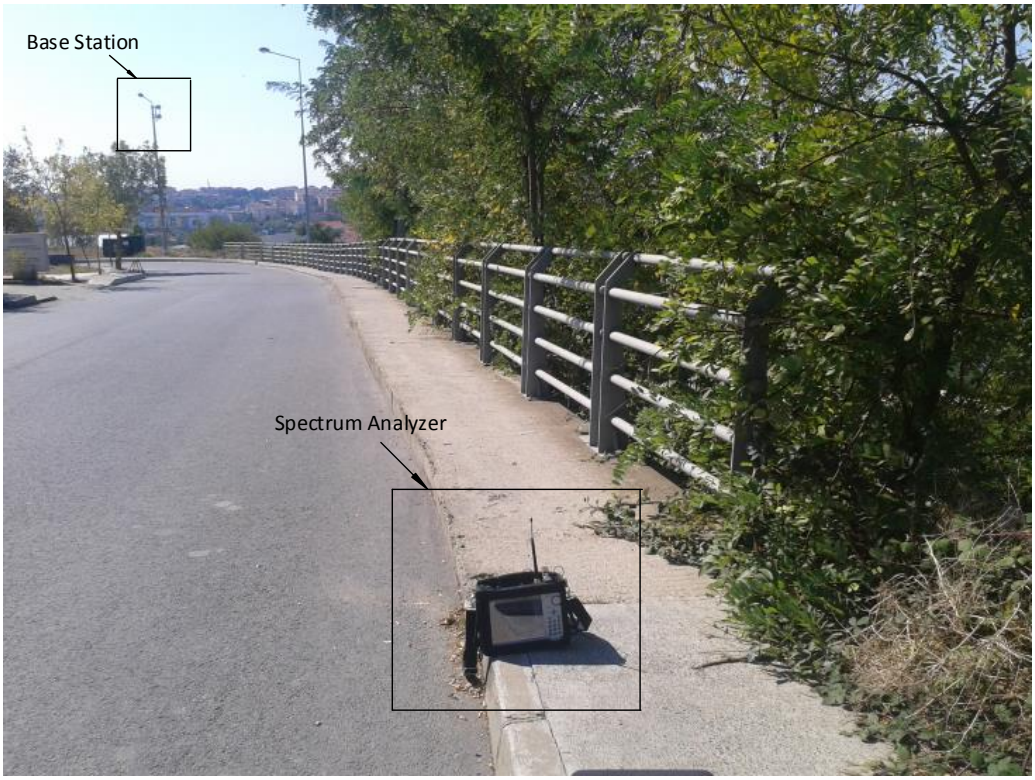

Figure 3.1: Measurement setup of Dataset I at Istanbul Technical University Maslak Campus. Spectrum analyzer and base station.

The scatter plot of the measurements are shown in Fig. 3.3(a). The dataset, resembling the scatter plots of [1,40,41], is also used to calculate the path loss exponent, $n = 2.52$. The average path loss values are also plotted in Fig. 3.3(a). In our analysis we will concentrate on two analysis of Dataset I with normalized means (means are shifted to zero); the complete set of all 10041 measurements, and the track point 5 (P5) that is composed of the 2016 measurements at 40 meters away form the transmitter. Their normalized histograms are also shown in Fig. 3.3(b) and 3.3(c), respectively.

In order to verify that Dataset I measurements are subject to independent shadow fading, the normalized autocorrelation functions (ACFs) of P5 and complete set of measurements in Fig. 3.4(a) and 3.4(b) are plotted. This is consistent with the reported results in [40–45]. Additionally, some of the literature about shadow fading is focused on determination of the spatial correlation properties, instead of the statistical distributions [42]. According to the Gudmundson model [44], the spatial correlation model for shadow fading is determined with an exponential decay function with distance. To demonstrate consistency of Dataset I measurements, we plot the

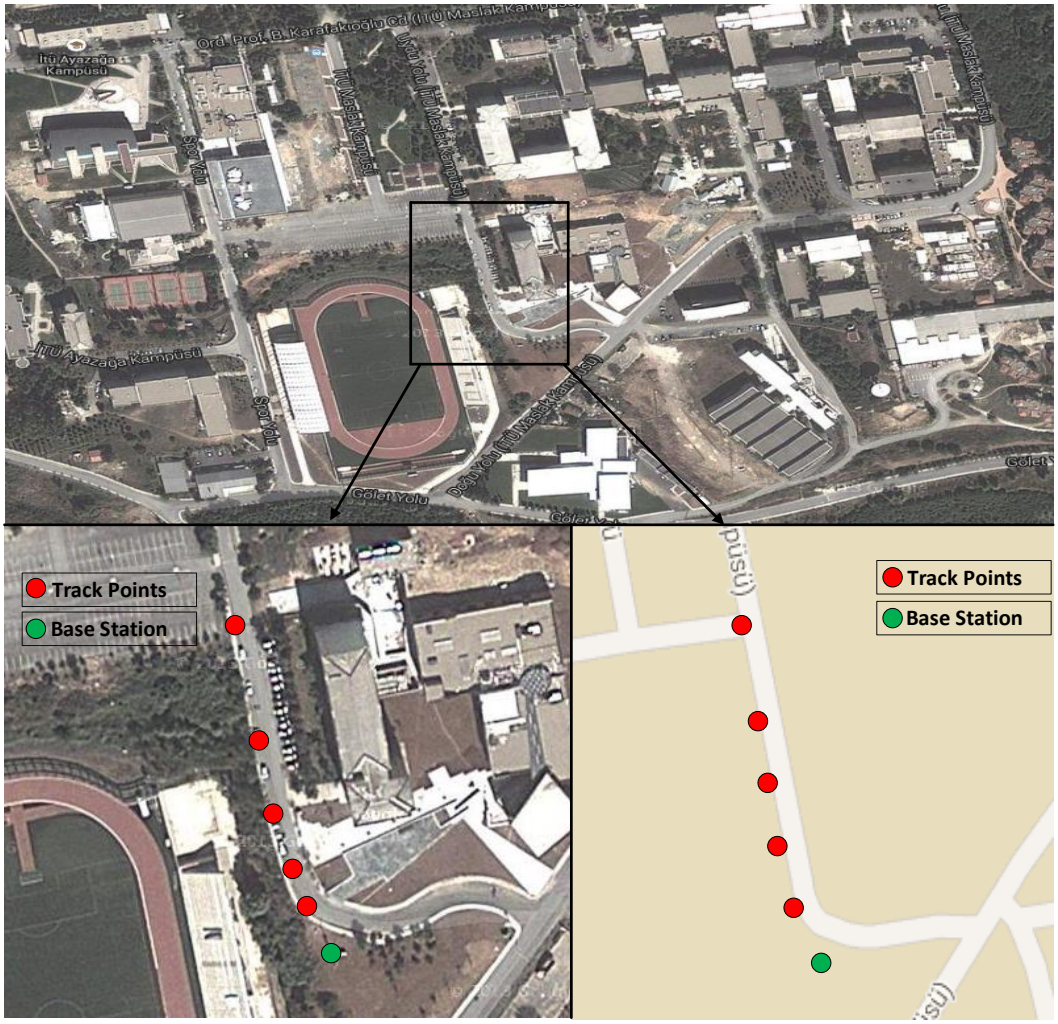

Figure 3.2: Measurement setup of Dataset I at Istanbul Technical University Maslak Campus. The control channel of a real-life functioning GSM network is used as the reference transmitter.

cross-correlation between track points and the first track point in Fig. 3.4(c) where we can observe that the measurements are uncorrelated. This obtained decorrelation distance of approximately 5 meters is a consistent result with the literature [42,43,45].

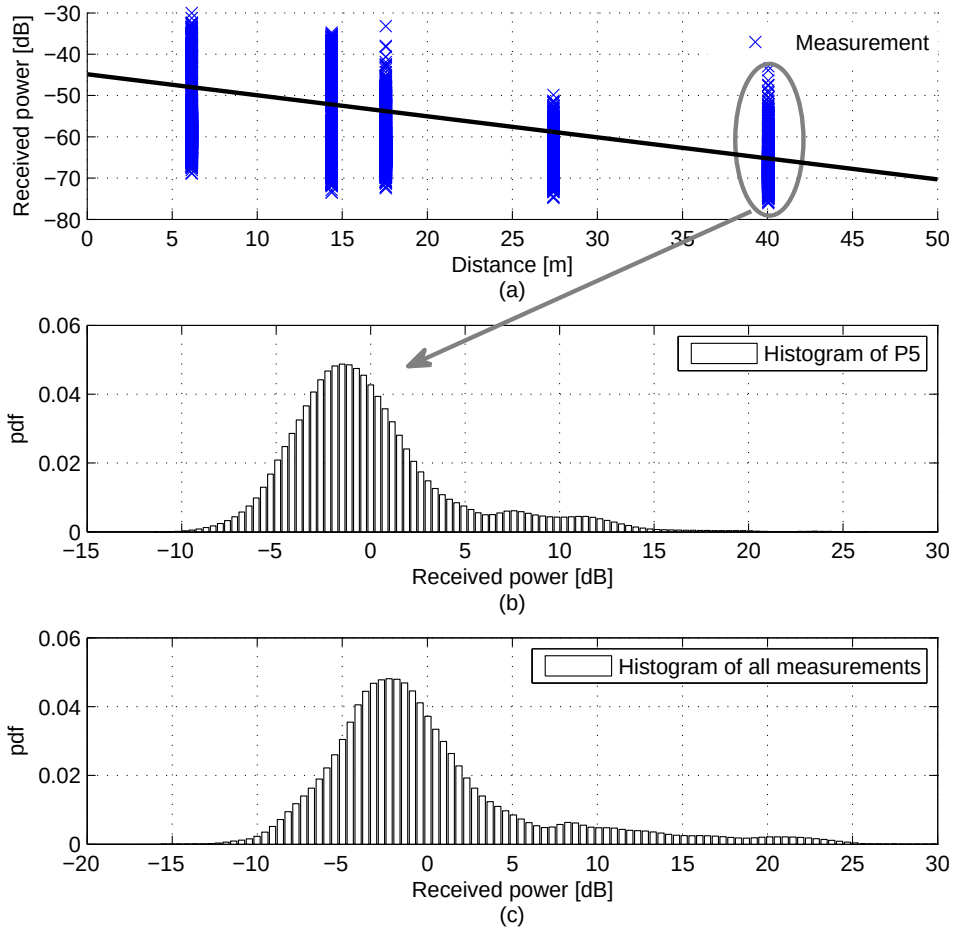

Figure 3.3: (a) Scatter plot of the 10041 measurements at 5 distinct track points. (b) Histogram of 2016 measurements at 40 meters (P5). (c) Histogram of all measurement points for Dataset I.

Dataset II, also is an urban macrocell channel dataset [27]. According to the dataset properties [27], channel measurement is performed in the 3GPP Long Term Evaluation (LTE) band in Ilmenau city center at Germany. By using this dataset, collected received powers from 3 base station 26250 non-line-of-sight (NLOS) and 5391 line-of-sight (LOS) track points at 2.4966 GHz are investigated. According to [8], the dataset is convenient for large scale parameter analysis and measurements in dataset are analyzed for shadow fading as well. Fig. 3.5 shows the sample scatter plot and the normalized histogram of received powers that are collected from Base Station 1 (BS1) in LOS environment, while Fig. 3.6 shows the same plots for the received powers that are collected from Base Station (BS3) in NLOS environment.

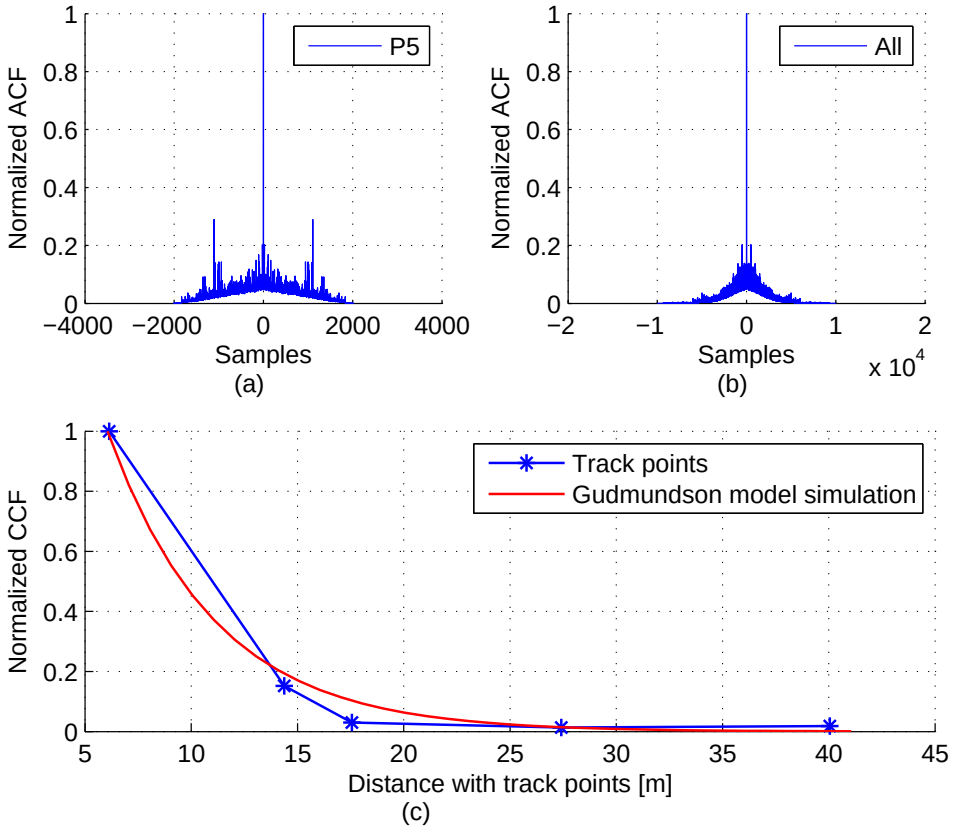

Figure 3.4: (a) Normalized autocorrelation function (ACF) of Dataset I-P5. (b) Normalized ACF of Dataset I all measurements. (c) Normalized cross-correlation function (CCF) of the Dataset I track points showing the decorrelation distance.

In order to demonstrate that Dataset II measurements are subject to independent shadow fading, the normalized autocorrelation functions (ACFs) of LOS and NLOS measurements in Fig. 3.7(a) and 3.7(b) are plotted. This is consistent with the reported results in [40–45].

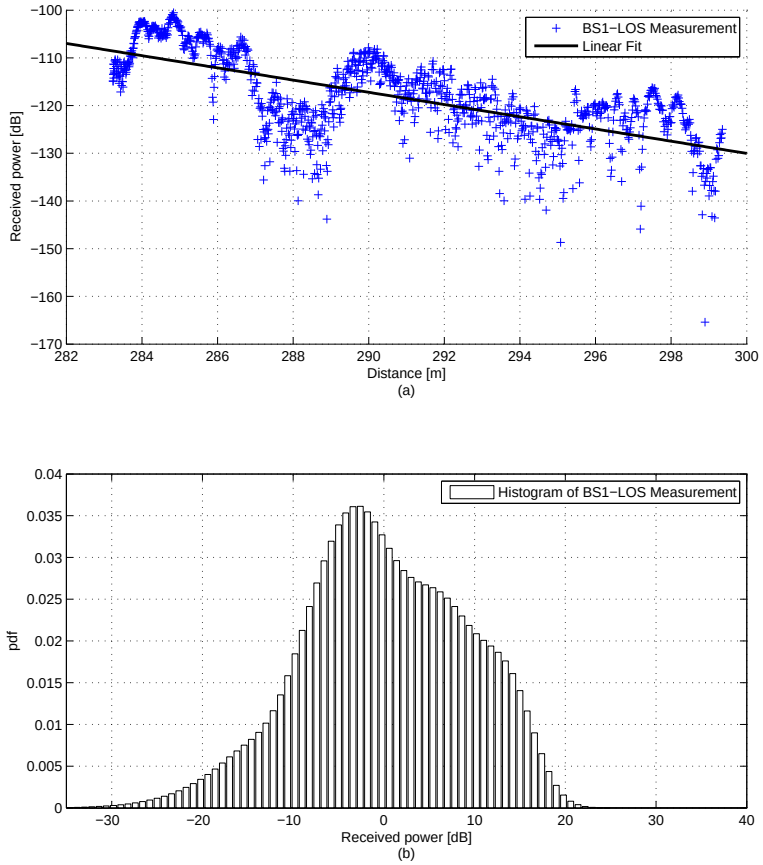

Figure 3.5: (a) Scatter plot of the Base Station 1 (BS1) measurements in Dataset II-LOS environment (b) Histogram of the Base Station 1 (BS1) measurements in Dataset II-LOS environment.

3.2 Evaluation Metrics: MRD, WMRD and KL Divergence

In this study, error vector norm techniques that are used frequently in literature are used to measure difference between the obtained and the actual pdfs [38]. These include the mean relative difference (MRD) the weighted MRD (WMRD) and the Kullback-Leibler (KL) divergence. Additionally Kolmogorov-Smirnov (K-S) goodness-of-fit test with the confidence level $\alpha = 0.05$ (corresponding to the null-hypothesis rejection level of 5%) is used to determine whether the mixture model is suitable for the actual histogram.

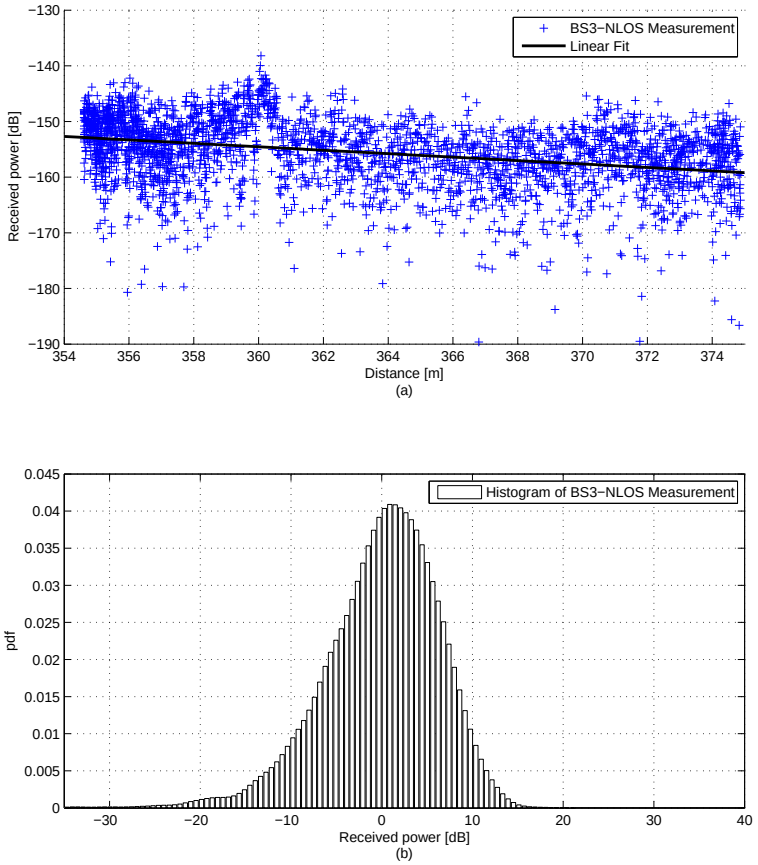

Figure 3.6: (a) Scatter plot of the Base Station 3 (BS3) in Dataset II-NLOS environment (b) Histogram of the Base Station 3 (BS3) measurements in Dataset II-NLOS environment.

The values of MRD, WMRD and KL divergence are respectively calculated by using the following equations

$$MRD = \frac{1}{T}\sum_{t}\frac{|y_t - \hat{y}_t|}{(y_t + \hat{y}_t)\times 0.5}, \tag{3.1}$$

$$WMRD = \frac{\sum_{t}|y_t - \hat{y}_t|}{\sum_{t}(y_t + \hat{y}_t)\times 0.5}, \tag{3.2}$$

$$D_{KL}(\hat{y}||y) = \sum_{t}\hat{y}_t log_2\frac{\hat{y}_t}{y_t}. \tag{3.3}$$

where D_{KL} denotes the KL divergence distance, T is maximum P_{Rx}, y_t represents the number of the t^{th} value P_{Rx} observations and $\hat{y}_t$ is the estimated value of y_t.

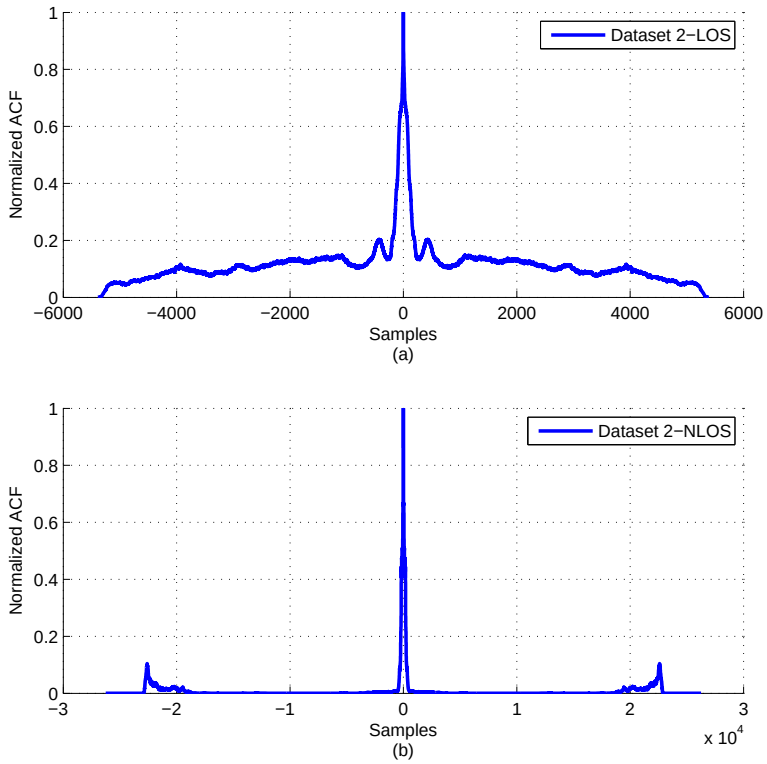

Figure 3.7: (a) Normalized autocorrelation function (ACF) of Dataset II-LOS environment. (b) Normalized ACF of Dataset II-NLOS environment.

3.3 Numerical Results

According to [8], histogram of the Dataset II that is also used in this thesis, corresponds to a lognormal distribution (normal distribution in dB) as expected. Fig. 3.10 and Fig. 3.11 show the normalized dataset histogram of P_{Rx} in dB, the Gaussian mixture pdfs (in dB scale) with several mixtures which are found by using EM and DPM algorithms, Gamma and Weibull pdf estimates for LOS and NLOS environment respectively for Dataset II, while Fig. 3.8 and Fig. 3.9 show the same results for one point (P5) and all measurements for Dataset I. It can be clearly seen that one Gaussian density function (i.e. the universally accepted lognormal shadowing model) is not a good fit for the measured data density function, while mixture models fit conceivably better.

In the Table 3.2 and 3.3, the accuracy of the proposed shadow fading models are shown that the results obtained from both goodness-of-fit tests, error vector norm based techniques (WMRD, MRD) and KL divergence. Results are consistent with the literature [26,46]. According to the Table 3.2 and Table3.3, proposed mixture modeling

has better MRD, WMRD and D_{KL} values than other models. The obtained parameters and the outcomes of error metrics analysis for Dataset II-LOS measurements are summarized in Table 3.4, from which it can be seen that the model obtained by using EM consisting of four mixture components is the most accurate representation of the measured power levels, with the smallest WMRD (0.3232,) value, while five component mixture has smallest D_{KL} (0.0021) value. Table 3.5 represents the error metrics for Dataset II-NLOS measurements. It can be seen that the mixture models are obtained by using EM and DPM have the smallest evaluation metric values.

Both Dataset I and Dataset II measurements demonstrate that shadow fading can be modeled more accurately with mixture models according to some evaluation metrics to determine the optimum mixture component numbers for shadow fading.

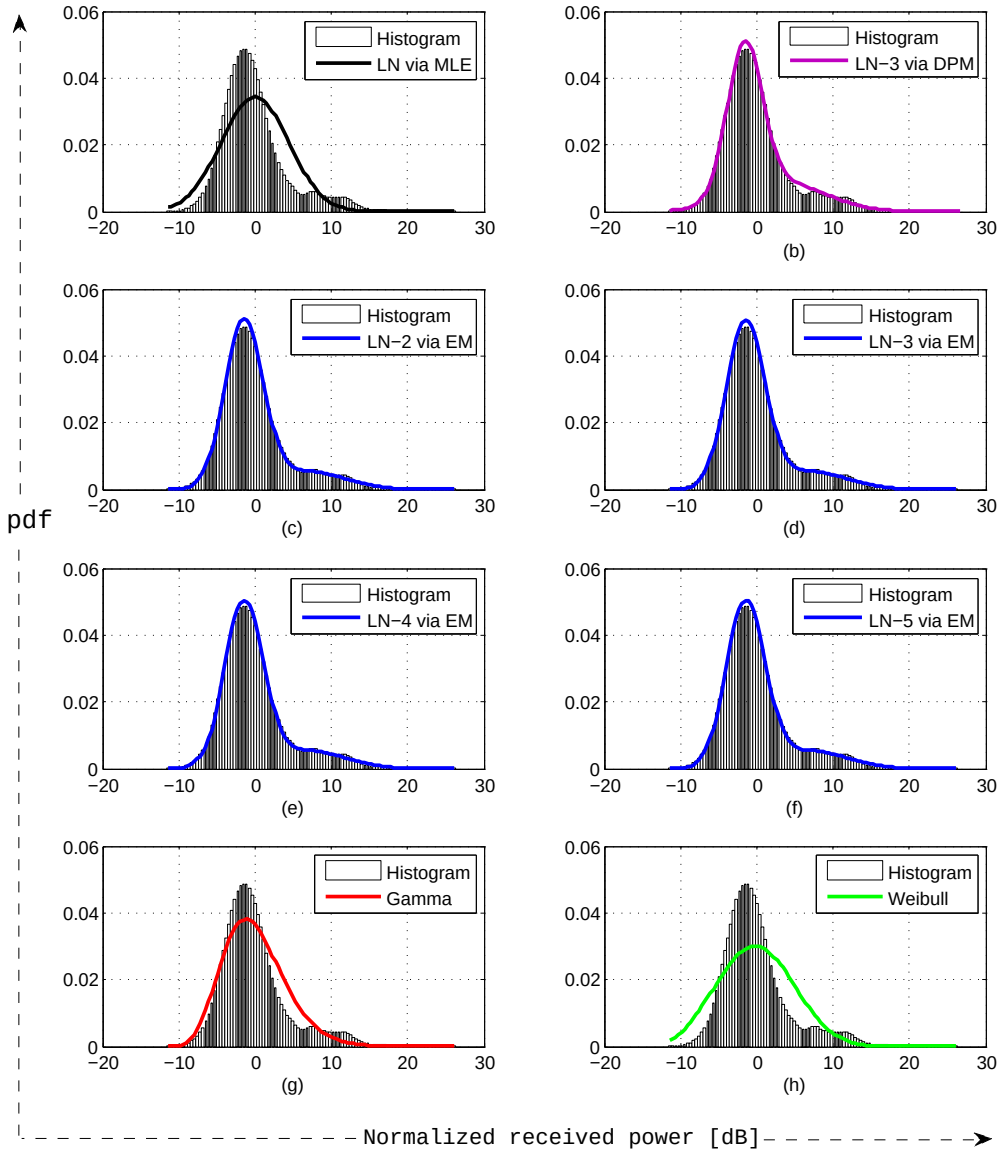

Figure 3.8: Normalized histograms of P_{Rx} and pdf estimates for Dataset I one point (P5). (a) Lognormal, (b) DPM (results in a three component mixture), (c) Lognormal-2 obtained with EM, (d) Lognormal-3 obtained with EM (e) Lognormal-4 obtained with EM (f) Lognormal-5 obtained with EM, (g) Gamma pdf estimate, (h) Weibull pdf estimate.

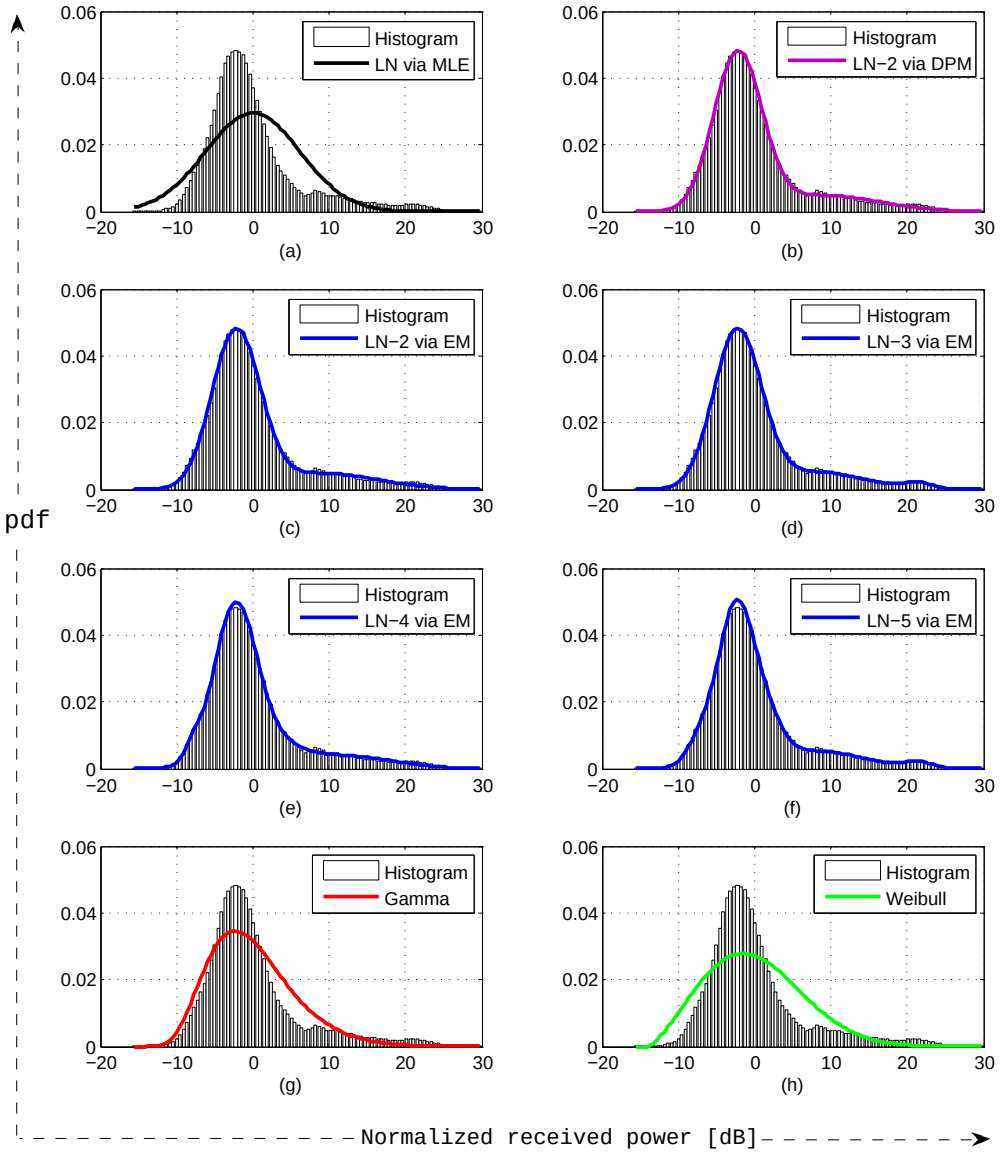

Figure 3.9: Normalized histograms of P_{Rx} and pdf estimates for Dataset I all measurements. (a) Lognormal, (b) DPM (results in a two component mixture), (c) Lognormal-2 obtained with EM, (d) Lognormal-3 obtained with EM (e) Lognormal-4 obtained with EM (f) Lognormal-5 obtained with EM, (g) Gamma pdf estimate, (h) Weibull pdf estimate.

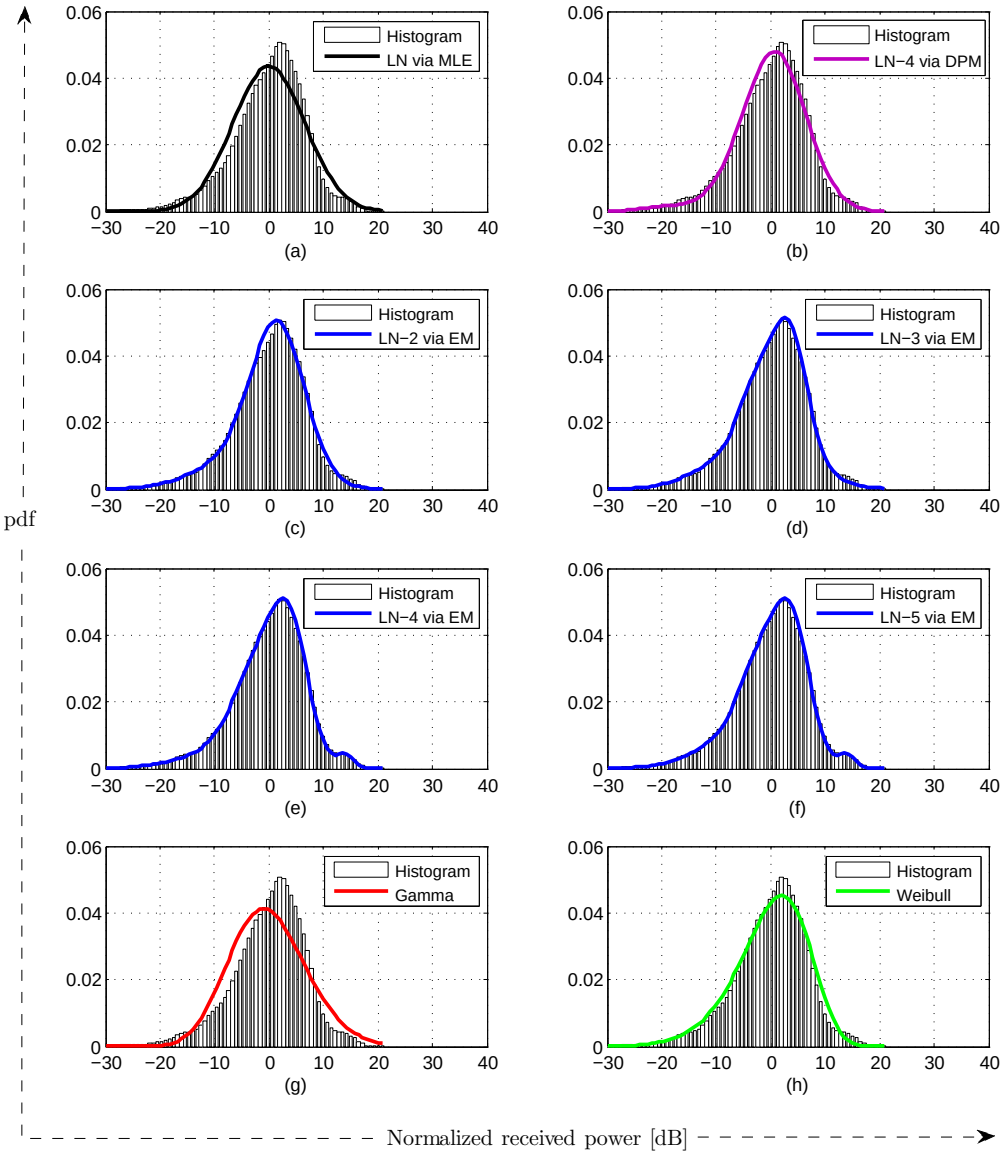

Figure 3.10: Normalized histograms of P_{Rx} and pdf estimates for Dataset II-LOS environment. (a) Lognormal, (b) DPM (results in a four component mixture), (c) Lognormal-2 obtained with EM, (d) Lognormal-3 obtained with EM (e) Lognormal-4 obtained with EM (f) Lognormal-5 obtained with EM, (g) Gamma pdf estimate, (h) Weibull pdf estimate.

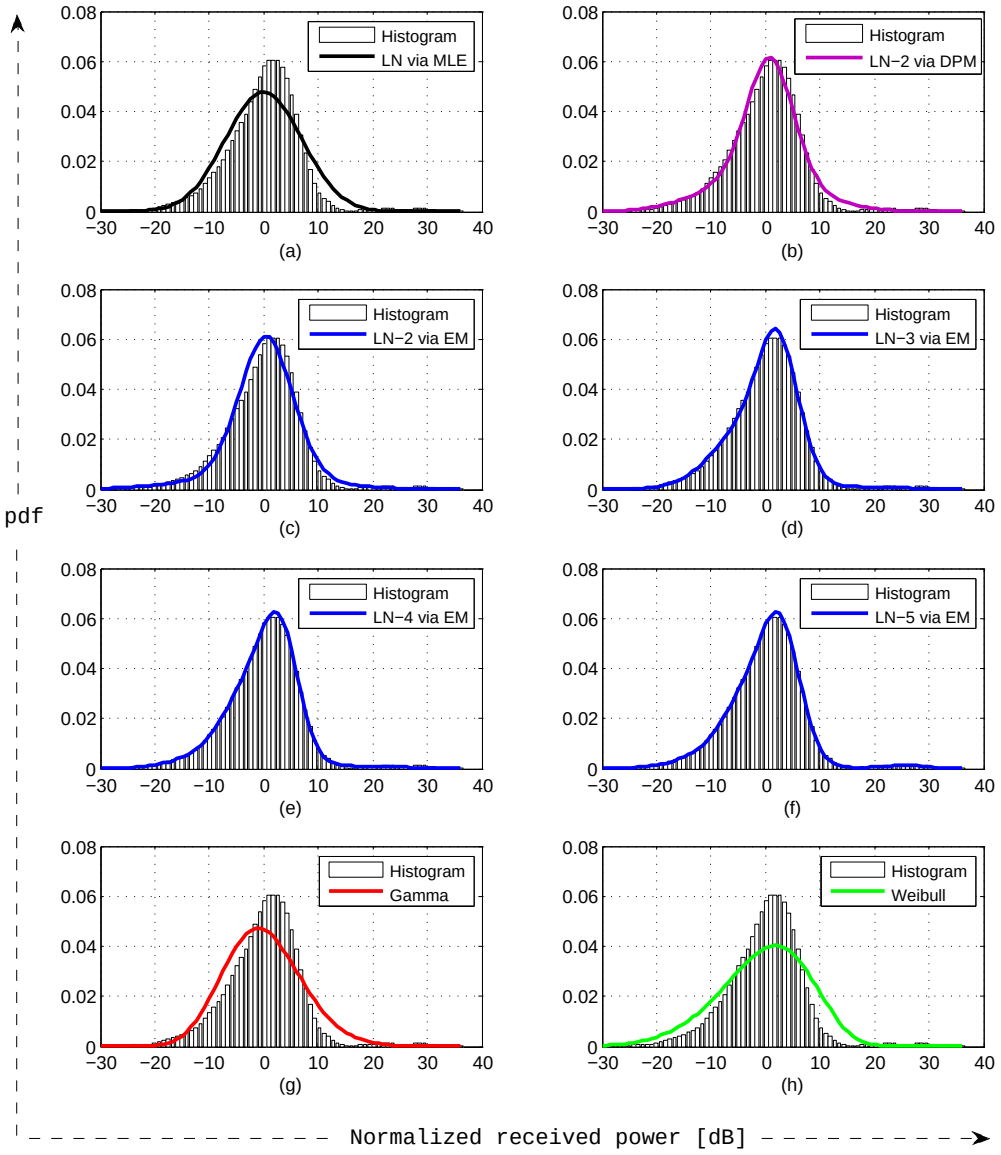

Figure 3.11: Normalized histograms of P_{Rx} and pdf estimates for Dataset II-NLOS environment. (a) Lognormal, (b) DPM (results in a two component mixture), (c) Lognormal-2 obtained with EM, (d) Lognormal-3 obtained with EM (e) Lognormal-4 obtained with EM (f) Lognormal-5 obtained with EM, (g) Gamma pdf estimate, (h) Weibull pdf estimate.

Table 3.2: Parameters and error metrics for pdf estimates for Dataset I-one point (P5).

	Parameters			MRD	WMRD	D_{KL}	K-S
LN	$\mu = 0$	$\sigma^2 = 19.54$	$\omega = 1$	0.0314	0.9182	0.1616	*Passed*
LN-3 via DPM	$\mu_1 = -0.6837$	$\sigma_1^2 = 10.5749$	$\omega_1 = 0.4437$	0.0301	0.8872	0.0103	*Passed*
	$\mu_2 = -1.426$	$\sigma_2^2 = 5.7561$	$\omega_2 = 0.4576$				
	$\mu_3 = 10.344$	$\sigma_3^2 = 12.4891$	$\omega_3 = 0.0986$				
LN-2	$\mu_1 = -1.4672$	$\sigma_1^2 = 6.2686$	$\omega_1 = 0.1788$	0.0314	0.9013	0.0039	*Passed*
	$\mu_2 = 6.7407$	$\sigma_2^2 = 25.1841$	$\omega_2 = 0.8212$				
LN-3	$\mu_1 = 6.6531$	$\sigma_1^2 = 25.4421$	$\omega_1 = 0.1818$	0.0349	0.9007	0.0038	*Passed*
	$\mu_2 = -2.0764$	$\sigma_2^2 = 5.5541$	$\omega_2 = 0.5789$				
	$\mu_3 = 0.0314$	$\sigma_3^2 = 4.8294$	$\omega_3 = 0.2393$				
LN-4	$\mu_1 = 10.06$	$\sigma_1^2 = 6.7252$	$\omega_1 = 0.0734$	0.035	0.9007	0.0038	*Passed*
	$\mu_2 = 3.02$	$\sigma_2^2 = 12.82$	$\omega_2 = 0.1261$				
	$\mu_3 = -1.5443$	$\sigma_3^2 = 6.1475$	$\omega_3 = 0.7947$				
	$\mu_4 = 18.6747$	$\sigma_4^2 = 8.6$	$\omega_4 = 0.0058$				
LN-5	$\mu_1 = -0.3281$	$\sigma_1^2 = 4.6332$	$\omega_1 = 0.4648$	0.0349	0.8966	0.0037	*Passed*
	$\mu_2 = -2.7587$	$\sigma_2^2 = 3.5465$	$\omega_2 = 0.3305$				
	$\mu_3 = 6.6931$	$\sigma_3^2 = 25.1454$	$\omega_3 = 0.1812$				
	$\mu_4 = -5.043$	$\sigma_4^2 = 0.0042$	$\omega_4 = 0.0096$				
	$\mu_5 = -7.1629$	$\sigma_5^2 = 1.0845$	$\omega_5 = 0.0140$				
Gamma	$a = 137.0122$	$b = 0.3649$		0.0328	0.9041	0.0464	*Passed*
Weibull	$\lambda = 52.1583$	$k = 9.4865$		0.0309	0.9054	0.1098	*Passed*

Table 3.3: Parameters and error metrics for pdf estimates for Dataset I-all measurements.

	Parameters			MRD	WMRD	D_{KL}	K-S
LN	$\mu = 0$	$\sigma^2 = 38.482$	$\omega = 1$	0.0208	0.7740	0.2424	*Passed*
LN-2 via DPM	$\mu_1 = 5.4437$	$\sigma_1^2 = 66.8474$	$\omega_1 = 0.3009$	0.0250	0.7440	0.0120	*Passed*
	$\mu_2 = -2.3449$	$\sigma_2^2 = 8.0904$	$\omega_2 = 0.6990$				
LN-2	$\mu_1 = 9.217$	$\sigma_1^2 = 52.96$	$\omega_1 = 0.193$	0.0253	0.7456	0.0065	*Passed*
	$\mu_2 = -2.205$	$\sigma_2^2 = 9.86$	$\omega_2 = 0.807$				
LN-3	$\mu_1 = 21.3471$	$\sigma_1^2 = 3.7928$	$\omega_1 = 0.0194$	0.0253	0.7461	0.0017	*Passed*
	$\mu_2 = 8.3905$	$\sigma_2^2 = 34.3438$	$\omega_2 = 0.1675$				
	$\mu_3 = -2.2367$	$\sigma_3^2 = 9.8221$	$\omega_3 = 0.8132$				
LN-4	$\mu_1 = -7.7488$	$\sigma_1^2 = 0.8417$	$\omega_1 = 0.0213$	0.0254	0.7457	0.0058	*Passed*
	$\mu_2 = -1.2826$	$\sigma_2^2 = 14.2437$	$\omega_2 = 0.4006$				
	$\mu_3 = -2.4930$	$\sigma_3^2 = 6.0049$	$\omega_3 = 0.4148$				
	$\mu_4 = 10.4876$	$\sigma_4^2 = 47.778$	$\omega_4 = 0.1633$				
LN-5	$\mu_1 = 1.9663$	$\sigma_1^2 = 4.8722$	$\omega_1 = 0.1316$	0.0254	0.7457	0.0015	*Passed*
	$\mu_2 = -1.028$	$\sigma_2^2 = 2.9161$	$\omega_2 = 0.3169$				
	$\mu_3 = -5.641$	$\sigma_3^2 = 4.6428$	$\omega_3 = 0.2057$				
	$\mu_4 = -3.3965$	$\sigma_4^2 = 1.6073$	$\omega_4 = 0.1581$				
	$\mu_5 = 5.17$	$\sigma_5^2 = 52.89$	$\omega_5 = 0.1877$				
Gamma	$a = 49.9634$	$b = 1.0007$		0.0225	0.7647	0.1730	*Passed*
Weibull	$\lambda = 52.7591$	$k = 8.9187$		0.0201	0.7661	0.4865	*Passed*

Table 3.4: Parameters and error metrics for pdf estimates for Dataset II-LOS environment.

	Parameters			MRD	WMRD	D_{KL}	K-S
LN	$\mu = 0$	$\sigma^2 = 43.94$	$\omega = 1$	0.0137	0.3421	0.0350	*Passed*
LN-4 via DPM	$\mu_1 = 1.6741$	$\sigma_1^2 = 23.5$	$\omega_1 = 0.5350$	0.0154	0.3243	0.0090	*Passed*
	$\mu_2 = -1.3341$	$\sigma_2^2 = 50.2863$	$\omega_2 = 0.4516$				
	$\mu_3 = -20.5198$	$\sigma_3^2 = 19.0387$	$\omega_3 = 0.0132$				
	$\mu_4 = -45.9005$	$\sigma_4^2 = 18.3291$	$\omega_4 = 1.8379\text{x}10^{-4}$				
LN-2	$\mu_1 = -4.4899$	$\sigma_1^2 = 70.6211$	$\omega_1 = 0.2744$	0.0158	0.3276	0.0063	*Passed*
	$\mu_2 = 1.6982$	$\sigma_2^2 = 23.3400$	$\omega_2 = 0.7256$				
LN-3	$\mu_1 = -5.9958$	$\sigma_1^2 = 59.0546$	$\omega_1 = 0.2450$	0.0160	0.3236	0.0076	*Passed*
	$\mu_2 = 14.6231$	$\sigma_2^2 = 1.2766$	$\omega_2 = 0.0119$				
	$\mu_3 = 1.7433$	$\sigma_3^2 = 21.328$	$\omega_3 = 0.7431$				
LN-4	$\mu_1 = 4.1381$	$\sigma_1^2 = 10.5979$	$\omega_1 = 0.3316$	0.0159	0.3228	0.0023	*Passed*
	$\mu_2 = -8.1882$	$\sigma_2^2 = 64.5319$	$\omega_2 = 0.1387$				
	$\mu_3 = 14.0646$	$\sigma_3^2 = 2.1468$	$\omega_3 = 0.0205$				
	$\mu_4 = -1.0318$	$\sigma_4^2 = 23.2732$	$\omega_4 = 0.5092$				
LN-5	$\mu_1 = -8.6026$	$\sigma_1^2 = 56.3252$	$\omega_1 = 0.1398$	0.0160	0.3232	0.0021	*Passed*
	$\mu_2 = 14.0426$	$\sigma_2^2 = 2.2019$	$\omega_2 = 0.0213$				
	$\mu_3 = 4.1350$	$\sigma_3^2 = 10.9588$	$\omega_3 = 0.3706$				
	$\mu_4 = -1.3268$	$\sigma_4^2 = 20.9955$	$\omega_4 = 0.4681$				
	$\mu_5 = -47.4297$	$\sigma_5^2 = 1.0\text{x}10^{-6}$	$\omega_5 = 1.84\text{x}10^{-4}$				
Gamma	$a = 49.9634$	$b = 1.0007$		0.0038	0.3520	0.0733	*Passed*
Weibull	$\lambda = 52.7591$	$k = 8.9187$		0.0042	0.3368	0.0133	*Passed*

Table 3.5: Parameters and error metrics for pdf estimates for Dataset II-NLOS environment.

	Parameters			MRD	WMRD	D_{KL}	K-S
LN	$\mu = 0$	$\sigma^2 = 48.9412$	$\omega = 1$	0.0048	0.2708	0.0821	*Passed*
LN-5 via DPM	$\mu_1 = 0.8611$	$\sigma_1^2 = 17.5637$	$\omega_1 = 0.5935$	0.0045	0.1942	0.0333	*Passed*
	$\mu_2 = -1.2245$	$\sigma_2^2 = 91.1296$	$\omega_2 = 0.4051$				
	$\mu_3 = -33.0342$	$\sigma_3^2 = 6.2753$	$\omega_3 = 0.0005$				
	$\mu_4 = -1.6165$	$\sigma_4^2 = 0.6731$	$\omega_4 = 0.0003$				
	$\mu_5 = 5.4642$	$\sigma_5^2 = 0.7790$	$\omega_5 = 0.0004$				
LN-2	$\mu_1 = 0.4351$	$\sigma_1^2 = 23.2315$	$\omega_1 = 0.8031$	0.0050	0.2100	0.0296	*Passed*
	$\mu_2 = -1.7753$	$\sigma_2^2 = 149.9086$	$\omega_2 = 0.1969$				
LN-3	$\mu_1 = 1.7250$	$\sigma_1^2 = 17.1176$	$\omega_1 = 0.6499$	0.0050	0.1790	0.0075	*Passed*
	$\mu_2 = 25.2884$	$\sigma_2^2 = 17.9321$	$\omega_2 = 0.0168$				
	$\mu_3 = -4.6355$	$\sigma_3^2 = 53.0853$	$\omega_3 = 0.3333$				
LN-4	$\mu_1 = -6.3740$	$\sigma_1^2 = 89.1687$	$\omega_1 = 0.0997$	0.0050	0.1824	0.0058	*Passed*
	$\mu_2 = 25.3196$	$\sigma_2^2 = 17.9268$	$\omega_2 = 0.0166$				
	$\mu_3 = -2.3159$	$\sigma_3^2 = 31.2184$	$\omega_3 = 0.4400$				
	$\mu_4 = 2.7827$	$\sigma_4^2 = 12.5057$	$\omega_4 = 0.4438$				
LN-5	$\mu_1 = -4.4541$	$\sigma_1^2 = 133.8335$	$\omega_1 = 0.0576$	0.0050	0.1777	0.0057	*Passed*
	$\mu_2 = 2.4799$	$\sigma_2^2 = 13.9270$	$\omega_2 = 0.6391$				
	$\mu_3 = -4.3617$	$\sigma_3^2 = 13.1887$	$\omega_3 = 0.1995$				
	$\mu_4 = -9.6098$	$\sigma_4^2 = 26.3708$	$\omega_4 = 0.0886$				
	$\mu_5 = 25.6297$	$\sigma_5^2 = 16.5555$	$\omega_5 = 0.0153$				
Gamma	$a = 48.5313$	$b = 1.0303$		0.0026	0.3225	0.1067	*Passed*
Weibull	$\lambda = 52.9913$	$k = 6.8521$		0.0026	0.3509	0.1519	*Passed*

In this Chapter, dataset properties are given along with the evaluation metrics to measure difference between obtained and the actual pdfs. With the help of Dataset I and Dataset II measurements, it is shown that shadow fading can be modeled more accurately with mixture models according to mentioned evaluation metrics.

After proposed mixture model fits shadow fading more accurately, in the next Chapter, coverage area and outage probability expressions will be derived.

4. APPLICATION: THE EFFECT OF SHADOW FADING DISTRIBUTIONS ON OUTAGE PROBABILITY AND COVERAGE AREA

Shadow fading has a fundamental role in wireless communication network design. Coverage area, ($C(\phi)$), and outage probability, (p_{out}), performances of the networks are directly affected by the shadow fading characteristics. In this Chapter, $C(\phi)$ and p_{out} for the received power are investigated with the lognormal, the lognormal mixture, Gamma and Weibull shadow fading models by using two datasets.

4.1 Outage Probability and Coverage Area with Lognormal Shadowing

In order to determine the percentage of coverage area, we should calculate the probability that the received signal level will exceed or fall below a desired threshold value ϕ. The probability that the received power at distance d, falls below ϕ is defined as outage probability $p_{out} = p(P_{Rx}(d) \leq \phi)$.

When X is a random variable with a lognormal distribution (Gaussian distribution in dB), then the P_{Rx} will also be a normal random variable in dB. Under this model, the error function ($erf(x)$) can be used to determine the probability that the P_{Rx} will exceed or fall below ϕ. According to the [1, 3], probability that the P_{Rx} will exceed a desired threshold value of ϕ is defined as

$$p(P_{Rx} > \phi) = 1 - p_{out} = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\phi - P_{Rx}(d)}{\sqrt{2}\sigma}\right)\right], \tag{4.1}$$

By using equation (**4.1**), p_{out} can be written as

$$\begin{aligned} p_{out} &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{\phi - P_{Rx}(d)}{\sqrt{2}\sigma}\right)\right], \\ &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{\phi - [P_{Tx}(d) - 10n\log(d) + X_\sigma]}{\sqrt{2}\sigma}\right)\right]. \end{aligned} \tag{4.2}$$

According to the analysis of Jakes and Rappaport [1, 47], the probability of coverage area of a single cell with radius R can be determined as

$$C(\phi) = \frac{1}{\pi R^2}\int_0^{2\pi}\int_0^R p(P_{Rx}(r) > \phi) r \mathrm{d}r\mathrm{d}\theta. \tag{4.3}$$

by continuing with the calculations, $C(\phi)$ can be written as

$$C(\phi) = \frac{1}{2}\left[1 - \text{erf}(a) + \exp\left(\frac{1-2ab}{b^2}\right) \times \left(1 - \text{erf}\left(\frac{1-ab}{b}\right)\right)\right], \tag{4.4}$$

where

$$a = \frac{(\phi - P_{Rx}(R))}{\sigma\sqrt{2}} = \frac{(\phi - P_{Tx}(R) + 10n\log(R))}{\sigma\sqrt{2}},\ b = \frac{10n\log(e)}{\sigma\sqrt{2}}. \tag{4.5}$$

4.2 Outage Probability and Coverage Area with Lognormal Mixtures

Since we are analysing P_{Rx} in dB, then analysis can be done with Gaussian random variables and Gaussian mixture model. Assuming X is a Gaussian mixture, then by using equation (**4.1**), the probability that the P_{Rx} will exceed ϕ for Gaussian mixture shadow fading model can be written as

$$p[P_{Rx} > \phi] = \sum_{k=1}^{K} \frac{\omega_k}{2}\left[1 - \text{erf}\left(\frac{\phi - (P_{Rx}(d) - \mu_k)}{\sqrt{2}\sigma_k}\right)\right], \tag{4.6}$$

where K is the number of mixture component and ω_k is the weight of k^{th} component for $k = 1, \ldots, K$.

With the help of equation (**4.6**), p_{out} for Gaussian mixture shadow fading model can be defined as

$$p_{out} = \sum_{k=1}^{K} \frac{\omega_k}{2}\left[1 + \text{erf}\left(\frac{\phi - (P_{Rx}(d) - \mu_k)}{\sqrt{2}\sigma_k}\right)\right]. \tag{4.7}$$

In order to investigate coverage for mixture shadow fading, we note that $C(\phi)$ is a weighted superposition of the coverage probabilities of individual coverage probabilities. By using this information $C(\phi)$ for Gaussian mixture shadow fading model can be written as

$$C(\phi) = \sum_{k=1}^{K} \omega_k C_k(\phi), \tag{4.8}$$

for $k = 1, \ldots, K$. Note that by setting $k = 1$ and $\mu_1 = 0$ we obtain the same expressions as in [1, 47]. By using equation (**4.8**), $C(\phi)$ for Gaussian mixture shadow fading model can be defined as

$$C(\phi) = \sum_{k=1}^{K} \frac{\omega_k}{2}\left[1 - \text{erf}(a_k) + \exp\left(\frac{1-2a_kb_k}{b_k^2}\right) \times \left(1 - \text{erf}\left(\frac{1-a_kb_k}{b_k}\right)\right)\right], \tag{4.9}$$

where

$$a_k = \frac{(\phi - P_{Rx}(R))}{\sigma_k\sqrt{2}} = \frac{(\phi - P_{Tx}(R) + 10n\log(R))}{\sigma_k\sqrt{2}},\ b_k = \frac{10n\log(e)}{\sigma_k\sqrt{2}}, \tag{4.10}$$

for $k = 1,\ldots,K$.

4.3 Outage Probability and Coverage Area with Gamma Random Variable

Under the assumption of the X is a Gamma random variable in dB, then the P_{Rx} will also be a Gamma random variable. In order to calculate the probability that P_{Rx} will exceed and fall below ϕ for Gamma fading model, maximum likelihood estimation (MLE) of NLOS and LOS received powers for Gamma pdf, $f_\phi(t|a,b)$, can be estimated with a shape parameter and b scale parameter. By using pdf of Gamma random variable, p_{out} can be written as

$$\begin{aligned} p_{out} &= p(P_{Rx}(d) \leq \phi) = \int_{-\infty}^{\phi} f_\phi(t|a,b))\,\mathrm{d}t, \\ &= \frac{1}{b^a\Gamma(a)}\int_{-\infty}^{\phi} t^{a-1}e^{-\frac{t}{b}}\,\mathrm{d}t = \Gamma_{inc}\left(\frac{P_{Rx}}{b}, \frac{a}{\Gamma(a)}\right), \end{aligned} \tag{4.11}$$

where $\Gamma_{inc}(a,x) = \frac{1}{(a-1)!}\int_0^x t^{a-1}e^{-t}\,\mathrm{d}t$.

With the help of equation (**4.3**), the coverage area probability for Gamma fading model can be calculated by using following equation

$$C(\phi) = \frac{1}{\pi R^2}\int_0^{2\pi}\int_0^R \left[1 - \Gamma_{inc}\left(\frac{P_{Rx}}{b}, \frac{a}{\Gamma(a)}\right)\right] r\mathrm{d}r\mathrm{d}\theta. \tag{4.12}$$

4.4 Outage Probability and Coverage Area with Weibull Random Variable

If X is a Weibull random variable in dB, then the P_{Rx} will also be a Weibull random variable. To calculate p_{out} for Weibull fading model, MLE of NLOS and LOS received powers for Weibull pdf, $f_\phi(\phi|\lambda,k)$, can be estimated with k shape parameter and λ scale parameter. By using pdf of Weibull random variable, p_{out} can be written as

$$p_{out} \quad = p(P_{Rx}(d) \leq \phi) = 1 - e^{(-\phi/\lambda)^k}. \tag{4.13}$$

By using of equation (**4.13**), the coverage area probability for Weibull fading model can be calculated with following equation

$$C(\phi) = \frac{1}{\pi R^2} \int_0^{2\pi} \int_0^R e^{(-\phi/\lambda)^k} r \mathrm{d}r \mathrm{d}\theta. \tag{4.14}$$

In this chapter, coverage area and outage probability expressions are derived. By using the derived equations, in the next chapter, coverage area and outage probability figures will be exhibited for the proposed model and candidate models.

5. NUMERICAL RESULTS FOR OUTAGE AND COVERAGE ANALYSIS

In this Chapter, Gamma, Weibull, lognormal, and the proposed lognormal mixture distributions for shadow fading model are investigated to determine coverage area, $(C(\phi))$, and outage probability, (p_{out}), both empirically and theoretically with by using two real-life datasets. Empirical Dataset I includes only LOS environments, while empirical Dataset II include both LOS and NLOS environment. Since Gamma and Weibull distribution can not have a negative random variable, normalized P_{Rx} is shifted 50 dB for $C(\phi)$ and p_{out} analysis. Noting that number of mixture component for proposed model is randomly chosen.

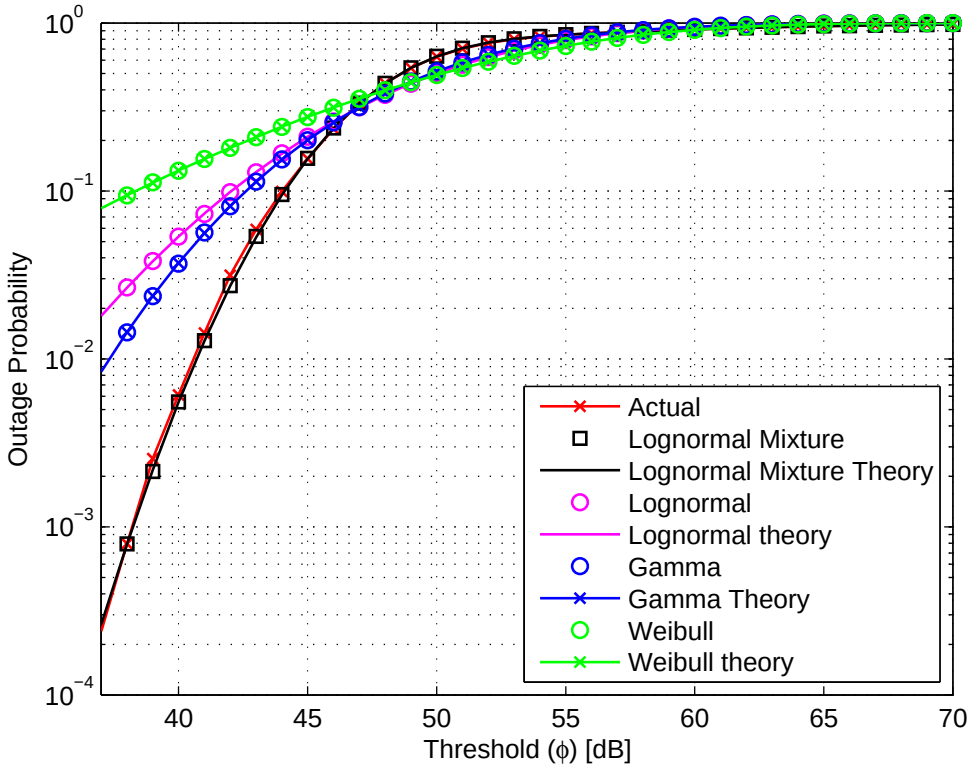

Figure 5.1: p_{out} of actual data and theoretical and empirical p_{out} for Lognormal Mixture (with 3 components), Lognormal, Gamma and Weibull shadow fading models for Dataset I.

Fig. 5.1 shows the p_{out} result for Dataset I. It is clearly seen that proposed mixture shadow fading model better fits than other models. For example, with the threshold

(ϕ)=40 dB, p_{out} is equal to 0.005528 for proposed model and 0.006106 is the actual value, while p_{out} values for other models are 0.05351 (lognormal), 0.03724 (Gamma), 0.1322 (Weibull).

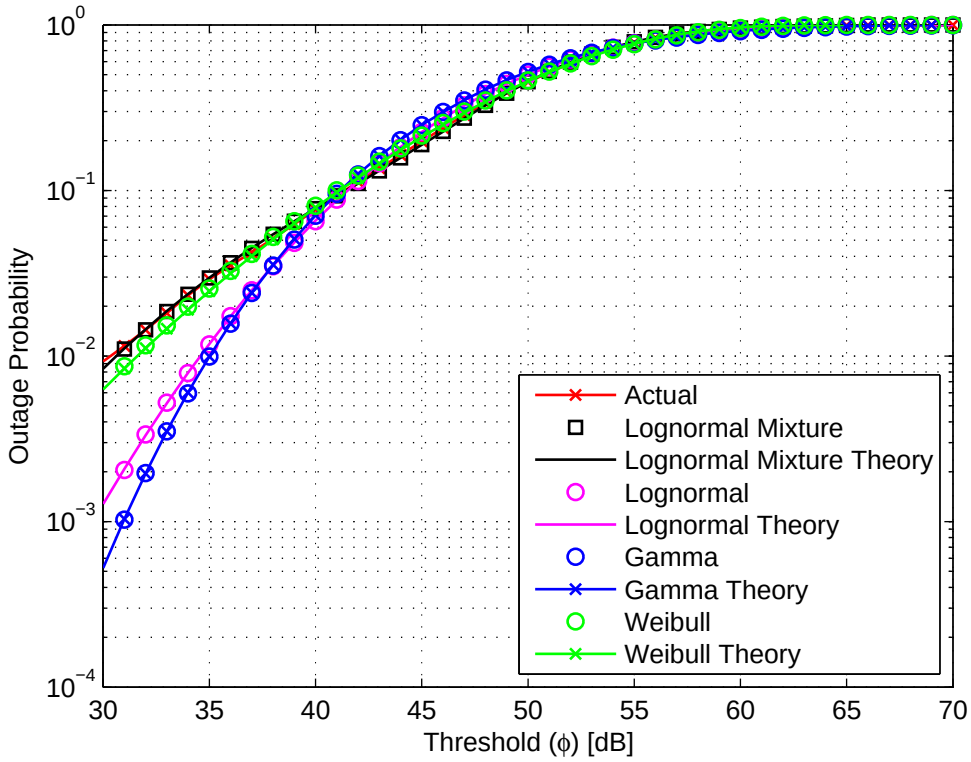

Figure 5.2: p_{out} of actual data and theoretical and empirical p_{out} for Lognormal Mixture (with 3 components), Lognormal, Gamma and Weibull shadow fading models in Dataset II-LOS environment.

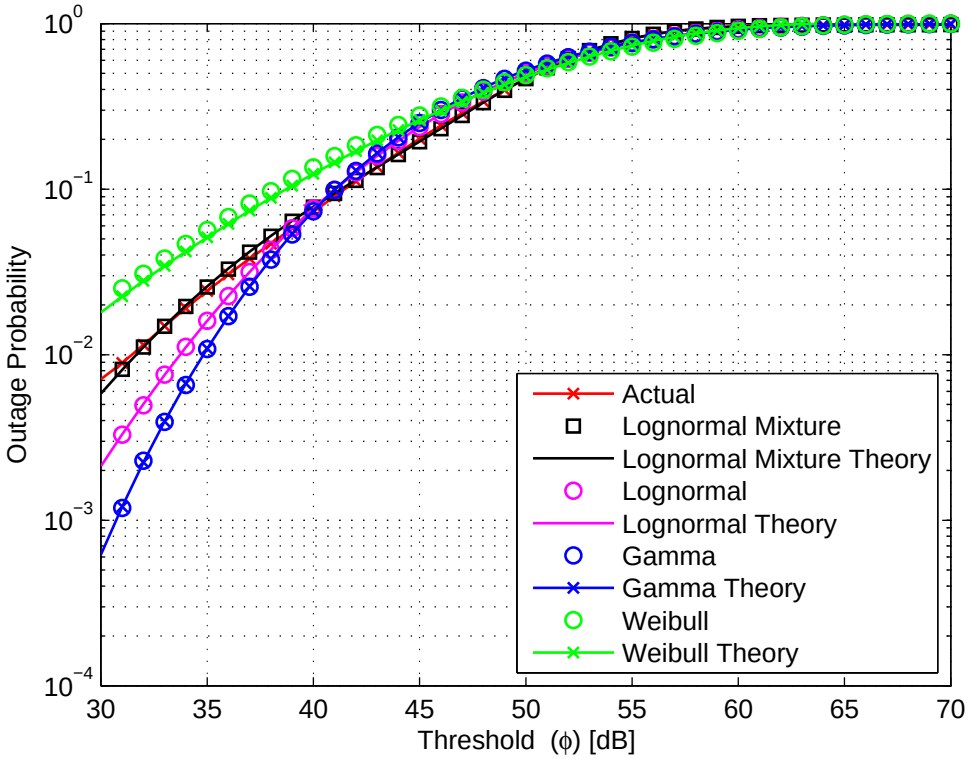

Figure 5.3: p_{out} of actual data and theoretical and empirical p_{out} for Lognormal Mixture (with 3 components), Lognormal, Gamma and Weibull shadow fading models in Dataset II-NLOS environment.

Fig. 5.2 and Fig. 5.3 show the p_{out} results for Dataset II-LOS and Dataset II-NLOS environment respectively. It can be easily seen that the proposed mixture shadow fading model with three lognormal random variable components is a better fit for p_{out} of actual data than lognormal, Gamma and Weibull models both theoretically and empirically.

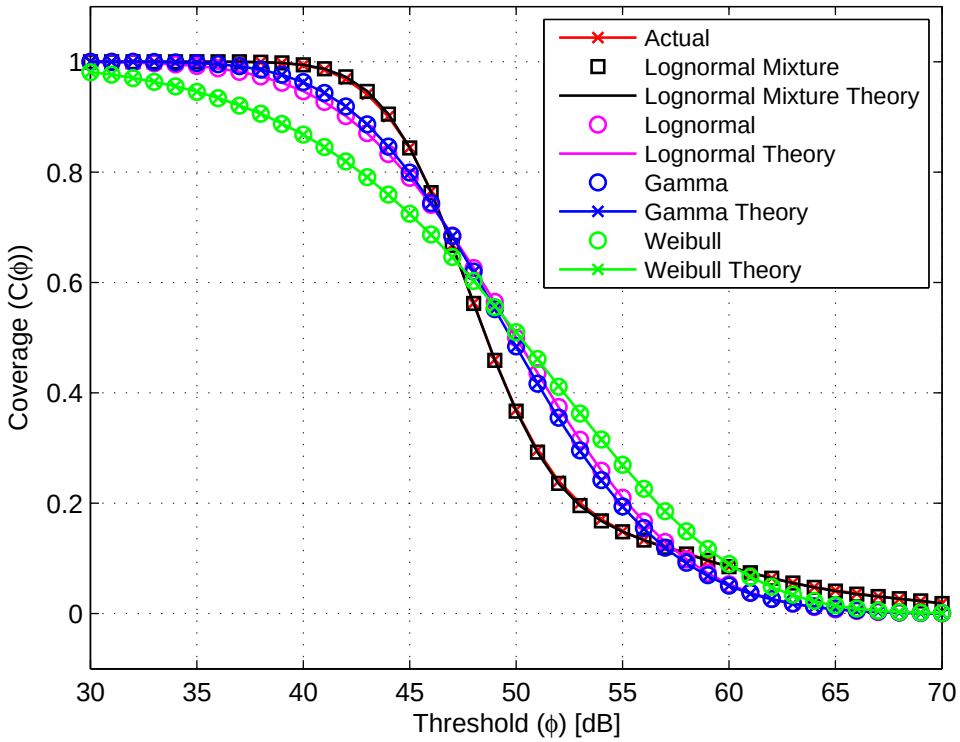

Figure 5.4: $C(\phi)$ of actual data and theoretical and empirical $C(\phi)$for Lognormal Mixture (with 3 components), Lognormal, Gamma and Weibull shadow fading models for Dataset I.

Fig. 5.4 shows the $C(\phi)$ result for Dataset I. With the threshold (ϕ)=45 dB, $C(\phi)$ is equal to 0.8445 for proposed model and 0.8444 is the actual value, while $C(\phi)$ values for other models are 0.7898 (lognormal), 0.7988 (Gamma), 0.7242 (Weibull). It is clearly seen that proposed mixture shadow fading model better fits than other models.

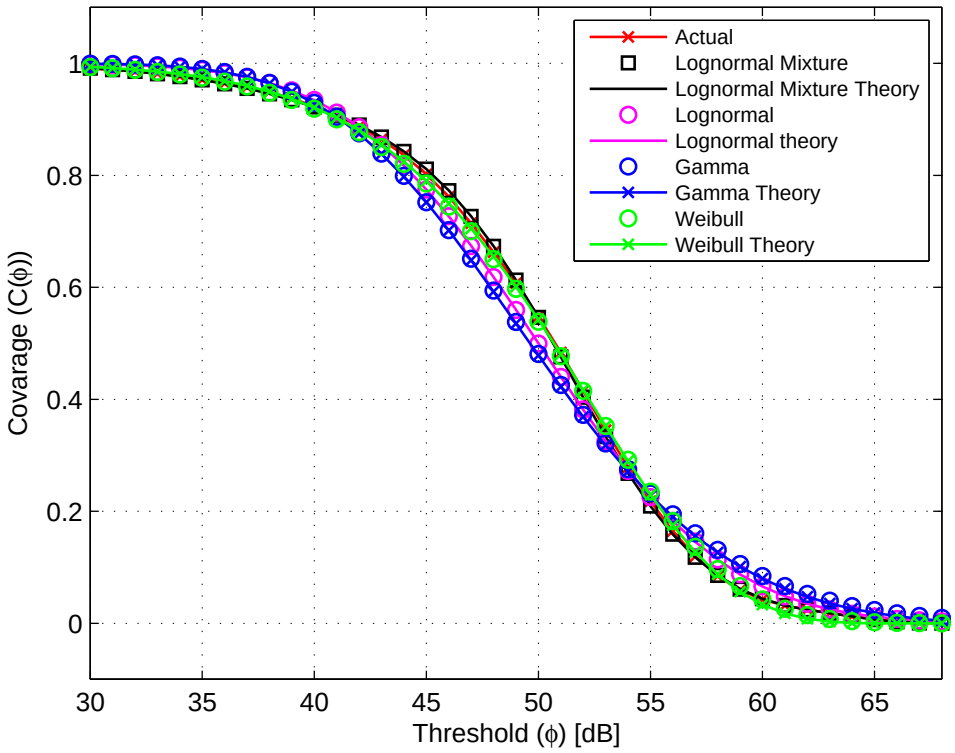

Figure 5.5: $C(\phi)$ of actual data and theoretical and empirical $C(\phi)$ for Lognormal Mixture (with 3 components), Lognormal, Gamma and Weibull shadow fading models in Dataset II-LOS environment.

Fig. 5.5 and Fig. 5.6 show the $C(\phi)$ results for Dataset II-LOS and Dataset II-NLOS environment.

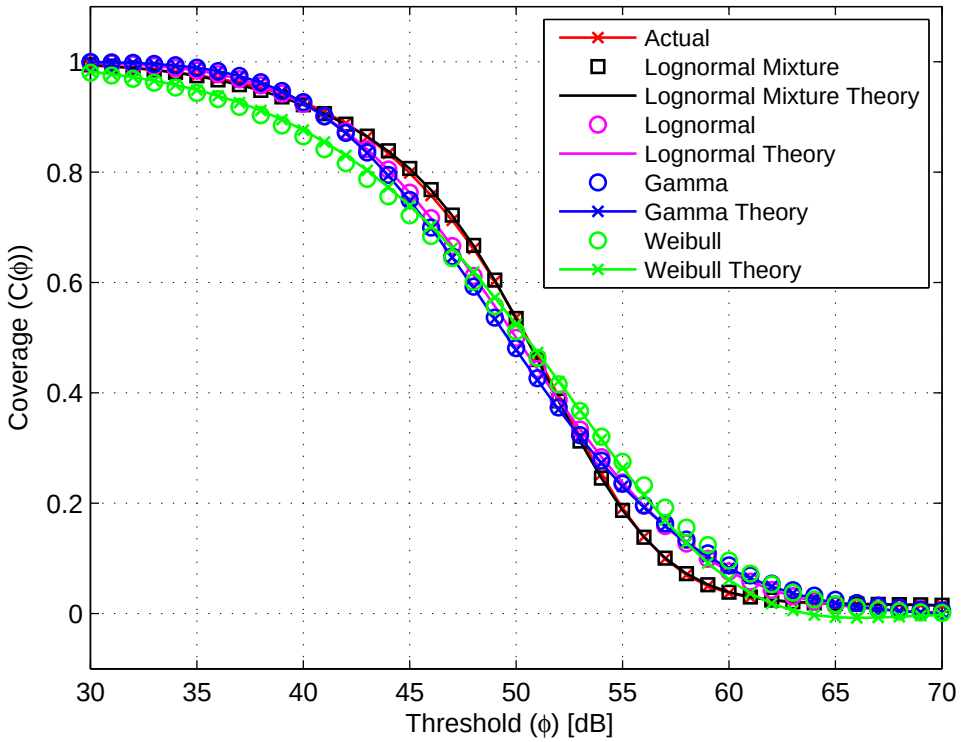

Figure 5.6: $C(\phi)$ of actual data and theoretical and empirical $C(\phi)$ for Lognormal Mixture (with 3 components), Lognormal, Gamma and Weibull shadow fading models in Dataset II-NLOS environment.

According to these figures, the proposed mixture shadow fading model with three lognormal random variable components better fit for $C(\phi)$ of actual data than other mentioned models above. As it can be easily seen that lognormal mixture shadow fading model gives more accurate p_{out} and $C(\phi)$ results than other models.

In this Chapter, as an application to show that better fitting models can provide more accurate (realistic) estimates in terms of the coverage and outage probability, results for lognormal, lognormal mixture, Gamma and Weibull shadow fading models are demonstrated.

6. CONCLUSIONS

In this thesis a lognormal mixture model for shadow fading based on cluster concepts is proposed to address the inaccuracy of single lognormal and other shadow fading models. The flexibility and the applicability of the proposed model is demonstrated by proving that an arbitrary probability density function can be modeled as lognormal mixtures by using positive definite kernels.

A nonparametric approach based on DPM models and a more practical parametric approach based on EM, where it is assumed that the number of components are known a priori are provided to find the parameters of the lognormal mixture components.

Two real-life urban macrocell empirical datasets are investigated to compare the proposed lognormal mixture shadow fading model and other candidate shadow fading models. The accuracy of the proposed lognormal mixture shadow fading model is shown that the results obtained from both goodness-of-fit tests, error vector norm based techniques (WMRD, MRD) and KL divergence. According to numerical results, proposed mixture modeling has better WMRD, MRD and KL divergence values than lognormal, Gamma and Weibull models. With the help of two datasets measurements, it is shown that shadow fading can be modeled more accurately with mixture models according to mentioned evaluation metrics.

The results of mixture modeling are extended to provide outage and cellular coverage probabilities. By using two real-life datasets, outage and cellular coverage probabilities for the received powers are investigated with the proposed lognormal mixture model and lognormal, Gamma, Weibull shadow fading models. Mathematical expressions for the proposed lognormal mixture shadow fading model and all considered shadow fading models are derived. According to numerical results, simulations and theoretical expressions are consistent and it is verified that more accurate shadow fading models lead to more realistic analysis outcomes.

Due to the fact that shadow fading has a fundamental role in wireless communication network design and performance metrics of the networks are directly effected by the shadow fading characteristics, other performance metrics should be investigated as a future work to verify that more accurate shadow fading models lead to more realistic analysis outcomes.

REFERENCES

[1] **Rappaport, T.**, 2001. Wireless Communications: Principles and Practice, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition.

[2] **Parsons, J.D.**, 2000. The Mobile Radio Propagation Channel, 2nd Edition, Wiley, 2 edition.

[3] **Goldsmith, A.**, 2005. Wireless Communications, Cambridge University Press, New York, NY, USA.

[4] **Marvin K. Simon, M.S.A.**, 2005. Digital Communication over Fading Channels, Wiley, 2nd edition.

[5] **Egli, J.**, 1957. Radio Propagation above 40 MC over Irregular Terrain, *Proc. of the IRE*, **45(10)**, 1383–1391.

[6] **Salo, J.**, **Vuokko, L.**, **El-Sallabi, H. and Vainikainen, P.**, 2007. An Additive Model as a Physical Basis for Shadow Fading, *Vehicular Tech., IEEE Trans. on*, **56(1)**, 13–26.

[7] **Suikkanen, E.**, **Hentila, L. and Meinila, J.**, 2010. Wideband radio channel measurements around 800 MHz in outdoor to indoor and urban macro scenarios, Future Netw. and Mobile Summit, 2010, pp.1–9.

[8] **Boettcher, A.**, **Schneider, C.**, **Narandzic, M.**, **Vary, P. and Thomae, R.**, 2010. Power and delay domain parameters of channel measurements at 2.53 GHz in an urban macro cell scenario, Antennas and Propagation (EuCAP), 2010 Proc. of the Fourth European Conf. on, pp.1–5.

[9] **Karedal, J.**, **Johansson, A.**, **Tufvesson, F. and Molisch, A.**, 2008. A Measurement-Based Fading Model for Wireless Personal Area Networks, *Wireless Comm., IEEE Trans. on*, **7(11)**, 4575–4585.

[10] **Jaldén, N.**, **Zetterberg, P.**, **Ottersten, B. and Garcia, L.**, 2007. Inter-and Intrasite Correlations of Large-scale Parameters from Macrocellular Measurements at 1800 MHz, *EURASIP J. Wirel. Commun. Netw.*, **2007(3)**, 10:1–10:12, `http://dx.doi.org/10.1155/2007/25757`.

[11] **Miller, J. and Miller, S.**, 1997. Smart antenna adaptive performance in the presence of imperfect power control, multipath and shadow fading, Global Telecomm. Conf., 1997. GLOBECOM '97., IEEE, volume 1, pp.384–388 vol.1.

[12] **Weitzen, J. and Lowe, T.**, 2002. Measurement of angular and distance correlation properties of log-normal shadowing at 1900 MHz and its application to design of PCS systems, *Vehicular Tech., IEEE Trans. on*, **51(2)**, 265–273.

[13] **Vuokko, L.**, **Vainikainen, P. and Takada, J.i.**, 2005. Clusters extracted from measured propagation channels in macrocellular environments, *Antennas and Propagation, IEEE Trans. on*, **53(12)**, 4089–4098.

[14] **Kalliola, K.**, **Laitinen, H.**, **Vainikainen, P.**, **Toeltsch, M.**, **Laurila, J. and Bonek, E.**, 2003. 3-D double-directional radio channel characterization for urban macrocellular applications, *Antennas and Propagation, IEEE Trans. on*, **51(11)**, 3122–3133.

[15] **Coulson, A.**, **Williamson, A. and Vaughan, R.**, 1998. A statistical basis for lognormal shadowing effects in multipath fading channels, *Comm., IEEE Trans. on*, **46(4)**, 494–502.

[16] **Jaldén, N.**, **Zetterberg, P.**, **Ottersten, B. and Garcia, L.**, 2007. Inter-and intrasite correlations of large-scale parameters from macrocellular measurements at 1800 MHz, *EURASIP J. Wirel. Commun. Netw.*, **2007(3)**, 10:1–10:12.

[17] **Jalden, N.**, **Zetterberg, P. and Ottersten, B.**, 2008. Directional Dependence of Large Scale Parameters in Wireless Channel Models, Wireless Comm. and Networking Conf., 2008. WCNC 2008. IEEE, pp.1223–1228.

[18] **Abdi, A. and Kaveh, M.**, 1999. On the utility of gamma PDF in modeling shadow fading (slow fading), Vehicular Tech. Conf., 1999 IEEE 49th, volume 3, pp.2308–2312 vol.3.

[19] **Ismail, M. and Matalgah, M.M.**, 2005. Outage probability analysis in cellular systems with noisy Weibull-faded lognormal-shadowed links, Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on, pp.269–274.

[20] **Atapattu, S.**, **Tellambura, C. and Jiang, H.**, 2011. A Mixture Gamma Distribution to Model the SNR of Wireless Channels, *Wireless Comm., IEEE Trans. on*, **10(12)**, 4193–4203.

[21] **J. Salo, L.V. and Vainikainen, P.**, 2005. Why is shadow fading lognormal?, Proc. of the Int. Symposium on Wireless Personal Multimedia Comm., Aalborg, Denmark, p.522–526.

[22] **Teh, Y.W.**, 2010. Dirichlet Process, Encyclopedia of Machine Learning, pp.280–287.

[23] **Escobar, M.D. and West, M.**, 1995. Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, **90**, 577–588.

[24] **Rasmussen, C.**, 2000. The Infinite Gaussian Mixture Model, **T.L.K.R.M. Solla S.A.**, editor, Advances in Neural Information Processing Systems 12, Max-Planck-Gesellschaft, MIT Press, Cambridge, MA, USA, pp.554–560.

[25] **Lo, A.Y.**, 1984. On a Class of Bayesian Nonparametric Estimates: I. Density Estimates, *The Annals of Statistics*, **12(1)**, 351–357.

[26] **Buyukcorak, S.**, **Kurt, G. and Cengaver, O.**, 2013, A Probabilistic Framework for Estimating Call Holding Time Distributions.

[27] **Schneider, C.** *et al.*, 2009. Multi-User MIMO Channel Reference Data for Channel Modelling and System Evaluation from Measurements, *in Int. ITG Workshop on Smart Antennas*.

[28] **Bilmes, J.**, 1997, A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.

[29] **McLachlan, G.J. and Krishnanl, T.**, 2008. The EM Algorithm and Extensions, Wiley Series in Probability and Statistics.

[30] **Cuturi, M.**, 2010. Positive Definite Kernels in Machine Learning, `arXiv/0911.5367`.

[31] **Hofmann, T.**, **Schölkopf, B. and Smola, A.J.**, 2008. Kernel methods in machine learning, *Annals of Statistics*, **36(3)**, 1171–1220.

[32] **Bertinet, A. and Agnan, T.C.**, 2004. Reproducing Kernel Hilbert Spaces in Probability and Statistics, Kluwer Academic Publishers.

[33] **Garcia, A.**, 1994. Probability and Random Processes for Electrical Engineering, Addison-Wesley.

[34] **Fasshauer, G.E.**, 2011. Positive definite kernels: past, present and future, *Dolomites Research Notes on Approximation*, **4(Special Issue on Kernel Functions and Meshless Methods)**, 21–63, `http://drna.di.univr.it/volume04.html`.

[35] **Mehta, N.**, **Wu, J.**, **Molisch, A. and Zhang, J.**, 2007. Approximating a Sum of Random Variables with a Lognormal, *Wireless Comm., IEEE Trans. on*, **6(7)**, 2690–2699.

[36] **Fearnhead, P.**, 2004. Particle filters for mixture models with an unknown number of components, *Statistics and Computing*, **14(1)**, 11–21.

[37] **Aldous, D.**, 1985. Exchangeability and related topics, Ecole d'Ete de Probabilities de Saint-Flour XIII 1983, Springer, pp.1–198.

[38] **Kumar, A.**, **Sung, M.**, **Xu, J. and Wang, J.**, 2004. Data streaming algorithms for efficient and accurate estimation of flow size distribution, Joint Int. Conf. on Measurement and Modeling of Computer Systems, SIGMETRICS/Performance.

[39] **Bilmes, J.**, 1997. A Gentle Tutorial of the EM algorithm and its application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical ReportTR-97-021, ICSI.

[40] **Wei, H.**, **Zhong, Z.**, **Xiong, L.**, **Ai, B. and He, R.**, 2011. Study on the shadow fading characteristic in viaduct scenario of the High-speed Railway, Comm. and Networking in China (CHINACOM), 2011 6th Int. ICST Conf. on, pp.1216–1220.

[41] **Algans, A.**, **Pedersen, K. and Mogensen, P.**, 2002. Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading, *Selected Areas in Comm., IEEE Journal on*, **20(3)**, 523–531.

[42] **Pätzold, M.**, **Avazov, N. and Nguyen, V.**, 2010. Design of measurement-based correlation models for shadow fading, Advanced Technologies for Comm. (ATC), 2010 Int. Conf. on, pp.112–117.

[43] **Glazunov, A. and Wang, Y.**, 2005. Decorrelation Distance Characterization of Long Term Fading of CW MIMO Channels in Urban Multicell Environment, Applied Electromagnetics and Comm., 2005. ICECom 2005. 18th Int. Conf. on, pp.1–4.

[44] **Gudmundson, M.**, 1991. Correlation model for shadow fading in mobile radio systems, *Electronics Letters*, **27(23)**, 2145–2146.

[45] **Bottcher, A.**, **Vary, P.**, **Schneider, C. and Thoma, R.**, 2012. De-correlation distance of the large scale parameters in an urban macro cell scenario, Antennas and Propagation (EUCAP), 2012 6th European Conf. on, pp.1417–1421.

[46] **Duffield, N.**, **Lund, C. and Thorup, M.** Estimating flow distributions from sampled flow statistics, *IEEE/ACM Trans. Netw.*, **13(5)**.

[47] **Jakes, W.C.**, 1974. Microwave mobile communications, Wiley New York.

[48] **Blackwell, D. and MacQueen, J.**, 1973. Ferguson distributions via Polya urn schemes, *Ann. Statist.*, **1**, 353–355.

[49] **Ulker, Y.**, **Gunsel, B. and Cemgil, A.**, 2011. Annealed SMC Samplers for Nonparametric Bayesian Mixture Models, *IEEE Signal Proc. Letters*, **18(1)**, 3–6.

[50] **Gershman, S.J. and Blei, D.**, 2012. A tutorial on Bayesian nonparametric models, *Journal of Mathematical Psychology*, **56(1)**, 1–12.

[51] **Kobayashi, H.**, **Mark, B. and Turin, W.**, 2011. Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance, Probability, Random Processes, and Statistical Analysis, Cambridge University Press.

[52] **Maceachern, S.N.**, 1994. Estimating normal means with a conjugate style dirichlet process prior, *Comm. In Statistics - Simulation and Computation*, **23(3)**, 727–741.

[53] **Liang, F.**, **Liu, C. and Carrol, R.**, 2010. Advanced Markov Chain Monte Carlo Methods, Learning From Past Samples, Wiley.

[54] **Casella, G. and George, E.I.**, 1992. Explaining the Gibbs Sampler, *The American Statistician*, **46(3)**, 167–174.

APPENDICES

APPENDIX A.1 : Dirichlet Process Mixture Model for Lognormal Mixtures

APPENDIX A.1

Dirichlet Process Mixture Model for Lognormal Mixtures

DPM models are useful tools for analysing mixture models with an unknown number of mixture components [36]. They can be applied for decomposing $S_f(y)$ into lognormal mixture components.

To be able to use DPM models, the parameters specifying the observation model are RVs and the observations need to be exchangeable. Although RVs are dependent among themselves, exchangeability demonstrates the assumption that RVs are independent of their observation orders [22, 37]. Noting that any P_{Rx} measurement set is exchangeable for the same d, so we can use DPM to determine mixture components.

Within the scope of our study, we use the Polya-Urn formulation [48] introduced by Blackwell and MacQueen for determining the mixture components in the DPM model [26, 36]. Taking this model into account, each observation of $y_{1:N}$ belongs to a mixture component. For any given mixture, all observations of a particular mixture are independent draws from the same distribution. Define $c_{1:N} = (c_1, \ldots, c_N)$ as the vector of mixture labels, K as the number of distinct mixtures in the assignment, where $c_i \in \{1, \ldots, K\}$ for $i = 1, \ldots, N$. Each mixture includes at least one observation. If we assume that each observation is a member of a cluster of lognormal distributions, $\theta_k = (\mu_k, \sigma_k^2)$, where μ_k represents the mean and σ_k^2 is the variance, for $k = 1, \ldots, K$. Let $S_f(y|\theta_k)$ denote the density of $y_{1:N}$ for a given value of $\theta_{1:K} = (\theta_1, \ldots, \theta_K)$, where $\theta_i = \theta_k$, only when $i = k$.

Under the assumption that mixture parameters are *iid*, the probability distribution for y_i has a hierarchical form [49]. First of all, we have a Dirichlet based prior for the mixture label c_i, $p(c_i)$ of k^{th} mixture. Then, conditional on having K mixtures under our assignment, we have K distinct θ_k, which are independently drawn from a known prior $p(\theta_k)$. Finally, conditional on c_i and θ_k, observation y_i is drawn independently from $S_f(y_i|\theta_{c_i})$. Hence, the joint distribution can be written as [36, 49]

$$S_f(y_i, c_i, \theta_k) = p(c_i) \prod_{k=1}^{K} p(\theta_k) \prod_{i=1}^{N} f_Y(y_i|\theta_{c_i}). \tag{A.1}$$

Bayesian nonparametric models aim to specify the number of mixture components with the assumption of $K \to \infty$ (i.e. possibly infinite number of components), specifying the prior over all likely distinct labelings. The prior over these labelings is originated from the Chinese restaurant process (CRP) [37]. The CRP provides a distribution over all possible labelings of the data [50], accordingly it can be used to specify the number of mixtures in the lognormal mixture model. The prior for the vector of mixture label c_i is parameterized by $\alpha > 0$, and can be calculated iteratively by using

$$p(c_{i+1} = k|c_i) = \begin{cases} N_k/(i+\alpha) & \text{for } k = 1, \ldots, k_i \\ \alpha/(i+\alpha) & \text{for } k = k_i + 1 \end{cases}, \tag{A.2}$$

where k_i is the number of mixtures in the assignment c_i, and N_k is the number of observations of mixture k [50]. Hereby with CRP, we reach a distribution over all possible distinct labelings amongst mixtures and their posterior probabilities. CRP can be used to determine shadow fading distributions of y_i with the highest posterior probability, so that we can model the experimental data in the MAP sense in an optimal manner [51].

In order to deal with the problem of random sampling from a collection of conditional distributions, the prior distribution in equation (**A.2**) is often difficult to work out analytically, necessitating the development of Monte Carlo procedures [29,52,53]. We use the Gibbs sampler to obtain the the optimum MAP labeling. This is a technique for indirectly generating RVs from a marginal distribution without having to calculate the density explicitly [54], and it can be used to implement a practically realizable framework to estimate mixture distributions for received powers.

By using the relation between Gaussian and lognormal RVs given in Proposition 1, each observation is drawn from a univariate Gaussian distribution with unknown mean and variance, $\theta_k = (\eta_k, \nu_k^2)$. The variance ν_k^2 is assumed an inverse Gamma prior. With $s_k = 1/\nu_k^2$, it can be written as [36]

$$p(s_k) = \frac{s_k^{\varsigma-1}\rho^{\varsigma}\exp(-\rho s_k)}{\Gamma(\varsigma)}, \tag{A.3}$$

where ς is the shape parameter, ρ is the scale parameter, which are known and $\Gamma(\varsigma)$ is the Gamma function. Conditional on s_k, η_k has a normal prior distribution with mean ψ and variance τ/s_k. The mixture parameters are *iid* with

$$\eta_k \sim \mathcal{N}(\psi, \tau/s_k), \tag{A.4}$$

$$\nu_k^2 \sim InvGamma(\varsigma, \rho), \tag{A.5}$$

where $InvGamma(\cdot)$ denotes the inverse Gamma function [49].

In order to obtain distribution parameters of mixture components, we can follow the formulation of Gaussian mixture model (based on Proposition 1). Conditional on vector of mixture labels c_i, the posterior distribution of the k^{th} mixture's parameter (i.e. s_k and η_k) depend on N_k and $\bar{y}_k$, $\hat{v}_k^2$ the mean and the variance of these observations. Hence, we obtain that $S_f(s_k|N_k, \bar{y}_k, \hat{v}_k^2)$ is Gamma density with shape parameter $\varsigma + N_k/2$ and scale parameter $\rho + \frac{N_k}{2}\left(\hat{v}_k^2 + \frac{(\bar{y}_k-\psi)^2}{1+N_k\tau}\right)$ [36]. As a result, the unknown variance parameter for each mixture in (**2.4**) is estimated by using the following equation

$$\nu_k^2 = \frac{1}{s_k} = \left[\left(\varsigma + \frac{N_k}{2}\right)^{-1}\left(\rho + \frac{N_k}{2}\left(\hat{v}_k^2 + \frac{(\bar{y}_k - \psi)^2}{1+N_k\tau}\right)\right)\right]. \tag{A.6}$$

The $S_f(\eta_k|N_k, \bar{y}_k, \hat{v}_k^2, s_k)$, represented the conditional mean, can be written as a Gaussian distribution

$$S_f(\eta_k|N_k, \bar{y}_k, \hat{v}_k^2, s_k) \sim \mathcal{N}\left(\frac{\psi + N_k\tau\bar{y}_k}{1+N_k\tau}, \frac{\tau}{(s_k + s_k N_k\tau)}\right). \tag{A.7}$$

where the first and the second terms are the mean and the variance of $S_f(\cdot)$, respectively. The mean and variance of unknown mean parameter in each mixture can be estimated by using (**A.7**). Weights of each mixture can be calculated by $\pi_k = \frac{N_k}{N}$.

CURRICULUM VITAE

Name Surname: Metin Vural

Place and Date of Birth: Narman/ERZURUM 21.01.1988

Adress: İzzetpasaçiftliği Mah. Gürun Cad. Otluk Sok. No:9 Daire:4, 34387, Şişli/Istanbul, Turkey

E-Mail: vuralmtn@gmail.com, vuralmeti@itu.edu.tr

B.Sc.: Istanbul Technical University, Department of Electronics and Communication Engineering, Electronics Engineering Programme

List of Publications and Patents:

- **Metin Vural**, Pamir Erdem, and Onur Ağın, "Regression Clustering With Lower Error via EM Algorithm," *IEEE Conference on Signal Processing, Communications, and their Applications (SIU 2014)*, (accepted), Trabzon, April 2014.

- Saliha Büyükçorak, **Metin Vural**, Güneş Karabulut Kurt, "Lognormal Mixture Shadowing," *IEEE Transactions on Vehicular Technology*, (submitted on January 2014).

- **Metin Vural**, and Güneş Karabulut Kurt, "Effect of cooperative communications on power consumption in smart grid," *IEEE Conference on Signal Processing, Communications, and their Applications (SIU 2013)*, Cyprus, April 2013.

- Zeynep Özdemir, **Metin Vural**, Özge Cepheli, and İbrahim Hökelek, "Performance Analysis of Wireless Sensor Networks by Using NS-2," *2013 European Conference of Technology and Society (EuroTecS-2013)*, Sarajevo, Bosnia, June 2013.

- N. Uçar, G. Karabulut, B. Yağcı, B. Görgün, O. Ceylan, **M. Vural**, B. Badur, B .Başkan, S.Cakın, C. Bozkurt, E. G. Akay, "Monitoring The Vital Signals: Smart Garment For Temperature Sensing Sensor Networks," *5th International Technical Textiles Congress*, İzmir, November 2012.

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- Saliha Büyükçorak, **Metin Vural**, Güneş Karabulut Kurt, “Lognormal Mixture Shadowing," *IEEE Transactions on Vehicular Technology*, (submitted on January 2014).