DENSITY ESTIMATION METHODS IN NON-LINEAR RECEIVERS FOR INTERFERENCE MITIGATION

A THESIS

submitted by

BANDE MEGHANA
(EE09B008)

in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
JUNE 2013

THESIS CERTIFICATE

This is to certify that the thesis titled **DENSITY ESTIMATION METHODS IN NON-LINEAR RECEIVERS FOR INTERFERENCE MITIGATION**, submitted by **Bande Meghana (EE09B008)**, to the Indian Institute of Technology Madras, Chennai for the award of the degree of **Bachelor of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. K. Giridhar
Research Guide
Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 13th June 2013

ACKNOWLEDGEMENTS

I take this opportunity to thank all the people who have given me constant support and encouragement at all times. I am grateful to Dr. Giridhar for his continual support and invaluable advice throughout the project. I would also like to both the Vaishnavis for their help and guidance.

I am very grateful for being blessed with the most wonderful set of friends and classmates at IIT Madras. I am thankful to Madhavi, Maneesha, Himaja, Sunaina, Sabah, Sohini, Pranitha, Kolan, Chitra, Karthik, Sarak, Malla, Abhinav for all the fun we had together. I am thankful to Ganesh and Sujan for being there for me and pushing me to work on my project. I would like to thank all the people who made my stay at IIT Madras special.

Most importantly, I would like to thank my parents and my brother for their never-ending support and understanding at each and every point of my life.

ABSTRACT

KEYWORDS: Interference Mitigation, Kernel Density Estimate, Gaussian Mixture, Expectation-Maximization, OFDM systems

Wireless communication systems employ universal frequency reuse to accommodate more users per area and hence co-channel interference and its mitigation have gained importance. Though Linear Minimum Mean Square Error (LMMSE) is commonly preferred, they may not be the optimal choice when the noise is not Gaussian. This is the case when the interference is high and the interference plus noise becomes non-Gaussian. We want to look at the performance of receivers that model the interference plus noise as non-Gaussian. The interference plus noise can be modeled as a probability density function that can be estimated using various machine learning techniques.

In this thesis, we use a parametric Gaussian Mixture Model (GMM) and a non-parametric Kernel Density Estimation (KDE) method to estimate the probability distribution function (pdf) of interference plus noise over a frequency selective channel OFDM channel without considering any temporal variation. We study non-linear receivers that use ML detection considering GMM and KDE and try to compare their performance.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

OFDM	Orthogonal Frequency Division Multiplexing
SNR	Signal to Noise Ratio
SINR	Signal to Interference Noise Ratio
MRC	Maximum Ratio Combining
LMMSE	Linear Minimum Mean Square Error
ML	Maximum Likelihood
MSE	Mean Square Error
CSI	Channel State Information
CCI	Co-channel Interference
KDE	Kernel Density Estimate
GMM	Gaussian Mixture Model
EM	Expectation Maximization
ISI	Inter Symbol interference
BER	Bit Error Rate
QPSK	Quadrature Phase Shift Keying

NOTATIONS

$CN(\mu,\sum)$	Complex normal
x^T	Transpose of x
$X_i[k]$	Element corresponding to kth frequency of ith user
$\|x\|$	Absolute value of x
$E[.]$	Expectation operator
$K(.)$	Kernel function
$argmax(.)$	The argument at which expression becomes maximum
$(d/dx)(.)$	Derivative with respect to x

CHAPTER 1
INTRODUCTION

Frequency reuse is deployed in wireless systems in order to increase the number of users that can be accommodated per square area at the same time. Multiple links using common resources are active simultaneously because of the demand for higher data rate made by the increased number of users. Therefore co-channel interference comes into the picture. Increasing the transmit power does not help in fighting inter-cell interference. Hence signal to interference noise (SINR) becomes the bottleneck in achieving good throughput performance or lower bit error rate (BER). So the need for receiver architecture that is capable of interference mitigation arises. Good interference mitigation at the receiver can improve quality of each link in the network thereby leading to a significant improvement in the overall system level performance.

1.1 Interference Mitigation:

In this thesis, we are mainly concerned with co-channel interference (CCI). Interference cancellation or mitigation techniques can be employed either at the transmitter or the receiver. This thesis is concerned with only receiver based techniques.

1.1.1 Receiver based Interference mitigation:

Some of the traditionally used receiver techniques in presence of interference are as follows

1. **Maximal Ratio Combining (MRC):**

   It is a diversity combining technique in which the signal at the receive antenna is weighed by a factor that is proportional to amplitude of the desired signal. It can be thought of as being equivalent to a spatial matched filtering at each antenna and then summing up all the signals. The received signals coherently add up to give the maximum output Signal-to-Noise Ratio (SNR). This combining technique is not very effective in presence of interference and is not an interference cancellation technique.

2. **Zero Forcing (ZF) combining:**
   In this method, the interference is cancelled by projecting the received signal into the subspace orthogonal to the interferers. It might lead to loss of amplitude of desired signal if the spatial signature is not orthogonal to interferer and is optimal at high SNR.
3. **Linear Mean Square Error Interference cancellation (LMMSE-IC):**
   The signals are weighted such that the Mean Square Error (MSE) between estimate and the actual symbol is minimized. It essentially does matched filtering at low SNR and Zero forcing at high SNR achieving the tradeoff between eliminating interference and maximizing desired signal power.

Vaishnavi(2012) showed that in presence of non-gaussian noise, LMBER methods outperform LMMSE and also that LMBER techniques do not work well in presence of heterogenous interferers. Performing ML detection on the conditional pdf gives optimal error rate performance compared to using linear receivers.

1.2 System Model:

In our model, we consider multiple transmitters and a single receiver antenna. Let the transmitters be $T_1, \ldots, T_n$ where the desired transmitter is T_1 and the remaining transmitters are interferers and R be the receiver antenna.

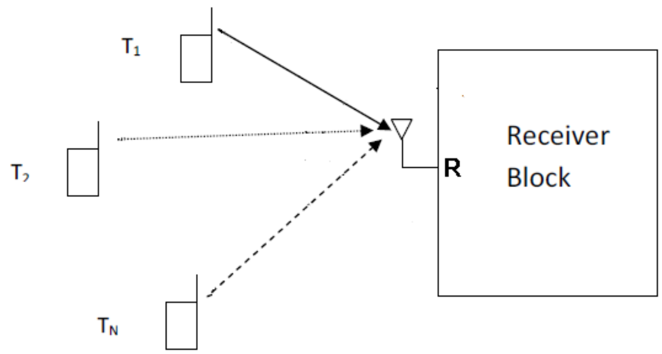


Figure 1.1: System Model

The communication system used is a 10MHz Orthogonal Frequency Division Multiplexing (OFDM) multi-carrier system. 1024 point IFFT/ FFT is used for OFDM modulation/demodulation. There are 1024 subcarriers out of which 212 in the beginning, 513^{th} (dc sub carrier) and the last 211 subcarriers are nulled. No transmissions in these subcarriers as these are guard bands. Data is transmitted only in the remaining 600 subcarriers. The guard band interval is sufficient to eliminate out of band radiation. Channel delay spread is less than the cyclic prefix (CP) length at the receiver and hence there is no inter symbol interference (ISI). Doppler is not considered in the model. So the model is essentially an OFDM system with no temporal variation.

The channel is modeled based on the Power Delay Profiles (PDP) of the ITUR Pedestrian outdoor channel A (Ped A), Pedestrian outdoor channel B (Ped B) and vehicular test environment channel A (Veh A) (Jain, 2007). In this thesis we only consider Ped A and Ped B channels.

Let $X_n[k], H_n[k], Y[k], N[k]$ be the transmitted symbol, Channel frequency response, received symbol and the noise corresponding to the k^{th} subcarrier in the frequency domain and n denotes the user. Here the noise $N[k]$ is taken to be zero mean Gaussian.

$$
\begin{aligned}
Y(k) &= \sum_{n=1}^{N} X_n[k]H_n[k] \ + N[k] \\
&= X_1[k]H_1[k] + \sum_{n=2}^{N} X_n[k]H_n[k] \ + N[k] \\
&= \ X_1[k]H_1[k] + N'[k]
\end{aligned}
$$

Here the noise $N'[k]$ is the net interference plus the Gaussian noise.

CHAPTER 2

DENSITY ESTIMATION METHODS

We have discussed that in presence of strong interferers, the interference plus noise cannot be modeled as Gaussian. We try to model the conditional pdf of the interference plus noise using GMM-EM or KDE. This chapter deals with these standard density estimation methods. First the GMM-EM algorithm is introduced and then non-parametric KDE method is discussed.

2.1 Gaussian Mixture Model-EM Algorithm:

In this method, we assume that the underlying distribution is a mixture of Gaussians. The GMM consisting of K components is defined as

$$p(x) = \sum_{k=1}^{K} \pi_k CN(x|\mu_k, \Sigma_k)$$

where π_k, μ_k *and* Σ_k *denote the probability*, mean and variance of the kth component

Assuming the means, variances and probabilities are known for each component; the likelihood can be calculated and maximized. Consider N data points for training. The log likelihood is given by

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln\{\sum_{k=1}^{K} \pi_k CN(x_n|\mu_k, \Sigma_k)\}$$

EM Algorithm:

The EM algorithm is an iterative algorithm that tries to calculate the parameters of the GMM while maximizing the log likelihood.

The following are the steps of the EM algorithm:

1. Initialize the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients $\boldsymbol{\pi}_k$, and evaluate the initial value of the log likelihood.

2. **Expectation step**. $\Upsilon_{n,k}$ denotes the responsibility kth component takes for explaining data point x_n. Evaluate the responsibilities using the current parameter values

$$\gamma_{n,k} = \frac{\pi_k CN(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j CN(x_n|\mu_j, \Sigma_j)}$$

3. **Maximization step**. Re-estimate the parameters using the current responsibilities

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{n,k} x_n$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{n,k} (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

$$where\ N_k = \sum_{n=1}^{N} \gamma_{n,k}$$

4. Evaluate the log likelihood

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln\{\sum_{k=1}^{K} \pi_k CN(x_n|\mu_k, \Sigma_k)\}$$

and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to step 2.

2.2 Kernel Density Estimation:

Kernel Density Estimation is a non-parametric density estimation method. In non-parametric density estimation, we do not make restrictive assumptions about the distribution of the underlying data and let the data speak for itself. Let us understand KDE through the naïve estimator.

Parzen window or the naive estimator:

. Consider X_1, X_2, ….., X_n are the data points . From the definition of a probability density, if the random variable X has density f, then

$$p(x) = \lim_{h \to inf} P(x - h < X < x + h)$$

For any given h, we can estimate $P(x-h<X<x+h)$by the proportion of the sample falling in the interval (x-h, x+h) .Thus a density estimate can be given by choosing a small number h and setting

$$p(x) = \frac{1}{2nh}\{No. of\ X_1, X_2 \ldots X_n\ falling\ \ in\ (x - h, x + h)\}$$

To find the number of data points that fall within the region we define Kernel function *K(x)*.

$$K(x) = \frac{1}{2} \qquad if\ |x| < 1$$

$$= 0 \qquad otherwise$$

This kernel which corresponds to a unit hyper-cube centred around origin is called the Parzen window or naive estimator. The estimator can be written as

$$p(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K(\frac{x - X_i}{h})$$

This idea can be extended to d dimensions. Assume the region R is a d-dimensional hypercube with h_i being the length of an edge in i^{th} dimension. The volume of the hypercube is given by

$$V = h_i^d$$

Parzen window estimate can be thought of as sum of windows or boxes which are centred around the data points.

Consider the histogram constructed from the data using bins of width h. Assume that no observations lie exactly at the edge of a bin. If x happens to be at the centre of one of the histogram bins, the naive estimate p(x) will be exactly the ordinate of the histogram at x . The naive estimate tries to construct a histogram where every point is the centre of a sampling interval, freeing the histogram from a particular choice of bin positions.

The choice of bin width remains and is governed by the parameter h which controls the smoothing. It is not a continuous function, but has jumps at the points X_i+h and X_i-h and has zero derivative everywhere else. All points have equal weight, irrespective of their distance from the estimation point x.

It is easy to generalize the estimator to overcome some of the difficulties discussed above. Use a kernel function *K* which satisfies the condition

$$\int_{-inf}^{inf} K(x)dx = 1$$

Usually, but not always, *K* will be a symmetric probability density function, the normal density, for instance, or the weight function used in the definition of the naive estimator. By analogy with the definition of the naive estimator, the kernel estimator with kernel *K* is defined by

$$p(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

where *h* is the window width , also called the smoothing parameter or bandwidth .

Let $X_1, X_2, \ldots, X_n$ be a sample of d-variate random vectors drawn from a common distribution described by the density function *f*. The multivariate kernel density estimate is defined to be

$$p_H(x) = \frac{1}{nh}\sum_{i=1}^{n} K_H(\frac{x - X_i}{h})$$

H is the bandwidth (or smoothing) dxd matrix which is symmetric and positive definite *K* is the kernel function .

Just as the naive estimator can be considered as a sum of boxes centred at the observations, the kernel estimator is a sum of `bumps' placed at the observations. The kernel function *K* determines the shape of the bumps while the window width h determines their width.

Kernel estimation of pdfs is characterized by the kernel, *K*, which determines the shape of the weighting function, and the bandwidth, *h*, which determines the "width" of the weighting function and the amount of smoothing. The two components determine the properties of p(x). *K* and *h* are to be selected in order to optimize the properties of p(x).

Density estimation aims at taking a finite sample of data and makes inferences about the underling probability density function everywhere including where no data are observed. In kernel density estimation, the contribution of each data point is smoothed out from a single point into a region of space surrounding it. Aggregating the individually smoothed contributions gives an overall picture of the structure of the data and density function.

Complexity:

If the density is being estimated at m points and the given data contains n samples of training data, the complexity of KDE is $O(mn)$.

Optimizing the kernel density estimate:

Consider the mean squared error (MSE) and its two components, namely bias and variance.

$$MSE(\,p(x)) = E\big(p(x) - f(x)\big)^2$$

$$= \big(Ep(x) - f(x)\big)^2 + E\Big(p(x) - E\big(p(x)\big)\Big)^2$$

$$= bias^2 p(x) + variance(p(x))$$

A measure of the global accuracy of p(x) is the mean integrated squared error (MISE)

$$MISE(\,p(x)) = \int E\big(p(x) - f(x)\big)^2 dx$$

$$= \int bias^2 p(x) dx + \int variance(p(x)) dx$$

It can be derived that

$$MISE(\,p) = \frac{1}{4} h^4 {k_2}^2 \;\; \beta(f) + \frac{1}{nh} j_2$$

where $k_2 = \int z^2\, K(z) dz$, $j_2 = \int K(z)^2\, dz$ and $\beta(f) = \int f''^2(x) dx$

Of central importance is the way in which MISE(p)changes as a function of the bandwidth h. For very small values of h the second term becomes large but as h gets larger so the first term increases. There is an optimal value of h which minimizes MISE(p).

Optimal bandwidth Expression is the measure that we use to quantify the performance of the estimator. We can find the optimal bandwidth by minimizing with respect to h. The first derivative is given by

$$d(MISE\big(p(x)\big))/dh = h^3 {k_2}^2 \beta(f) - (\frac{1}{nh})^2 j_2$$

Setting this equal to zero yields the optimal bandwidth, h_{opt} , for the given pdf and kernel

$$h_{opt} = (\frac{1}{n}\frac{\gamma(K)}{\beta(f)})^{1/5}$$

where $\gamma(K) = j_2 {k_2}^{-2}$

We note that h_{opt} depends on the sample size, n, and the kernel, K. However, it also depends on the unknown pdf f, through the functional $\beta(f)$. Thus as it stands expression is not applicable in practice.

Choice of Kernel function:

The MISE can also be minimized with respect to the kernel used. It can be shown (see, e.g., Wand and Jones, 1995) that Epanechnikov kernel is optimal in this respect.

$$K(x) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}x^2) \qquad if\ |x| < \sqrt{5}$$

$$= 0 \qquad otherwise$$

The efficiency of a kernel, K, relative to the optimal Epanechnikov kernel, K_{EP} , is defined as

$$Eff(K) = MISE_{opt}\left(p(x)\right) using\ K_{EP} / MISE_{opt}\left(p(x)\right)^{5/4}$$

The selection of kernel has rather limited impact on the efficiency as mentioned in [5].

Bandwidth selection:

The problem of choosing h is crucial in density estimation. A large h will over-smooth the pdf and mask the structure of the data. A small h will yield a spiky pdf which would be difficult to interpret. We would like to find a value of h that minimizes the error between the estimated density and the true density.

Univariate:

Subjective Selection

The natural way for choosing h is to plot out several curves and choose the best estimate. It is not practical in pattern recognition since we typically have high-dimensional data.

Selection with reference to a standard distribution:

Assume a standard density function and find the value of the bandwidth that minimizes the integral of the square error MISE. If we assume that the true distribution is Gaussian and we use a Gaussian kernel, it can be shown that the optimal value of h is $h_{opt} = 1.06\sigma N^{-1/5}$

Cross validation:

The ML estimate of h is degenerate since it yields h=0 , a density estimate with Dirac delta functions at each training data point. A practical alternative is to maximize the "pseudo-likelihood" computed using leave-one-out cross-validation.

$$h_{opt} = argmax\{\frac{1}{n}\sum_{i=1}^{n} p_{-n}(x_i)\}$$

$$p_{-n}(x_i) = \frac{1}{h(n-1)} \sum_{m=1, m!=i}^{n} K(\frac{x_i - x_m}{h}) \}$$

Multivariate data:

Kernel density estimation for multivariate data is an important technique that has a wide range of applications. However, it has received significantly less attention than its univariate counterpart. The lower level of interest in multivariate kernel density estimation is mainly due to the increased difficulty in deriving an optimal data- driven bandwidth as the dimension of data increases.

The choice of the kernel function K is not crucial to the accuracy of kernel density estimators while the choice of the bandwidth matrix H is the single most important factor affecting its accuracy since it controls the amount of and orientation of smoothing induced. One basic difference between multivariate and univariate case is that the bandwidth matrix also induces an orientation which is not defined for 1D kernels.

We use subjective selection to fix the bandwidth in our case by using trial and error.

CHAPTER 3
OPTIMAL SINGLE ANTENNA RECEIVERS

Introduction:

As discussed earlier, we try to model the conditional pdf of the interference and noise mixture as a non-Gaussian pdf. After the conditional pdf is obtained, the probability of correct decision is maximized by using the Maximum Likelihood (ML) rule and not the minimum distance rule. Performing ML detection on the conditional pdf should give optimal error rate performance compared to linear receivers. We first study the pdf of the interference in the presence of a single antenna receiver .This can later be extended for the case of multiple antenna receivers.

ML estimate is given by

$$\hat{x} = argmax\{p(y - h_1 x)\}$$

Single Antenna :

The system in this case involves a single QPSK interferer. The SNR is varied between 5-25 dB. The graphs plotted are for Ped B channel. The conditional pdf is estimated using KDE and GMM-EM considering a range of frequencies. The size of the training data is taken as a multiple of the frequency range assigning a fixed number of data points per each frequency.

We first try to fix number of points in training data. The following graphs indicate that the BER floors after training data per frequency of 10 for interference values of -3 dB, 0 dB and 3 dB.

3.1 BER Vs Training data per frequency :

a) Frequency Range=8

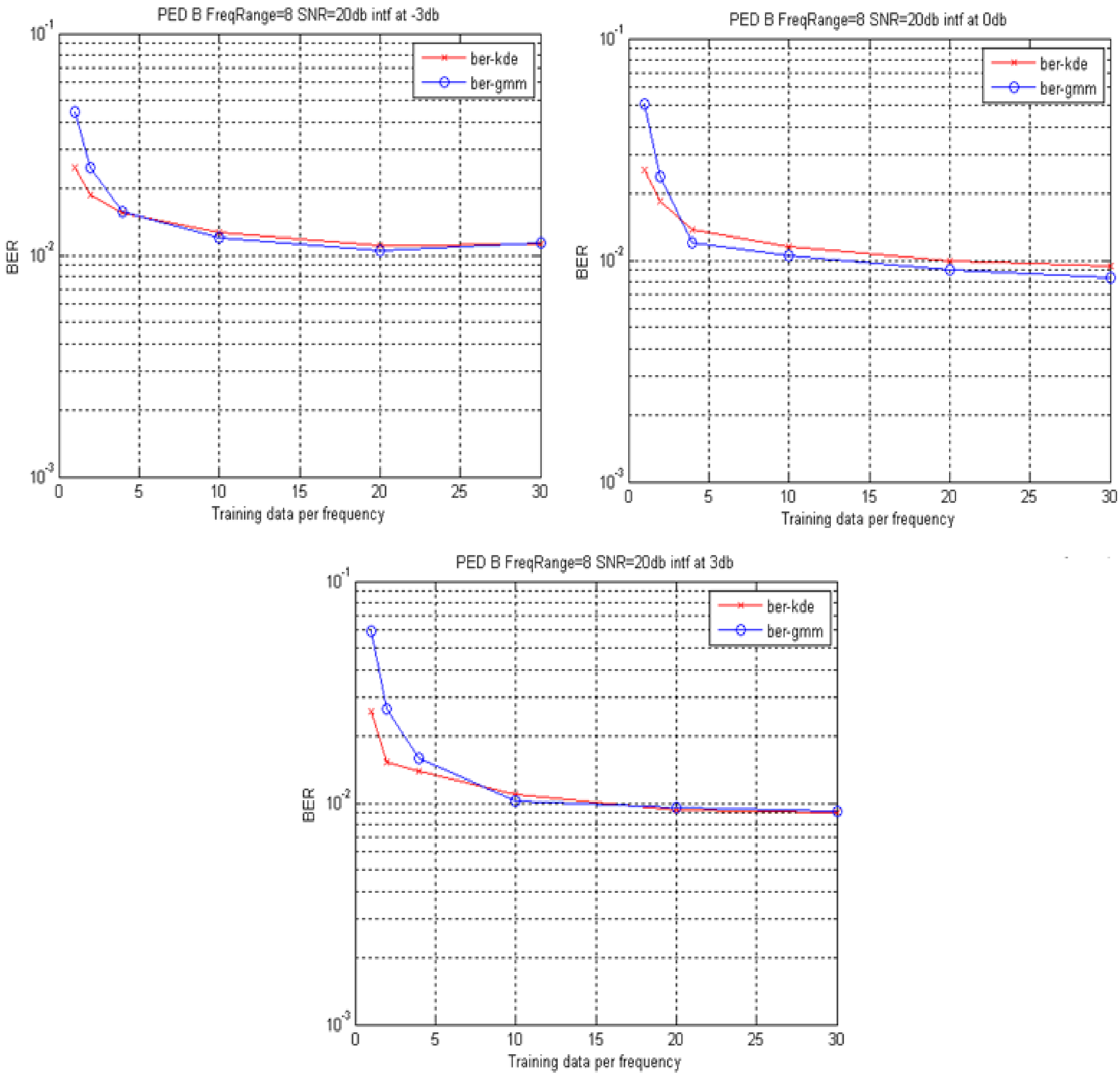


Fig 3.1 BER Vs Training data per frequency for GMM-EM and KDE for frequency range=8 at SNR =20 dB

b) Frequency Range=12

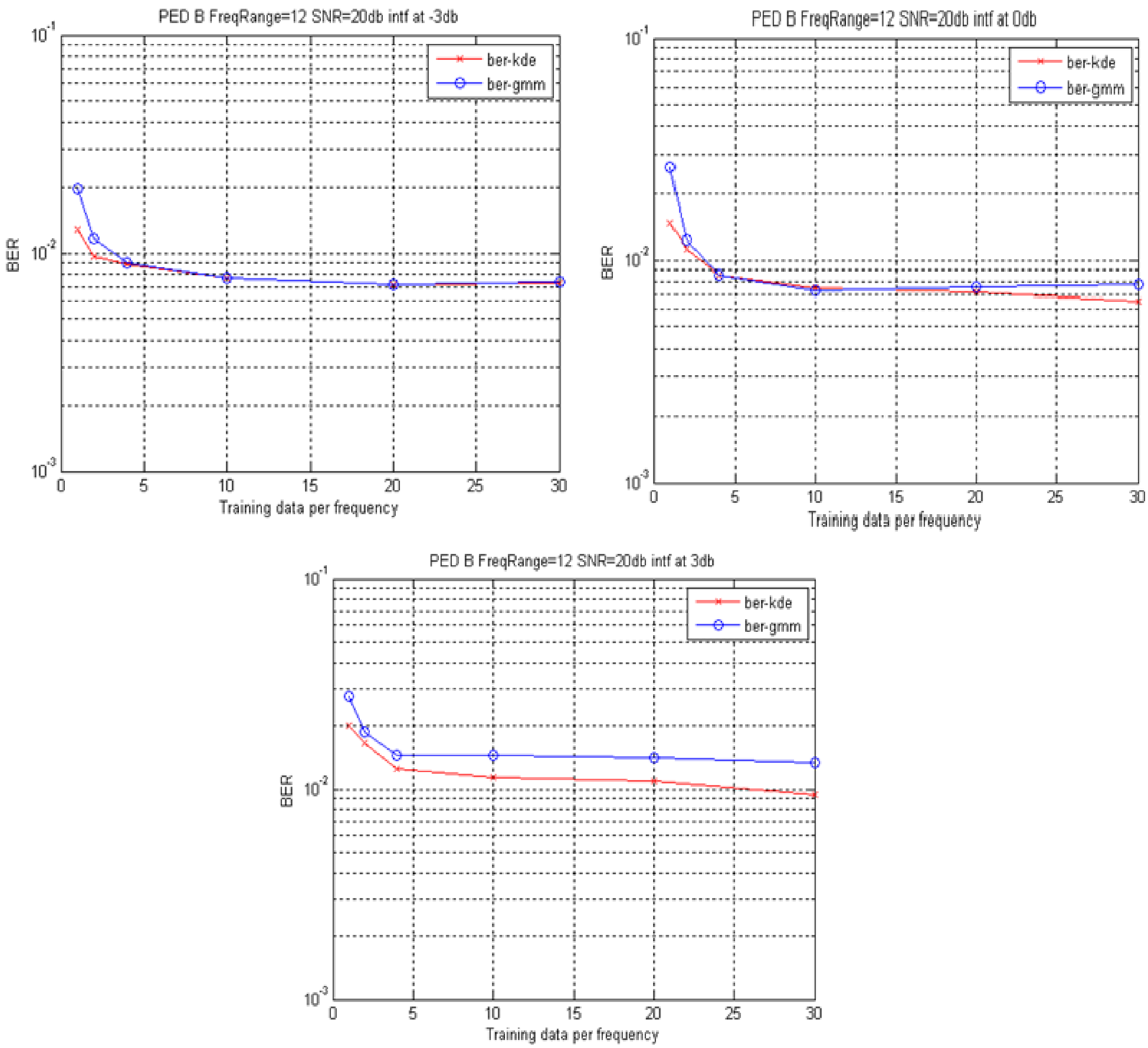


Fig 3.2 BER Vs Training data per frequency for GMM-EM and KDE for frequency range=12 at SNR =20 dB

c) Frequency Range=18

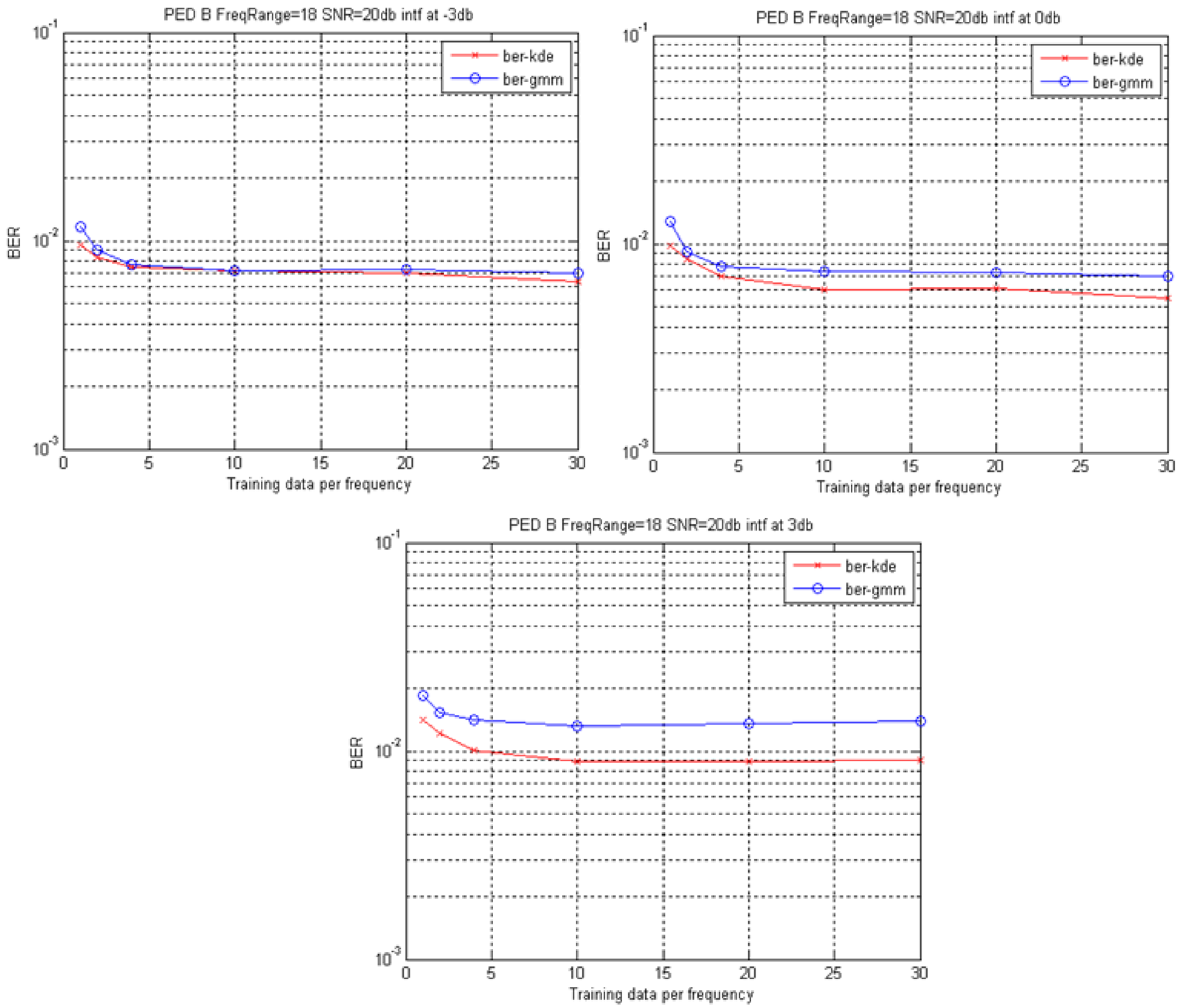
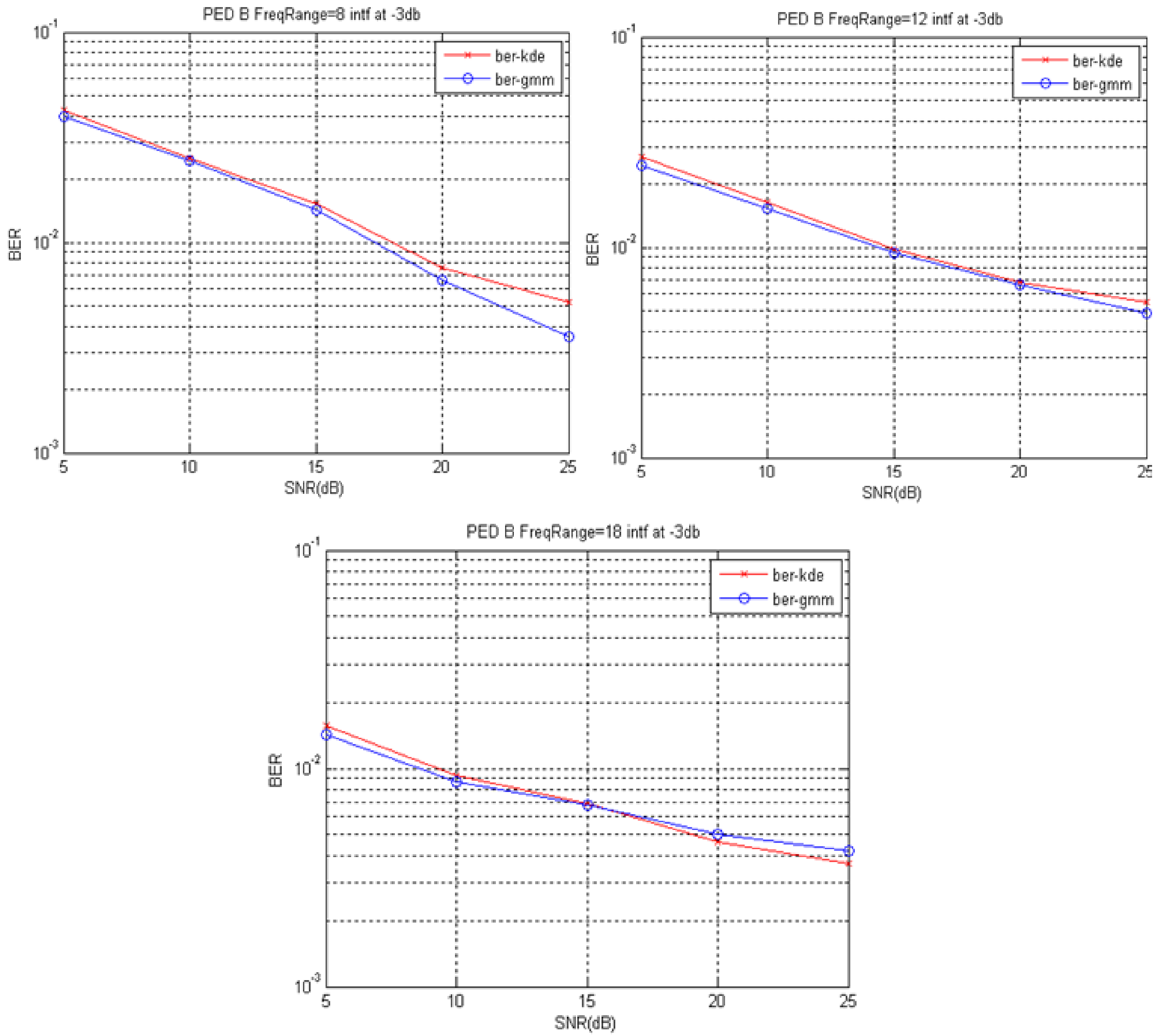


Fig 3.3 BER Vs Training data per frequency for GMM-EM and KDE for frequency range=18 at SNR =20 dB

.

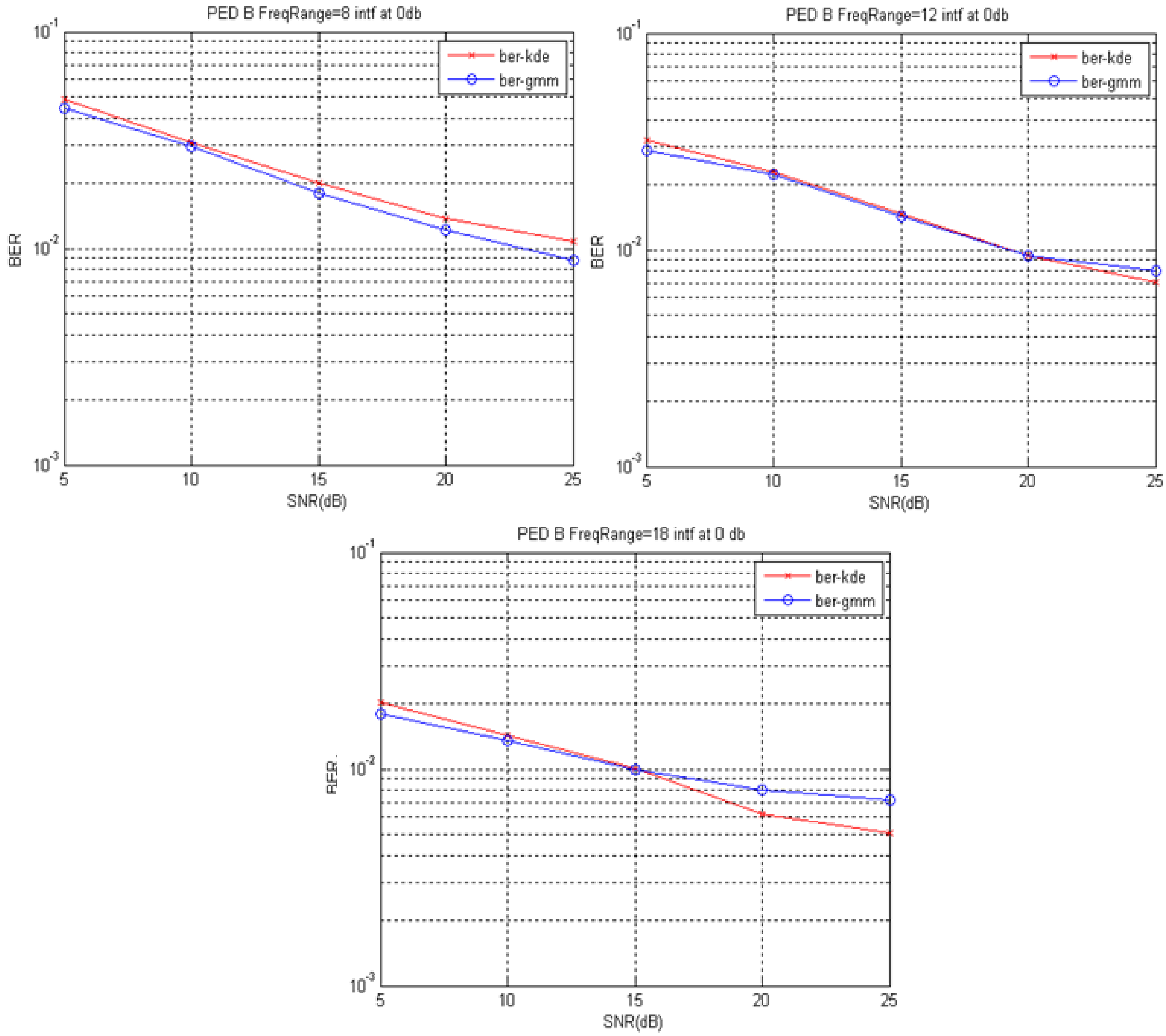
3.2 BER Vs SNR:

a) Interferer at -3 dB



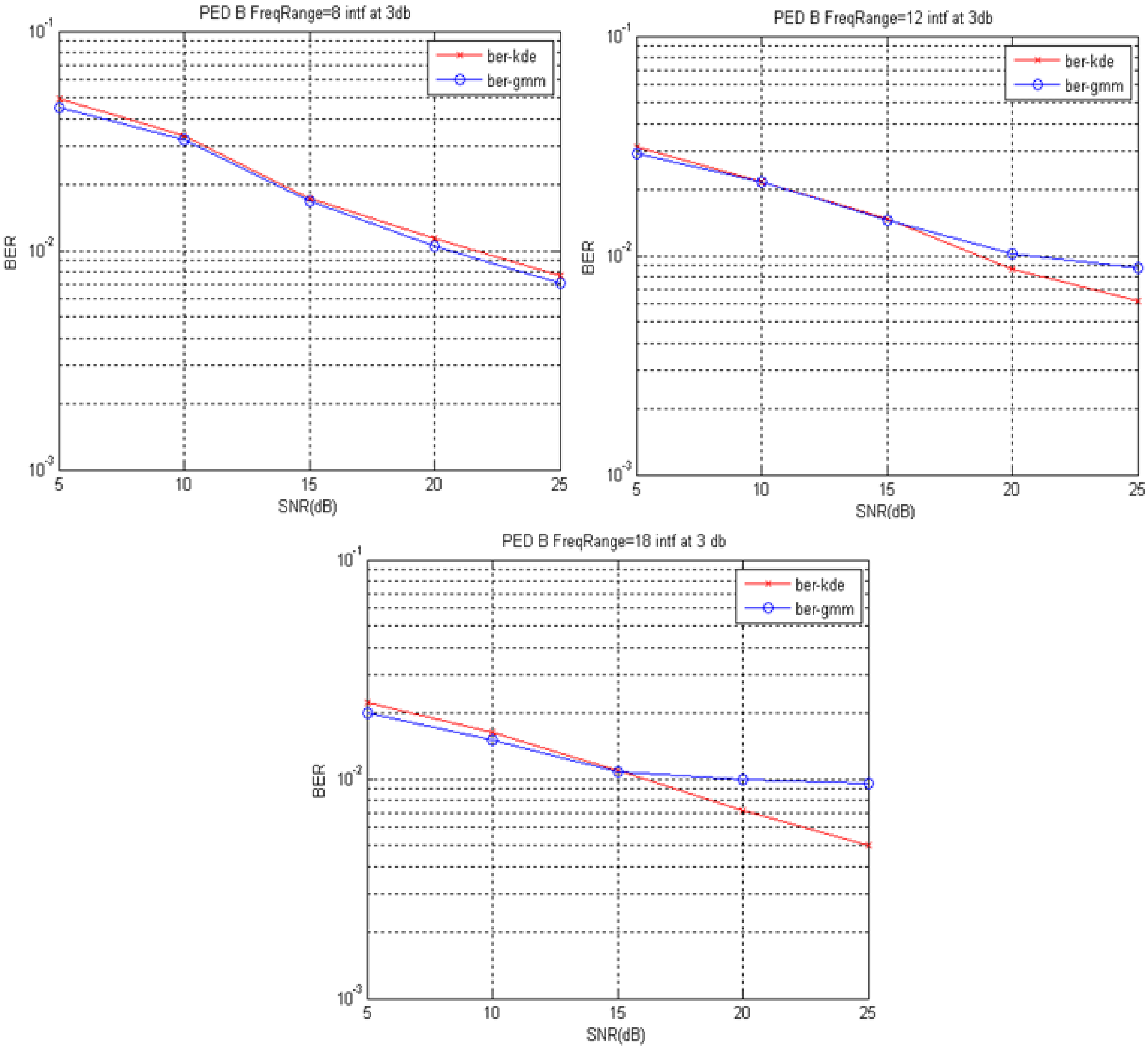
3.4: Performance (BER Vs SNR) Comparison of GMM-EM and KDE for interferer at -3 dB

b) Interferer at 0 dB



3.5: Performance (BER Vs SNR) Comparison of GMM-EM and KDE for interferer at 0 dB

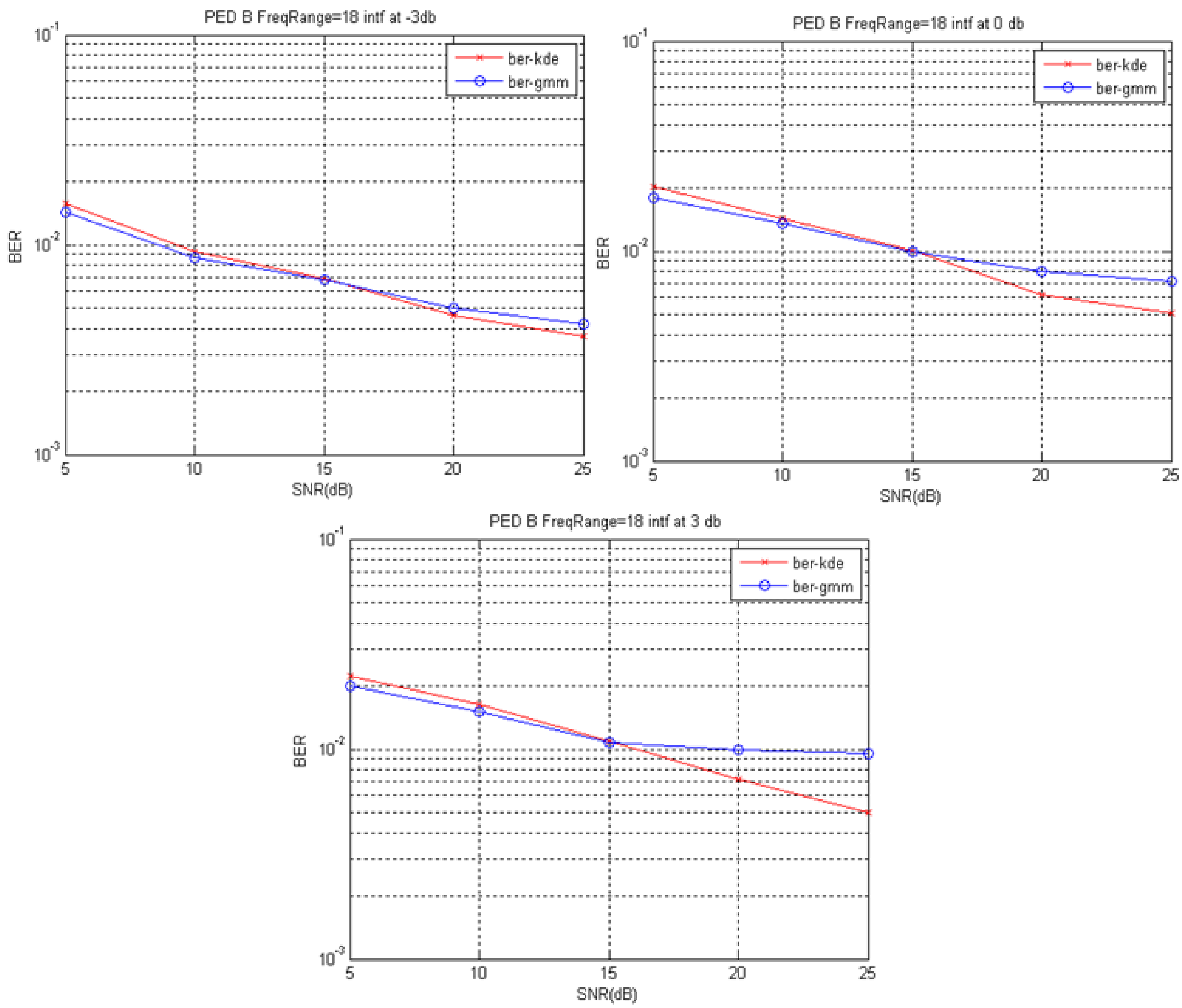
c) Interferer at 3 dB



3.6: Performance (BER Vs SNR) Comparison of GMM-EM and KDE for interferer at 3 dB

The value of training data per frequency=10 is used to plot BER curves. For each interference value, GMM-EM and KDE are compared for frequency range of 8, 12 and 18 carrier frequencies We observe that at all three values of SINR, for a range of frequencies around 8, the performance of GMM is better. As the range increases to 12 and 18, KDE is observed to perform better at higher SNR. The crossover is typically seen to occur at SNR=15dB and in between frequency range of 8-10.

We now compare the SNR for the different interference values.



3.7: Performance (BER Vs SNR) Comparison of GMM-EM

and KDE for frequency range=18

We observe that the BER decreases with the increase in the power of the interferer. This is seen for SNR values of 5 to 20. However, at SNR value of 25, it seems a bit ambigious.

CHAPTER 4
CONCLUSION AND FUTURE WORK

4.1 Conclusion:

In the presence of strong interference, the interference plus noise is no longer Gaussian and the commonly used methods like LMMSE are not very effective. There is a need to design more sophisticated receivers than can handle interference mitigation.

The Gaussian Mixture Model-Expectation Maximization algorithm was studied for a simple static flat faded Rayleigh channel. We now try to extend the GMM-EM to a frequency selective OFDM channel (without temporal variation). Over the frequency selective channel, small ranges of frequency are considered and the average pdf is estimated using both the methods. Over a slightly higher range of frequencies, GMM component parameters might tend to a mean of its parameters for each channel. So another method is considered that does not assume any parameters for the underlying pdf and sees the data as it is. This method is the non-parametric Kernel Density Estimation.

The performance of an ML receiver with both the methods is studied with BER as the performance indicator. It is observed that the performance of GMM-EM over smaller frequency range (<8-10) of the frequency selective channel was better than that of KDE but as the frequency range is slightly increased (12-18) , both the methods perform equally well at SNR upto 15 dB with KDE surpassing the performance of GMM after 15 dB. This was observed over 3 different values of the SINR.

4.2 Future Work:

The work can be extended to the case of multiple interferers and also multiple receiver antenna case.

The complexity of the KDE algorithm can be reduced but at the cost of accuracy. The accuracy and cost tradeoff can be studied.

We have considered only frequency selective OFDM system and that too varying only across a few resource blocks. It would be interesting to study the behavior of KDE and GMM when temporal variation is introduced.

It would be useful to study the performance of the channel when coding is introduced. Instead of finding the mean pdf of interference plus noise as the channel changes slowly, it would be useful to track the pdf over the channel change.

.

REFERENCES

[1] Tse, D. and P. Viswanath, Fundamentals of Wireless communication, Cambridge University Press, 2005.

[2] Vaishnavi, J. (2012). Advanced Receiver Techniques based on Density Estimation for Inter-cell Interference Mitigation. Masters thesis, Department of Electrical Engineering, IIT-Madras.

[3] Cho, Y. S., J. Kim, W. Y. Yang, and C.-G. Kang, *MIMO-OFDM* Wireless Communications with *MATLAB*. John Wiley & Sons, 2010.

[4] Bishop, C., Pattern Recognition and Machine Learning. Springer, 2006.

[5] B. Silverman, Density estimation for statistics and data analysis. Chapman & Hall, 1986, vol. 26.

[6] Walter Zucchini (October 2003), Applied Smoothing Techniques. http://isc.temple.edu/economics/Econ616/Kernel/ast_part1.pdf