

Using RBF Neural Networks and Kullback-Leibler distance to classify channel models in Free Space Optics

Geetha Prakash

PES Institute of Technology, Bangalore
geetha.prakash@pes.edu

Muralidhar Kulkarni, U Sripati

National Institute of Technology, Suratkal
mkuldce@gmail.com, sripati_acharya@yahoo.co.in

Abstract—Free Space Optical (FSO) communication systems offer a license free and cost effective access performance. FSO systems provide virtually unlimited bandwidth. Since the laser beams used in these systems are spatially confined, the links are very secure. However FSO links perform well only in clear weather conditions. Clouds, fog, aerosols, and turbulence drastically affect the performance of FSO systems and lead to fluctuations in both the intensity and phase of the received signal. FSO links can suffer from data packet corruption and erasure. Various statistical models have been proposed to describe the atmospheric turbulence channels. The choice of the appropriate model for varying level of turbulence is dependent on the atmospheric parameters. In this paper we classify the channels using Radial Basis Function Neural Networks to decide the best fit. We also use Kullback-Leibler distance as a measure between the reference distribution and the distribution of observed data.

Key words: Free Space Optics, Distribution functions, Channel models, Radial Basis Functions, Kullback-Leibler distance.

I. INTRODUCTION

Free space optical communication (FSO) is the technology that can take care of connectivity needed in optical networks and is a viable solution for the last mile bottleneck problem. FSO uses modulated optical beams to establish wireless communication. With frequencies in the optical range, data rates of the order of giga bits per second are achievable. Also, FSO links are difficult to intercept and are immune to interference or jamming from external sources [1].

The main disadvantage of communications over the FSO channel is the effect of the atmosphere on the optical signal. Atmospheric effects can degrade free space optical links through two different mechanisms. Atmospheric attenuation leading to reduction in the detected optical power level. Atmospheric turbulence leading to random optical power fluctuations, resulting in beam formation, beam wander and scintillation effects. Atmospheric attenuation is due to interaction of the laser beam with air molecules and aerosols during propagation [2, 3, 4, 5].

Several channel models have been proposed to characterize FSO channel models. In this paper we present a statistical FSO channel model classification using Radial Basis Neural Networks (RBFNN) considering two different distribution

functions. The presented novel approach to classification is based on Radial Basis functions as classifiers with Gaussian function as the activation function.

The paper is organized as follows. In Section II we review the basic scintillation process and two different channel models which can be used to approximate the behavior of the FSO link. The channel is modeled using Gamma-Gamma Distribution and the K distribution. Section III classifies the distribution functions using Radial Basis Function Neural Networks as classifiers. Section IV gives the Kullback-Leibler distance and Sections V and VI the simulation results and conclusions.

II. FSO CHANNEL MODEL DESCRIPTION

A. Atmospheric turbulence and the scintillation process.

We consider a terrestrial FSO system with the laser beams propagating along a horizontal path through a turbulent channel. Propagation of optical waves through the atmosphere is affected by atmospheric turbulence, scattering off aerosols, and atmospheric absorption. Atmospheric turbulence is a result of localized variations of temperature, humidity, and pressure in the atmosphere. Turbulence is by nature a random process, and may be described using statistical quantities. Air packets at different temperatures when mixed by wind and convection lead to atmospheric turbulence. This effect produces fluctuations in the density, leading to variations in the air refractive index.

Scintillations are random fluctuations in the intensity of the received signal caused by the inhomogeneities in the atmospheric temperature and pressure along the optical beam propagation path. These random fluctuations lead to increased bit error probabilities, thus degrading the system performance. Beam wandering is due to relatively slower motion of the turbulent eddies in the beam path [6].

The received signal at the detector fluctuates as a result of the thermally induced changes in the index of refraction of the air. Air packets at different temperatures when mixed by wind and convection lead to atmospheric turbulence. This effect produces fluctuations in the density, leading to variations in the air refractive index. Scintillations and beam wandering occur

across the receiver plane due to the turbulence in the atmosphere [6, 7, 8].

Rytov approximation or the method of smooth perturbations is widely used to describe line-of-sight propagation in turbulent media when the amplitude variations are small. The Rytov approximation is fundamentally an enlargement of geometrical optics that includes diffraction effects. The essence of this method is to express the field strength as the product of the unperturbed field and the exponential of a surrogate function. It is a complex function that describes the important influence of diffraction [9].

The Rytov method provides a closed form expression for the intensity fluctuations of a plane wave in weak turbulence condition for a point receiver. The received signal scintillations variance is given by

$$\sigma_I^2 = 1.23 \times C_n^2 \times k^{7/6} \times L^{11/6} \quad (1)$$

Where C_n^2 is the refractive index structure parameter, 'k' is the wave number of the transmitted beam and L is the link length in meters.

The Rytov method predicts that the signal variance due to turbulence can increase without limit for stronger turbulence [6]. Relative variation of optical intensity σ_{Ir}^2 can be expressed as

$$\sigma_{Ir}^2 = \frac{\langle I^2 \rangle - \langle I \rangle^2}{\langle I \rangle^2} = \frac{\langle I^2 \rangle}{\langle I \rangle^2} - 1 \quad (2)$$

where I is the optical intensity of the received signal and $\langle \cdot \rangle$ signifies mean value. When $\sigma_{Ir}^2 \ll 1$, then we can use Rytov approximation which ties relative variation of optical intensity and refractive index structure parameter [8].

The transmitter modulates data on to the instantaneous intensity of an optical beam. Intensity modulated direct detection channels using OOK modulation, is widely employed in practical systems. The received photocurrent signal is related to the incident optical power by the detector responsivity R. It is assumed that the receiver integrates the photocurrent for each bit period and removes any constant bias due to background illumination. The received signal 'y' suffers from a fluctuation in signal intensity due to atmospheric turbulence and misalignment, as well as additive noise, and can be well modeled as

$$y = h R x + n$$

where x is the transmitted intensity, h is the channel state, y is the resulting electrical signal, and n is signal-independent additive white Gaussian noise with variance σ_n^2 . The channel state h models the random attenuation of the propagation channel [10].

B. Channel Model represented by different distributions.

Let the independent and identically distributed symbols be denoted by x_k , each with energy E_s . Let x_k be subject of an independent fading process s_k , which accounts for the intensity

fluctuations due to scintillation effects. Hence if the received signal y_k is assumed to be distorted by zero-mean white Gaussian noise n_k with power spectral density $N_0/2$, we have

$$y_k = s_k x_k + n_k \quad (3)$$

where s_k follows the Gamma-Gamma distribution

For a wide range of turbulence conditions, the fading gain in FSO systems can be modeled by a Gamma-Gamma distribution function

$$f(I) = \frac{2(\alpha\beta)^{(\alpha+\beta)/2}}{\Gamma(\alpha)\Gamma(\beta)} I^{\frac{\alpha+\beta}{2}-1} K_{\alpha-\beta}(2\sqrt{\alpha\beta}I), I > 0 \quad (4)$$

Where $\Gamma(\cdot)$ is the standard Gamma function $K_a(\cdot)$ is the modified Bessel function of the second kind of order 'a'. Here $\alpha > 0$ and $\beta > 0$ are parameters linked to the scintillation index. If spherical wave propagation is assumed, α and β can be directly linked to the physical parameters

$$\alpha = \left[\exp\left(\frac{0.49\chi^2}{(1+0.18d^2+0.56\chi^{12/5})^{7/6}}\right) - 1 \right]^{-1}$$

$$\beta = \left[\exp\left(\frac{0.51\chi^2(1+0.69\chi^{12/5})^{-5/6}}{(1+0.9d^2+0.62d^2\chi^{12/5})^{5/6}}\right) - 1 \right]^{-1} \quad (5)$$

where $\chi^2 = 0.5C_n^2 k^{7/6} L^{11/6}$, $d = (kD^2 / 4L)^{1/2}$ and $k=2\pi/\lambda$. Here λ , D, C_n^2 , and L are the wavelength in meters, the diameter of the receiver's aperture in meters, the index of refraction structure parameter, and the link distance in meters respectively.

In general, the turbulence can be classified into 3 types – Weak turbulence, moderate turbulence and strong turbulence.

To obtain the plot of Gamma-Gamma distribution, we set the value of the Rytov variance $\chi=0.2, 1$ and 3 to obtain the values of parameters α and β as in Table I.

TABLE I. PARAMETERS FOR GAMMA-GAMMA DISTRIBUTION

Turbulence	χ	α	β
Weak	0.2	51.9	49.1
Moderate	1	4.39	2.56
Strong	3	5.49	1.12

The plot of Gamma- Gamma distribution is shown in Fig 1.

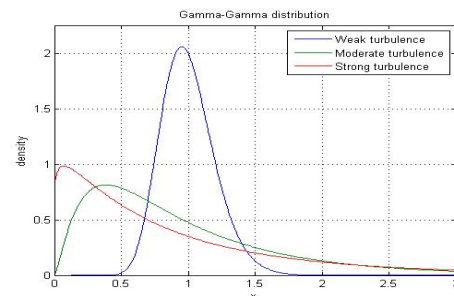


Fig. 1. Gamma-Gamma Distribution

From the plot in Figure 1, we find that the Gamma distribution is symmetric for weak turbulence. But for moderate turbulence the symmetric property of the distribution decreases. For strong turbulence the distribution is not at all symmetric. In view of this, the effect of weak turbulence is equal on both 0 and 1 but as the turbulence increases the effect increases and as a result the scintillation effect increases with the increasing turbulences.

The K Distribution is suitable for strong turbulence channels.

If h_a is the attenuation due to atmospheric turbulence,

$$f_{h_a}(h_a) = \frac{2\alpha^{\frac{\alpha+1}{2}}}{\Gamma(\alpha)} h_a^{\frac{\alpha-1}{2}} K_{\alpha-1}(2\sqrt{\alpha h_a}), \quad h_a > 0 \quad (6)$$

Where α is a channel parameter related to the effective

number of discrete scatterers, $\Gamma(\bullet)$ is the well known Gamma function [11]. When $\alpha \rightarrow \infty$, the distribution approaches the negative exponential distribution [11].

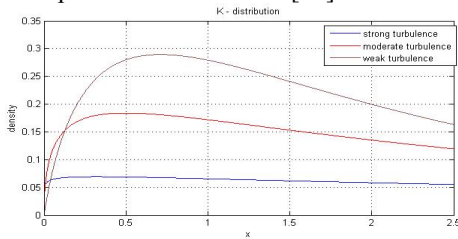


Fig. 2. K Distribution

The plot of K distribution for different values of α is shown in Figure 2

III. CLASSIFICATION OF THE DISTRIBUTION FUNCTIONS USING RADIAL BASIS FUNCTIONS.

A. Radial Basis Function Neural Networks.

The FSO channel models represented by different distribution functions are similar in shape for the same coefficient of variation, making it difficult to identify the differences between the three densities. It is therefore very essential to classify the channel models before approximating the channel model and estimating the performance of the FSO link. Radial Basis Function neural networks have been widely applied to pattern classification. This paper applies this model to the discrimination, from data, between three similar probability density functions. The analysis is based on three key parameters pertaining to a distribution function namely kurtosis, skewness and mean [12].

In the classical approach to radial basis function RBF network implementation, the basis functions are usually chosen as Gaussian data. The weights connecting the hidden units to the output layer are normally determined by a least mean squares algorithm [13,14].

The construction of an RBF network involves three layers. The input layer is made up of source nodes that connect the

network to the environment. The second layer applies a non linear transformation from the input space to the hidden space. The output layer is linear supplying the response of the network to the activation pattern applied to the input layer.

The RBF technique consists of choosing a function F that has the form

$$F(x) = \sum_{i=1}^N w_i \varphi(\|x - x_i\|) \text{ where } \{\varphi(\|x - x_i\|) \mid i = 1, 2, \dots, N\} \quad (7)$$

is a set of N arbitrary functions known as radial basis functions and $\|\cdot\|$ denotes a norm that is Euclidean. The known data points $x_i \in \mathfrak{R}^m, i = 1, 2, \dots, N$ are taken to be centers of the radial basis functions.

For the unknown coefficients (weights), a set of simultaneous equations are obtained as

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2N} \\ \dots & \dots & \dots & \dots \\ \varphi_{N1} & \varphi_{N2} & \dots & \varphi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_N \end{bmatrix} \quad (8)$$

where $\varphi_{ji} = \{\varphi(\|x - x_i\|), (j, i) = 1, 2, \dots, N\}$

Let

$$\bar{d} = [d_1, d_2, \dots, d_N]^T$$

$$\bar{w} = [w_1, w_2, \dots, w_N]^T$$

The N by 1 vectors \bar{d} and \bar{w} represent the desired response vector and linear weight vector respectively, where N is the size of the training sample. Let Φ denote an N -by- N matrix with elements φ_{ji}

$$\bar{\Phi} = \{\varphi_{ji} \mid (j, i) = 1, 2, \dots, N\} \quad (9)$$

This is called the interpolation matrix

$$\bar{\Phi} \bar{w} = \bar{x} \quad (10)$$

Assuming that $\bar{\Phi}$ is non singular and therefore the inverse matrix $\bar{\Phi}^{-1}$ exists and the weight vector can be obtained as

$$\bar{w} = \bar{\Phi}^{-1} \bar{x} \quad (11)$$

The error at the n th time step between estimated output $\hat{y}(n)$ and the desired output $y(n)$ is

$$e(n) = |y(n) - \hat{y}(n)| \quad (12)$$

The least square principle is used to minimize the sum-squared error with respect to the weights of the model [13].

The characteristic feature of the radial function is that their response decreases or increases monotonically with distance from a central point. The center and the precise shape of the radial function are parameters of the model, all fixed if it is linear. In order to specify the middle layer of an RBFN Network, we have to decide the number of neurons of the layer and their kernel functions which are usually Gaussian functions. In this paper, we use a Gaussian function as a kernel function. A Gaussian function is specified by its center and width.

B. Design of RBF Networks as classifiers

The RBFN Network is designed by considering 200 x 500 data points and minimizing a specified goal. We iteratively create a radial basis network one neuron at a time. Neurons are added to the network until the sum-squared error falls beneath an error goal or a maximum number of neurons have been reached. At each iteration the input vector that results in lowering the network error the most, is used to create a neuron. The error of the new network is checked, and if it is low as specified, the simulation is complete. Otherwise the next neuron is added. This procedure is repeated until the error goal is met, or the maximum number of neurons is reached.

RBFNNs are used to classify the distribution schemes into two different classes, namely Gamma-Gamma and K distribution. We have specifically chosen these two different distributions because K -Distribution function gives good correlation between theoretical and practical results as in [15] and Gamma-Gamma Distribution works well for a wide range of turbulence, from weak to strong.

We use Gaussian function as the activation function for the RBF networks. A different value for the goal which is the mean squared error is set and the training is done for the neural network. Training is done using predefined values of a 200 x 500 matrix. The target matrix is T which is used to classify the distribution functions for Gamma-Gamma is $\begin{bmatrix} 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \end{bmatrix}$ for K distribution. Classification of the distribution function is done based on three features mean, kurtosis and skewness. The entries in the matrices will decide the classification.

Kurtosis a descriptor of the shape of the distribution function and is the degree of peakedness of a distribution, defined as a normalized form of the fourth central moment of a distribution. Skewness is a measure of the degree of asymmetry of a distribution [16].

Our simulations compute the values in the target matrix based on the skewness and kurtosis of these distribution functions, make a comparison and then classify them as either Gamma-Gamma or K-distribution functions

For RBFNNs, the number of hidden neurons are obtained as a result of the training which depends on mean squared goal and spread.

C. Performance of RBFNN Classifiers

The apparent error rate is the fraction of observations which are misclassified by RBFNN classifiers.

$$A_{per} = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k N_i} \tag{13}$$

where N_i is the total number of observations for $i=1,2,3\dots$ and n_i is the number of items misclassified.

IV. KULLBACK-LEIBLER DISTANCE

Fitting a parametric model to data is a solution for the proper choice of an appropriate model. There are several features that make system identification a difficult and nontrivial task.. The stochastic behaviour of data is usually the main source of uncertainty in system identification. In practice, data available for identification are rarely complete. Frequently some data items are completely missing. Data may also be corrupted by noise or systematic errors. The assumption that the true system belongs to a certain model class is commonly accepted. The use of approximation increases the uncertainty of system identification. Data are supposed to be a sequence of random variables with values in a finite set. The distribution is not known completely but it is assumed to belong to a family $\{S_\theta : \theta \in \Theta\}$. Kullback-Leibler distance is also known as relative entropy, cross entropy, informational divergence, or information for discrimination. Kullback-Leibler distance is defined by

$$D(R \| S) = \sum_{x \in X} R(x) \log \frac{R(x)}{S(x)} \tag{14}$$

When making the reference about the unknown θ , the probability mass function S_θ^k is considered as a function of θ

for given x. This is the likelihood function $L_x(\theta) = S_\theta^k(x)$.

is so chosen that it maximizes the likelihood $\max_{\theta} L_x(\theta)$. Maximum likelihood estimation is equivalent to minimum Kullback-Leibler distance estimation

$$\min_{\theta} D(R_x \| S_\theta) \tag{15}$$

Some properties of Kullback-Leibler distance demonstrate that the use of Kullback-Leibler distance for inference is consistent with common statistical approaches [17]. We have classified the different models using Radial Basis Function Neural Networks. However to decide if the channel model is the appropriate choice, we use Kullback-Leibler distance

V. SIMULATION RESULTS

Tables II and III give the classification of the different channel models using RBFN networks. Table IV gives the misclassification results. Table V and VI provide the sample values for K distribution and α and β for Gamma-Gamma and mean, kurtosis and skewness. Classification was done with 400 neurons in the network and a computation time of around 33

seconds. The differences between the channel distribution functions become significant in their tail behaviour. Distributions with a relatively large tail are called leptokurtic. The evaluation of mean, skewness and kurtosis of the distribution functions will describe the tail of the distribution functions and also the location of the mean. The skewness describes the asymmetry in the distribution. Thus the computation of these parameters will describe the nature of turbulence. Strong, moderate and weak turbulence of the FSO channel which are characterized by the kurtosis and the skewness can be easily classified by Fig 3 and 4 show plots of Kullback-Leibler distance. The proximity of the test distribution to the standard distribution can be seen from the graphs. The closer the test distribution is to the reference distribution, the Kullback-Leibler distance is very small in magnitude and approaches zero as the distribution approaches the reference distribution using RBFN Networks.

TABLE II. CLASSIFICATION OF THE GAMMA-GAMMA DISTRIBUTION FUNCTION

α	β	Nature of turbulence	Y
Goal 0.01			
5.49	1.12	Strong	45.6226
			-44.6226
4.39	2.56	Moderate	$3.7535 * 1e+003$
			$-3.7525 * 1e+003$
51.9	49.1	weak	3.4503
			-2.4502
Goal 0.02			
5.49	1.12	Strong	171.0052
			-171.0052
4.39	2.56	Moderate	$1.0e+003 * 1.4453$
			$1.0e+003 * -1.4443$
51.9	49.1	weak	0.4961
			0.5039

Plots in Fig 3 and 4 show the Kullback-Leibler distance estimated to match the text data with the known distribution. Fig 3 contains two plots of test data which are compared against the reference data and it can be observed that the Kullback-Leibler distance has large variation since the test data deviate from the reference data. However in Fig 4 where one of the test data confirm with the reference data, hence giving the Kullback-Leibler distance which is small in magnitude and very close to zero.

TABLE III. CLASSIFICATION OF THE K-DISTRIBUTION FUNCTION

a	Y
Goal 0.01	
0.1	$1.0e+005 * -1.6874$
	$1.0e+005 * 1.6875$
0.5	$1.0e+005 * -1.1082$
	$1.0e+005 * 1.1083$
2	$1.0e+003 * -2.2570$
	$1.0e+003 * 2.2580$
Goal 0.02	
0.1	$1.0e+005 * -1.5554$
	$1.0e+005 * 1.5554$
0.5	$1.0e+005 * -1.0336$
	$1.0e+005 * 1.0336$
2	-228.1035
	229.1035

TABLE IV. MISCLASSIFICATION RESULTS

α	β	Y
Goal 0.01, misclassified as Gamma-Gamma		
13.4955	32.4998	7.3473
		-6.3472
Goal 0.02, misclassified as Gamma-Gamma		
13.4955	32.4998	1.2754
		-0.2754

TABLE V. SAMPLE VALUES OF MEAN, KURTOSIS AND SKEWNESS FOR K DISTRIBUTION OBTAINED AFTER CLASSIFICATION

Distribution features	Y
K distribution, 400 neurons, MSE 0.052247, time 33.551908	
mean_K = 0.3191	0.0509
kurtosis_K = 2.8467	
skew_K = 1.0117	0.9483
mean_K = 0.3191	

TABLE VI. SAMPLE VALUES OF GAMMA-GAMMA DISTRIBUTION OBTAINED AFTER CLASSIFICATION

Distribution features	Y
Gamma-Gamma distribution, 400 neurons, MSE 0.0463962, time 38.023126 seconds	
16.9499	0.6312
30.6703	0.3688

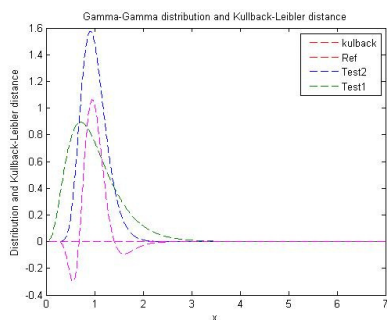


Fig. 3. Plot of test data , known fit and Kullback-Leibler distance

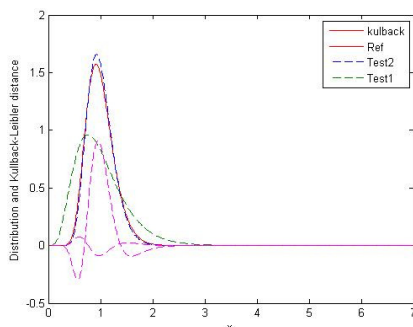


Fig. 4. Plot of test data , known fit and Kullback-Leibler distance

VI. CONCLUSIONS

The paper discusses the possible classification for selection of a most suitable channel distribution function using Radial Basis functions as classifiers. Computer simulations using Matlab have shown that with appropriate training, such classification has been possible with a specified mean square error. It was found that the Apparent Error rate was found to be 0.75 and the correct classification rate was 0.25., smaller value indicating better classification. Also the Kullback-Leibler distance was used to as a measure to estimate the proximity of the test data to the available reference data.

REFERENCES

[1] Nick Letzepis, Ian Holland and William Cowley, "The Gaussian Free space Optical MIMO channel with Q-ary Pulse Position Modulation", IEEE Transactions On Wireless Communications, May 2008, Vol. 7, No. 5, pp1744 – 1753

[2] Zeinab Hajjarian and Jarir Fadlullah, "MIMO Free Space Optical Communications in Turbid and Turbulent Atmosphere Invited Paper", Journal Of Communications, Sept 2009 Vol. 4, No. 8, pp 524 -532

[3] Deva K. Borah and David G. Voelz, , "Estimation of laser beam pointing parameters in the presence of atmospheric turbulence", Applied Optics , Aug 2007, Vol. 46, No. 23, pp 6010 -6018

[4] Hennes Henniger, Otakar Wilfert, , "An Introduction to Free-space Optical Communications", Radio Engineering, June 2010, Vol. 19, No.2, pp 203-212

[5] Zdenek. Kolka, Viera Biolková , Dalibor Biolek, "Simulation of FSO Transmission Channel", Proceedings of the 7th WSEAS

International Conference on Circuits Systems, Electronics, Control And Signal Processing (CSECS'08) , 2008 pp 186- 190

[6] Latsa Babu P and Balaji Srinivasan , "Characterizing the Atmospheric effects on Laser beam propagation for free space optical communication", 14th National Conference on Communications , NCC-2008 ,IIT Bombay , pp 332-334

[7] Ales Prokes, "Modelling of Atmospheric Turbulence effect on Terrestrial FSO Link", Radio Engineering ,2009 ,Vol 18, No.,1, pp 42- 47

[8] Lucie Dordov'a and Otakar Wilfert, "Laser Beam Attenuation Determined by the Method of Available Optical Power in Turbulent Atmosphere", Journal of Telecommunications and Information Technology, 2009, pp 53-57

[9] Albert D Wheelon,, "Electromagnetic Scintillation", Volume 2, Weak Scattering, Cambridge University Press, 2003(Online 2009),pp10-32

[10] Ahmed A. Farid, Steve Hranilovic , "Outage Capacity Optimization for Free-Space Optical Links With Pointing Errors", Journal Of Lightwave Technology, 2007,Vol. 25, N. 7, pp 1702-1710.

[11] Gradshteyn and I.M. Ryzhik, Table of Integrals, Series, and Products ,Seventh Edition, Elsevier Inc, 2007

[12] Edwirde Luiz Silva and Paulo Lisboa, "Analysis of the characteristic features of the density functions for gamma, Weibull and log-normal distributions through RBF network pruning with QLP", Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece, Feb 2007,pp 223-228

[13] Simon Haykin, Neural Networks, A comprehensive Foundation, Second edition, Pearson Prentice Hall, 2005.

[14] Mark J Orr , "Introduction to Radial Basis Function Networks", Center for cognitive Science, University of Edinburgh, Scotland, April 1966, pp1-67

[15] Harilaos G. Sandalidis, Theodoros A. Tsiftsis, George K.Karagiannidis,and Murat Uysal, , "BER Performance of FSO Links over Strong Atmospheric Turbulence Channels with Pointing Errors", IEEE Communications Letters, Jan 2008,Vol. 12, No. 1, pp 44-46

[16] <http://mathworld.wolfram.com> > Probability and Statistics > Moment

[17] Rudolf Kulhavy , "A Kullback-Leibler distance approach to system Identification, Institute of Information Theory and Automation" ,Academy of Sciences of the Czech Republic