RAST - 17

A NON-GAUSSIAN MODEL FOR INDIAN MONSOON RAINFALL

Kokila Ramesh[*1], R.N.Iyengar[2]
[*1] Assistant Professor, Department of Basic Sciences, Jain University, India
[2] Distinguished Professor, Center for Disaster Mitigation, Jain University, India

Abstract

A non-Gaussian model as a function of Gaussian process is developed in this paper for Indian monsoon rainfall time series. The functions of a Gaussian process are the Hermite polynomials. The unknown coefficients of the Hermite polynomials are found with the help of the first four moments of the given data. Since the probability density function of the Gaussian process is known, the non-Gaussian density function for the rainfall process is found by using the transformation on the known Gaussian density function numerically. Sample histogram of the data and the non-Gaussian density function are compared graphically along with the Gaussian density function. This clearly justifies that the non-Gaussian density better compares with the data distribution. This exercise has been done on the four broad regions of India identified by Indian Meteorological Department (IMD) and also for one subdivision of Karnataka. It has been observed that at 5% significance level, this model is able to reproduce the probability structure of the rainfall time series at different spatial scales studied.

Keywords: Non-Gaussian Model; Gaussian Process; Hermite Polynomials; Moment Estimation; Simulation.

1. Introduction

In many applications, we come across non-Gaussian time series such as rainfall, earthquake etc. Modeling of such time series will be of at most importance. Most commonly used assumption for this kind of series would be stationarity and Gaussianness. The advantage of this assumption will be using Gaussian model and its characteristics. Hence, when any time series is given it is assumed that the data is stationary and Gaussian to make further analysis simple. However with the nature of skewness and the kurtosis, it can be easily shown that the time series data is non-Gaussian. Thus the non-Gaussianess cannot be overlooked, if sometimes stationarity is. If a non-Gaussian model is to be proposed for the given time series, then it is preferable to completely

define it in terms of the first few important moments and the power spectral density function of the given time series. But this can be done only with Gaussian process. Hence the transformations on the Gaussian process can be considered as a non-Gaussian model. This chapter presents a non-Gaussian simulation technique developed during the course of this research. This simulation method is based on measured samples and known characteristics of the system input and output. First, static transformation techniques are developed to simulate non-Gaussian realizations by transforming the underlying Gaussian time or auto-correlation sample.

The target probability density function (PDF) is achieved by selecting a nonlinear static transformation, but unavoidably alters the power spectrum. The Correlation alteration methods appropriately identify the power spectral density (PSD) of the underlying Gaussian process; such that the desired PSD can be obtained using a proper nonlinear transformation along with the target PDF. The random fields were simulated numerically by Yamazaki and Shinozuka (1988) using the Gaussian PSD through iteration until the target PSD is achieved after transformation. A polynomial transformation of a Gaussian process was used by Ammon (1990). Using the same transformation he calculated back the Gaussian PSD based on the analytical relation between polynomial transformations on the power spectrum. Iyengar and Jaiswal (1993) have used a transformed process to develop a new non-Gaussian model for random excitations. They also found the level crossing and peak statistics of the non-Gaussian model and compared with the Gaussian model and observed data. This kind of level crossing detailed discussion of a non-Gaussian process, which is expressed in terms of a Gaussian process is studied by Grigoriu (1984). Gurley *et al* (1996) developed a non-Gaussian model using Hermite polynomial transformation through optimization of the parameters such as skewness and kurtosis. This is based on the minimization of sum of the squares of these two parameters as a condition to adjust the other parameters, which they called it as modified direct transformation. This method was shown to perform well compared to the regular direct transformation method. Sadek and Simiu (2002) identified that the gamma distribution and a normal distribution are appropriate for estimating the peaks corresponding to the longer and shorter tail of the time series respectively. The distribution of the peaks is then estimated by using the standard translation processes approach. A non-Gaussian peak factor analysis for univariate stationary non-Gaussian processes was discussed by Kwon and Kareem (2009). This was used for estimating expected positive and negative extremes of non-Gaussian processes of a time history of wind pressure fluctuations in the roof of a low-rise building. In all the above cases it is mentioned that the simulation of a non-Gaussian process can be achieved in terms of function of the Gaussian process. Therefore in the present study a non-Gaussian model to simulate the rainfall process using a Gaussian process is developed.

2. Rainfall Data

Six sets of SWM rainfall data are considered here for further work. The first is the All India rainfall value (AIRF) representing the whole country, which is spatial average based on the sub-regions. The other five data chosen are for the four sub-regions and one subdivision along the west coast with high variability. In Figure 1, the six regions are marked for clarity (Ref: Long range forecasting update, IMD, April 2013). The basic details of the data for the period (1901-2000) are shown in Table 1.

Table 1: Basic Statistics of Rainfall Data (1901-2000)

Name	Area (Sq Km)	LTA ($m_R$cm)	LTD ($\sigma_R$cm)	Skewness	Kurtosis
AIRF	32, 87,782	88.49	8.97	-0.22	2.67
CEIND	11, 06,673	96.41	12.89	0.03	3.13
NWIND	10, 10,719	61.60	11.53	0.14	3.18
NEIND	5, 24,065	143.97	12.29	-0.29	3.14
PEIND	6, 46,325	72.26	9.91	0.26	3.12
COKNT	18,717	289.19	46.80	0.66	4.92

AIRF-All India Rainfall CEIND–Central India NEIND-Northeast India NWIND-Northwest India PEIND-Peninsular India, COKNT-Coastal Karnataka.

The first five data series are from the official reports of the India Meteorology Department (IMD). The data for the Coastal Karnataka subdivision is taken from the data base of the Indian Institute of Tropical Meteorology (http://www.tropmet.res.in).
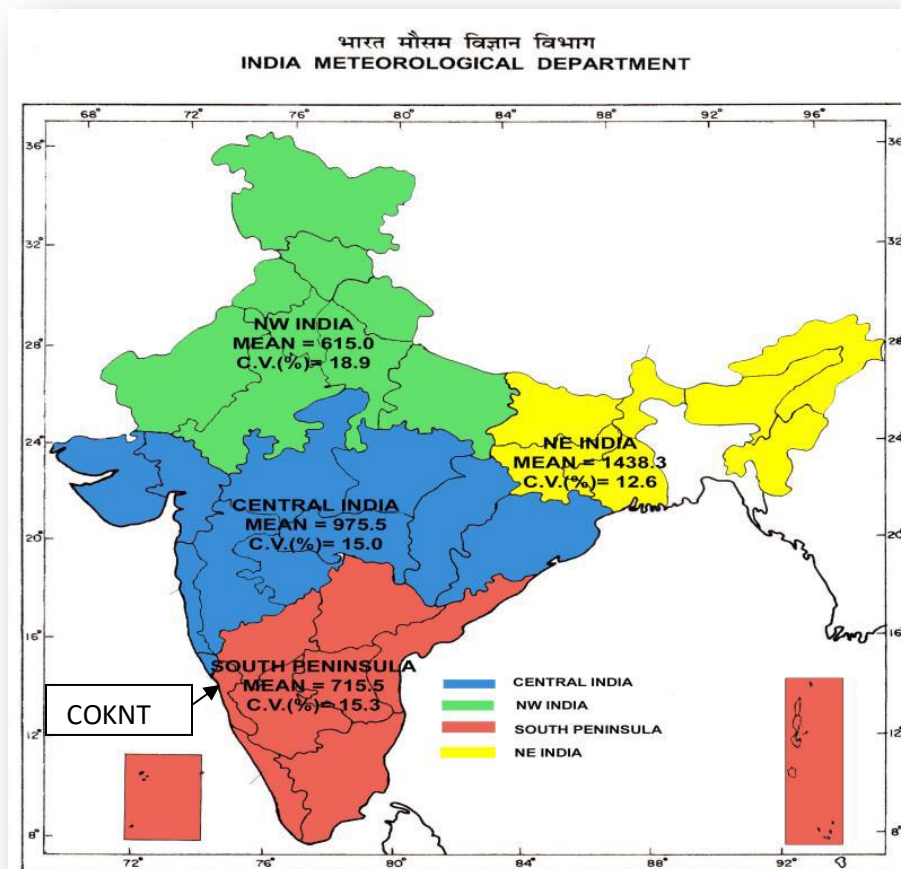

Figure 1: The four broad regions namely CEIND, NWIND, NEIND and PEIND of India and the one subdivision COKNT are marked in different colors

3. Methodology

A situation arises very often to model time series data which is non-Gaussian, but contains some properties of a Gaussian process. This can be modeled as a function of Gaussian process. Rainfall process is one such example. Hence in this chapter an attempt has been made to model Indian monsoon rainfall as a function of Gaussian process. It is observed that rainfall R_i, ($i = 1, 2, \ldots n$) as a random variable is non-Gaussian. But it can be approximately represented by a log-normal distribution. Hence it is advantageous to transform the data series R_i, for further analysis as

$$r_i = log\left(\frac{R_i}{m_R}\right) \tag{1}$$

The transformed series r_i still has some small mean in it along with some standard deviation. Therefore r_i is further standardized using the relation $x_i = \frac{(r_i - m_r)}{\sigma_r}$, where m_r is the long term average and σ_r the long term deviation of r to obtain zero mean and unit standard deviation process. This standardization is required for the model to develop as a function of zero mean, unit standard deviation Gaussian process in this chapter. The standardization always helps in finding the parameters involved in the non-Gaussian model used in an easy way. The basic statistics of x_i is presented in Table 2.

Table 2: Basic Statistics of normalized data x_i for the period 1901-2015

Name	LTA (m_x)	LTV (σ_x)	Skewness	Kurtosis	JB Statistic
AIRF	0	1	-0.4347	2.7487	3.92
CEIND	0	1	-0.4059	3.2621	3.52
NWIND	0	1	-0.6848	3.3838	9.69
NEIND	0	1	-0.3792	3.3427	3.31
PEIND	0	1	-0.3015	3.6038	3.26
COKNT	0	1	-0.1420	4.2659	7.99

Gaussianness of the random data x_i can be checked using skewness and kurtosis in the data. According to the Jarque-Bera (JB) test (Jarque and Bera 1987) the null hypothesis is that the sample comes from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution. For the null hypothesis H_0 accepting the data as Gaussian, the tabulated JB statistic at 90% confidence is 3.75. The calculated statistics listed in the above table exceed (3.74) in the case of AIRF, NWIND and COKNT, where as the other cases reject the null hypothesis at 85% significance level. Therefore even after normalization, the data cannot be treated as Gaussian. Hence a non-Gaussian method is required for modeling and simulation.

3.1. Modelling

In this chapter a new non-Gaussian model for rainfall process is proposed. Influenced by the work of Iyengar and Jaiswal (1993) $x(t)$ is expressed as a series with finite number of terms,

where the first term is a stationary Gaussian process with zero mean and unit standard deviation. The remaining terms would be the functions of the Gaussian process. The representation is as follows:

$$x(t) = a_0\varphi_0(t) + a_1\varphi_1(t) + a_2\varphi_2(t) + a_3\varphi_3(t) + a_4\varphi_4(t) + \qquad (2)$$

Here $\varphi_0 = 1$ and $\varphi_1 = u(t)$ which is a Gaussian random process with zero mean and unit standard deviation. If only the first two terms are selected in equation (2), then we have a Gaussian model with mean a_0 and variance a_1^2 with the auto-correlation function being $R_{xx}(\tau) = R_{uu}(\tau)$. But the data is non-Gaussian as shown earlier. Therefore in order to obtain non-Gaussianness in the process, the remaining φ's are selected as polynomials in $u(t)$ such that (φ_i, φ_j) $(i,j = 1,2,3 \dots n)$ are uncorrelated between each other. Hence these φ's are obtained using Gram-Schmidt orthogonalization procedure. The number of terms in equation (2) is decided based on the reflection of the first four moments. The reflection of the first four moments is possible if we consider four terms in the series. The following equations for φ's are considered as the polynomials in u to find the coefficients contained in them:

$$\varphi_0 = 1, \varphi_1 = u$$
$$\varphi_2 = C_{20}\varphi_0 + C_{21}\varphi_1 + C_{22}\varphi_2 u^2$$
$$\varphi_3 = C_{30}\varphi_0 + C_{31}\varphi_1 + C_{32}\varphi_2 u^2 + C_{33}\varphi_3 u^3 \qquad (3)$$

Using the orthogonality conditions of $\langle\varphi_i\varphi_j\rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$, and finding the unknowns in equation (3) we obtain the following equations for φ's

$$\varphi_2 = \frac{1}{\sqrt{2}}(u^2 - 1), \ \varphi_3 = \frac{1}{\sqrt{6}}(u^3 - 3u), \ \ \varphi_4 = \frac{1}{\sqrt{24}}(u^4 - 6u^2 + 3) \qquad (4)$$

It can be easily observed that φ's are the normalized Hermite polynomials which are given by

$$\varphi_k(u) = \frac{(-1)^k}{\sqrt{k!}}\exp\left(\frac{u^2}{2}\right)\frac{d^k}{du^k}\left[\exp\left(-\frac{u^2}{2}\right)\right] \qquad (5)$$

Since the data is standardized, the mean value is zero, therefore $a_0 = 0$ in equation (2). Substituting the values of φ's obtained in equation (4) in equation (2) we get

$$x(t) = a_0 + a_1 u(t) + \frac{a_2}{\sqrt{2}}(u^2(t) - 1) + \frac{a_3}{\sqrt{6}}(u^3(t) - 3u(t)) \qquad (6)$$

The variance of x can be used as one of the equations to find unknowns in equation (2), i.e

$$\langle x^2 \rangle = a_1^2 + a_2^2 + a_3^2 \qquad (7)$$

Similarly, we can use the higher order moments of x to find the remaining unknowns. Hence we get the following:

$$\langle x^3 \rangle = 2\sqrt{2}a_2^3 + 3\sqrt{2}a_1^2 a_2 + 9\sqrt{2}a_3^2 a_2 + 6\sqrt{3}a_1 a_2 a_3 + 12a_1 a_3 a_4 \qquad (8)$$
$$\langle x^4 \rangle = 3a_1^4 + 15a_2^4 + 93a_3^4 + 36\sqrt{6}a_3^3 a_1 + 4\sqrt{6}a_1^3 a_3 + 48\sqrt{6}a_2^2 a_1 a_3 + 30a_1^2 a_2^2 + 42a_1^2 a_3^2 +$$
$$186a_2^2 a_3^2 \qquad (9)$$

Suppose if the model requirement is of zero skewness and kurtosis$\neq$3, then a three term representation for x would have been sufficient. Also from the Gaussianness of u and the uncorrelatedness of φ's, the autocorrelation function of x reduces to

$$R_{xx}(\tau) = \sum_{i=1}^4 a_i^2 R_{uu}^i(\tau) \qquad (10)$$

Here a_i's are found using equations (7-9). Therefore if the one-dimensional statistics and the autocorrelation $R_{xx}(\tau)$ of the data are known then the modeling effort in the present approach reduces to finding the co-efficient a_i and $R_{uu}(\tau)$.

4. Results

For finding the values of a_i's the first four moments are used as shown in equations (7-9). It is observed that these equations are non-linear equations in a_i's. Therefore multiple roots for a_i's are possible which may be real or complex. The present approach is interested in only real roots of a_i's. All these real roots as possibilities are considered for modeling the PDF of the data. Hence the modeled non-Gaussian one dimensional PDF of x for all possible real roots is compared with the sample PDF to arrive at the final conclusion of selection of a_i's. The comparison of the non-Gaussian PDF with the sample PDF for all the possible real roots of the equation (7-9) is shown in Figure 1 for particular region. For example, for COKNT two sets of real roots are possible. They are:

Set I: $a_1 = 0.9955, \ a_2 = -0.0271, \ a_3 = 0.0906$
Set II: $a_1 = 0.8115, \ a_2 = -0.0645, \ a_3 = -0.5808$

For each of these sets of coefficients, the PDF of x is obtained by transforming the Gaussian PDF of u numerically. The sample histogram is also shown in Figure 2 as cross marks for comparison. From this comparison, the set of roots that follow the sample histogram is selected as the desired roots for modeling the data as non-Gaussian model. It is observed that Set I values follow the data distribution correctly, whereas the Set II roots show a sudden dip at the tails of the distribution. Therefore the distribution which is able to capture the extremes along with the other places is considered as the right model for the data. Hence the Set I roots are selected to model the data distribution and will be the right PDF for the observed data. Similarly the values of a_i's, which are suitable for the remaining regions are based on the sample histogram match with the PDF's obtained through the different set of roots of a_i's. The a_i's selected through this procedure using the first four moments for all the regions considered for the present study is shown in Table 3. Using the particular set of values of a_i's the desired non-Gaussian model is developed for the data. The visual comparison between the data distribution as the sample histogram and selected model for all the regions is shown in Figure 3. This comparison is shown along with the Gaussian distribution to show that this non-Gaussian model performs better than the Gaussian model. Specifically in the extremes or tails, the non-Gaussian model is able to capture the distribution better in comparison with the Gaussian model. Also it is observed that the PSD and the autocorrelation function of the estimated values can be exactly reproduced by solving equation (11).

Table 3: The values of a_1, a_2, a_3 using the first four moments of the data in equations (6-8)

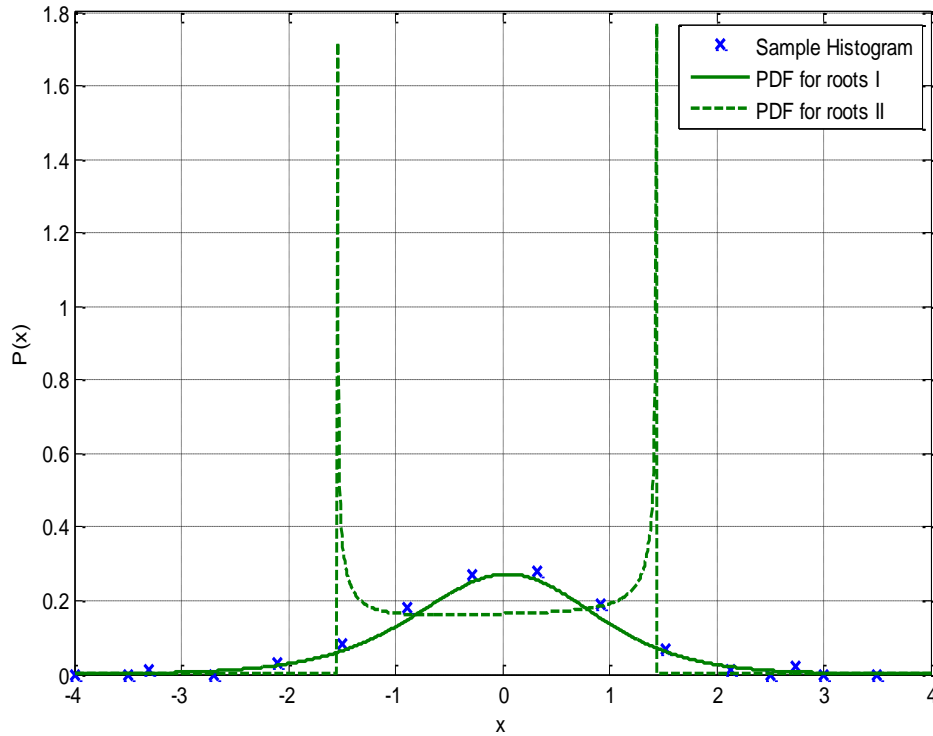| Name	AIRF	CEIND	NWIND	NEIND	PEIND	COKNT
a_1	0.9953	0.9958	0.9929	0.9962	0.9975	0.9955
a_2	-0.1113	-0.0806	-0.1718	-0.0865	-0.0541	-0.0271
a_3	-0.0207	-0.0237	-0.0015	0.0168	0.0449	0.0906

Figure 2: Sample histogram for the observed data of COKNT and the modeled solid line PDF for set I values and dotted line PDF for set II values

5. Discussion

In this paper, a new non-Gaussian model as a function of Gaussian process is developed for all India, its four regions and a subdivision from IMD. The functions of a Gaussian process are Hermite polynomials. The coefficients of the series are found using the first four moments of the standardized data also known as moment estimation method. The coefficients of the regions used for the present study is tabulated in Table 2. Using these coefficients with the help of Gaussian density, the non-Gaussian distribution obtained by using transformation numerically The comparison between the sample histogram of all the regions and subdivision and the non-Gaussian density function is shown in Figure 2. Along with this the Gaussian density also plotted to show that the non-Gaussian density performs better than the Gaussian density function, specifically at the extremes. This is clearly visible from the figure. For this kind of modeling the higher order moments such as skewness and kurtosis are important as it can the parameters to show that the probability distribution is Gaussian or not. Hence these are used to obtain the unknowns. It is clearly visible that the skewness and the kurtosis are matched in the figure shown.
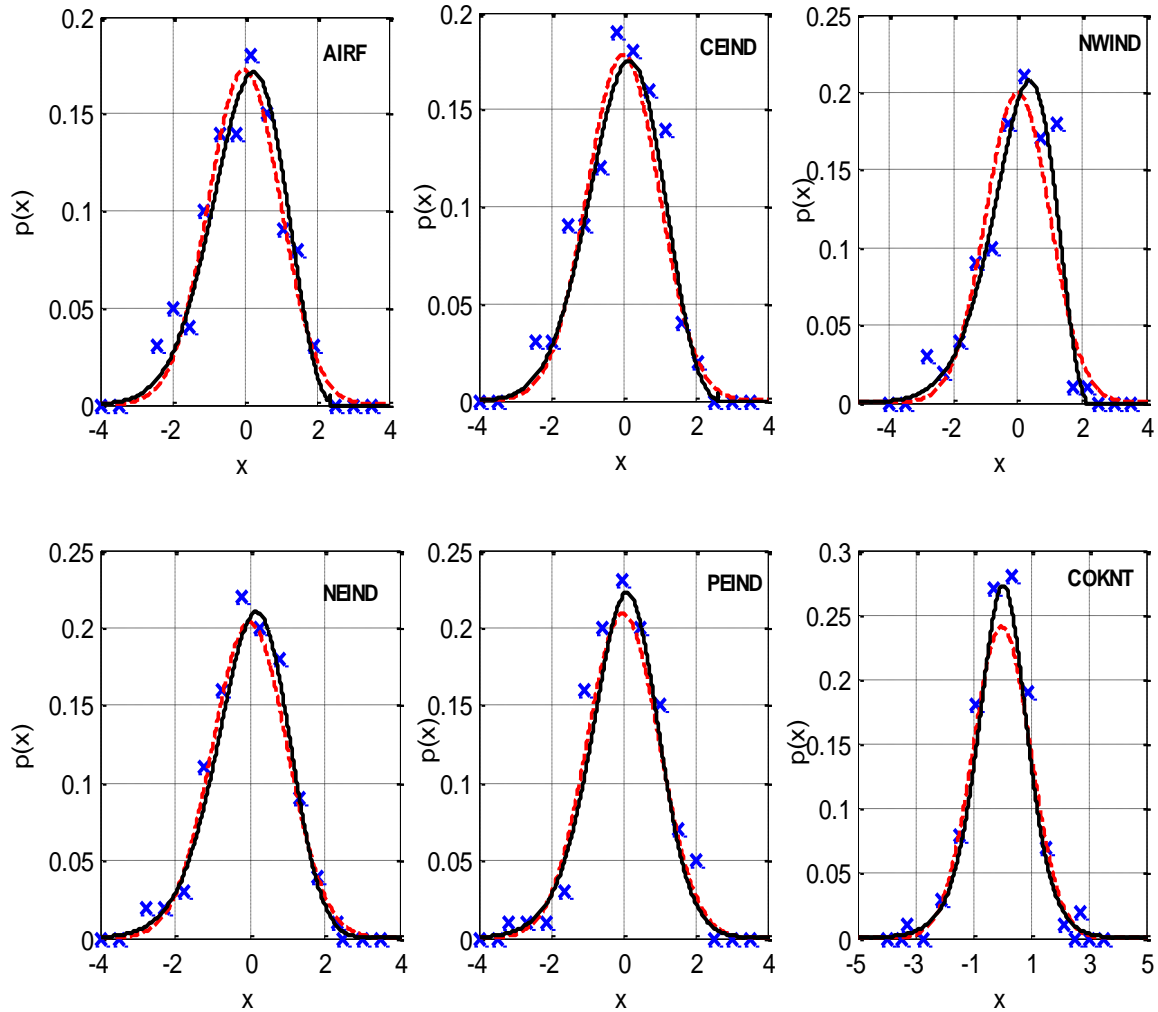
Figure 3: Comparison between the observed and the modeled PDF. x x x represents the sample histogram, dotted line is a Gaussian model and the solid line is a non-Gaussian model

6. Conclusions

The model developed for all India SWM rainfall data, its broad regions and a subdivision is a non-Gaussian model. This is expressed as function of Gaussian process, which are Hermite polynomials. The advantage of this model is that the properties of the Gaussian process can be used in identifying the properties of the non-Gaussian model. The first four moments of the data have been used to find the unknowns present in the non-Gaussian model. At 5% significance level this model fits to the rainfall data in comparison with the Gaussian model. This method is suggested for the smaller spatial scales such as seasonal data series, monthly data series and so on. It is expected that this model performs better in large size data which is obviously can be in smaller temporal scales.

References

[1] Yamazaki and Shinozuka (1988). "DIGITAL SIMULATION OF NON-GAUSSIAN STOCHASTIC FIELDS". J.Eng.Mech, ASEC, 114(7), 1183-1197

[2] Ammon, D. (1990). "APPROXIMATION AND GENERATION OF GAUSSIAN AND NON-GAUSSIAN STATIONARY PROCESS, Structural Safety, 8, 153-160

[3] Iyengar and Jaiswal (1993). "A NEW MODEL FOR NON-GAUSSIAN RANDOM EXCITATIONS", Journal of Probabilistic Engineering Mechanics, 8, 281-287

[4] Grigoriu (1984). "CROSSING OF NON-GAUSSIAN TRANSLATION PROCESS", j. Eng, Mech, ASCE, 110(4), 610-620.

[5] Gurley K, Kareem A, Tognarelli M A (1996). "SIMULATION OF A CLASS OF NON-NORMAL RANDOM PROCESS", International Journal of Nonlinear Mechanics, 31(5), 601-617

[6] Sadek and Simiu (2002). "PEAK NON-GAUSSIAN WIND EFFECTS FOR DATABASE-ASSISTED LOW-RISE BUILDING DESIGN", Journal of Engineering Mechanics, ASCE, 128(5), 530-539

[7] Kwon and Kareem (2009). "GUST-FRONT FACTOR: NEW FRAMEWORK FOR WIND LOAD EFFECTS ON STRUCTURES". Journal of Structural Engineering, ASEC 135, 717-732

*Corresponding author.

E-mail address: r.kokila@jainuniversity.ac.in