Gaussian Auto Regressive Model for Scat-Sat Data Set

Polaiah B[1], Sunitha M[2], Rajesh Anand B[3], Khadar Babu SK[4]

[1]Assistant Professor, Mallareddy University, Hyderabad

[2]Statistician cum Assistant Professor, Department of community Medicine, SVIMS, Tirupathi

[3]Department of Mathematics, Sri Venkateswara University, Tirupathi

[4]Associate Professor, VIT University, Vellore, TN, India
Corresponding Author Mail ID: *khadar.babu36[at]gmail.com*

Abstract: ***The Gaussian distribution plays a very important role in the statistical theory as well as methods. In this paper introduces the identification of the phenological stages of the rice crop located at Godavari region, Andhra Pradesh, India. The data taken from satellite images using ARCGIS platform and got the backscatter values. At present we make the following objectives for research work carried out. The major objective is to identify the phenological stages using graphical approach and to apply Gaussian probability approach for generation of the values using random Gaussian generator. Next step of the process is to fit an Auto-regressive model to predict and forecasting of the future observations and also obtain missing observations in the bulky data set. While fitting a Gaussian probability approach applied the method of moments and the maximum likelihood method of estimation for parameter estimation.***

Keywords: Probability distribution, frequency analysis, Gaussian distribution, Auto Regressive model etc

1. Introduction

The Gaussian distribution has a significant theoretical and practical impact on probability theory and mathematical statistics. The distribution can be found in a variety of sectors, such as high technology, natural phenomena, and industrial production. In general, a Gaussian distribution is an index that is influenced by numerous explanatory factors, each of which has a negligible impact. For example, the outcomes of quality indexes include tool size, fiber strength, and single feature groups like class vital capacity, plant strength of rice stem diameter in a particular place, recorded data on maximum air temperature, and average rainfall.The gaussian probability distribution's laws and regulations are being followed by all of the aforementioned indices. The phenological stages of the rice crop in the Godavari region of Andhra Pradesh, India, are identified in this research. The backscatter values were obtained from the data obtained from satellite photos using the ARCGIS platform. The main goal is to use a graphical approach to determine the phenological stages. The next step is to use the Gaussian probability technique and a random Gaussian generator to produce the values. The procedure then moves on to fitting an auto-regressive model to anticipate and predict future observations, as well as extract missing observations from a large amount of data. For time series data sets, the autoregressive model is a generator that fits the data the best.

2. Review of Literature

T. Mayooran and A laheetharan (2014) is to identify the best fit probability distribution of annual maximum rainfall flow time series in Colombo district for each period of study. In this paper for estimation of parameters they applied maximum likelihood method of estimation and goodness of fit tests were carried out In order to find the best fitting probability distribution among 45 probability distributions for annual rainfall flow time series for four seasons separately but the author cannot compare the parameters calculated by different methods.

Dr. Rafa H ALL-Suhili and Dr.Reza Khan Bilvardi (2014) studied about different frequency distribution models were fitted to the monthly rainfall data in solaimonia region north Iraq.

Haoge Liu and Jianhe (2016) studied about design of practical models of statistical events to optimize the Gaussian probability density function and to provide a significant method in statistics.

M Sunitha et al (2019) studied about novel research methodology to analyze the parameters of the big data sets such as rainfall flow time series, wind speed time series, hydrological datasets under fuzzy randomized approach.

Mahesh Palakuru et.al 2019, focused on the phenological parameters estimating using S map soil moisture active passive (S map), MODIS NDVI and Scat Sat – 1 scaterometer data and they adopted Gaussian distribution and two parameter Logistic distribution model for analysis.

3. Methods and Discussions

Let us consider the data, Zi, i=1,2, 3,…, n observations in a data set, for these values, by using the method of moments, to estimate the parameters of the probability distributions.

The Gaussian distribution plays a very important role in the statistical theory as well as methods. The great mathematician like Gauss, Laplace, Legendre and others are associated with the discovery and use of the distribution of errors of measurement.

Mathematically, a random variable X is said to have Gaussian distribution, if its probability density function is given by

$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-1}{2\sigma^2}(x-\xi)^2)$ -∞ < x < ∞; -∞ < μ < ∞, σ>0.

Where μ and σ are location and scale parameters respectively.

The probability density function of U = $(x-\xi)/\sigma$ is

$p_U(u) = (\sqrt{2\pi})^{-1}\exp(-\frac{1}{2}u^2)$,

Which does not depend on the parameters ξ,a. This is called the standard form of normal distribution. (it is also the standardized form.). the random variable U is called a standard, or unit, normal variable.

Since Pr[X≤x] = Pr[U ≤$\frac{x-\xi}{\sigma}$]

Such probabilities can be evaluated from tables of the cumulative distribution function of U, which is

$$\Phi(u) = Pr[U \le u] = (\sqrt{2\pi})^{-1}\int_{-\infty}^{u} e^{-\frac{x^2}{2}}dx\,.$$

The notation Φ(.) is widely used, further it is convent to have a systematic notation for the quantiles of the distribution of U. we use the system defined by

$$\Phi(U_\alpha) = \alpha$$

Method of Moments (MM)

If U has the unit Gaussian distribution, then, since the distribution is symmetrical about U= 0,

E(U) = 0,

and so

$$\mu_r = \mu_r^{|} = \boldsymbol{E}(\boldsymbol{U}^r) = (\sqrt{2\pi})^{-1}\int_{-\infty}^{\infty} x^r e^{-\frac{x^2}{2}}dx$$

If r is odd,

$$\mu_r = 0$$

If r is even,

$$\mu_r = (\sqrt{2/\pi})\int_{0}^{\infty} x^r e^{-\frac{x^2}{2}}dx$$

$$= (\sqrt{2/\pi})2^{\frac{(r+1)}{2}}\int_{0}^{\infty} t^{\frac{(r-1)}{2}}e^{-t}dt$$

$= 2^{r/2}\Gamma\left(\frac{1}{2}(r+1)\right)/\sqrt{\pi}$

= (r-1)(r-3) … 3.1

Hence

$$\text{Var}(U) = \mu_2 = 1$$

$$\alpha_3(U) = 0,$$

$$\beta_2(U) = \alpha_4(U) = 3$$

One of the simplest and oldest methods of estimation is the method of moments. The method of moments was discovered by Karl Pearson in 1894. It is a method of estimation of population parameters such as mean, variance etc. (which need not be moments), by equating sample moments with unobservable population moments and then solving those equations for the quantities to be estimated. The method of moments is special case when we need to estimate some known function of finite number of unknown moments. Suppose f (x;$\lambda_1,\lambda_2,\ldots,\lambda_p$) be the density function of the parent population with p parameters ($\lambda_1,\lambda_2,\ldots,\lambda_p$). Let $\mu^{|}_s$ be the s[th] moment of a random variable about origin and is given by

$\mu^{|}_s = \int_{-\infty}^{\infty} x^s\ f(x;\ \lambda_1,\lambda_2,\ldots\ldots,\lambda_p); r = 1,2,\ldots,p$

In general $(\mu_1', \mu_2', \ldots, \mu_p')$ will be the functions of parameters $(\lambda_1,\lambda_2,\ldots\ldots,\lambda_p)$.

Let (x_i; i=1,2, …, n) be a random sample of size n from the given population. The method of moments consists in solving the p-equations (i) for $(\lambda_1,\lambda_2,\ldots\ldots,\lambda_p)$ in terms of $(\mu_1', \mu_2', \ldots, \mu_p')$. Then replacing these moments $(\mu_s'\,; s = 1,2,3,\ldots,p)$ by the sample moments

e.g., $\hat{\lambda} i = \hat{\lambda}(\mu_1', \mu_2', \ldots, \mu_p') = \lambda_i(m_1', m_2', \ldots, m_p')$; i = 1,2, ..., p

where m_i is the i[th] moment about origin in the sample.

Then by the method of moments ($\hat{\lambda}_1,\hat{\lambda}_2, \ldots,\hat{\lambda}_p$) are the estimators respectively.

Autoregressive models can be effectively coupled with moving average models to form a general and useful class of time series models called autoregressive moving average models. However, they can only be used when the data are stationary. This differencing of the data series. For all three stages, they are identified using Anomaly approach for SCASAT data of Kharif season it counts approximately 100 observations. By using the method, it is divided into different stages sample values. It gives 16 observations for stage-I, 15 observations for stage – II, 13 observations for stage – III. For every stage applied Gaussian probability distribution for data simulation using R-studio and also for auto regressive model building using Excel sheet. The data analysis clearly given in the following tables and made different graphical methodology is given.

3.1 Rice Crop Phenology Stage II Statistical Analysis

It is the major stage in rice crop phenology called heading. In this stage seed yield buildup is constructed in this stage. Here needs continuous observation. For the data, calculated all normally distributed simulated values for the corresponding observations. Build a model for data at heading stage and analysis is given below.

Table 3.1: Auto regressive model building for stage – II analysis

Back scatter Values X_t	Estimated Y_t	$(X_t - \bar{X})$	$(X_t - \bar{X})^2$	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})^2$	$(X_t - \bar{X})(Y_t - \bar{Y})$	MA3 method
-9.4500	-9.7671	0.46360	0.21492	0.10270	0.01055	0.04761	---------
-9.5120	-9.9343	0.40160	0.16128	-0.06451	0.00416	-0.02591	-9.6027
-9.8460	-9.8173	0.06760	0.00457	0.05245	0.00275	0.00355	-9.5993
-9.4400	-9.5277	0.47360	0.22430	0.34211	0.11704	0.16202	-9.6780

-9.7480	-10.4949	0.16560	0.02742	-0.62511	0.39077	-0.10352	-9.6740
-9.8340	-9.6166	0.07960	0.00634	0.25317	0.06410	0.02015	-9.8060
-9.8360	-10.2140	0.07760	0.00602	-0.34424	0.11850	-0.02671	-9.7967
-9.7200	-9.9299	0.19360	0.03748	-0.06016	0.00362	-0.01165	-9.8453
-9.9800	-10.2240	-0.06640	0.00441	-0.35419	0.12545	0.02352	-9.7087
-9.4260	-9.9140	0.48760	0.23775	-0.04424	0.00196	-0.02157	-10.0907
-10.8660	-9.5178	-0.95240	0.90707	0.35195	0.12387	-0.33519	-10.4073
-10.9300	-10.5860	-1.01640	1.03307	-0.71625	0.51301	0.72799	-10.7053
-10.3200	-9.4738	-0.40640	0.16516	0.39593	0.15676	-0.16091	-10.3993
-9.9480	-9.6748	-0.03440	0.00118	0.19498	0.03802	-0.00671	-10.0387
-9.8480	-9.3543	0.06560	0.00430	0.51542	0.26566	0.03381	-------

Mean (X_t) = -9.9136 r_1 = 0.13467526
$\sigma_\epsilon^2 = 0.23179562$ $\sigma_\epsilon = 0.481451576$

Therefore AR (1) equation for the problem can be written as

$$X_t - \bar{X} = r_1(X_{t-1} - \bar{X}) + \sigma_\epsilon$$

The error σ_ϵ is also follows normal distribution with mean zero and variance 1.
(Law of large numbers)

$X_t - (-9.9136) = 0.13467526(X_{t-1} + 9.9136) + 0.481451576$
$X_t = 0.13467526(X_{t-1} + 9.9136) + 0.481451576 - 9.9136$
If $X_{t\_-1} = -9.8480$ then the predicted observation is
$X_{16} = 0.13467526(-9.8480 + 9.9136) + 0.481451576 - 9.9136$
$X_{16} =$ **-9.42331**

Table 3.3: t-Test: Paired Two Sample for Means

	Variable 1	*Variable 2*
Mean	-9.9136	-9.869766467
Variance	0.216805829	0.13830046
Observations	15	15
Pearson Correlation	0.13467526	
Hypothesized Mean Difference	0.043833533	
df	14	
t Stat	-0.611334275	
P(T<=t) one-tail	0.275388266	
t Critical one-tail	1.761310136	
P(T<=t) two-tail	0.550776532	
t Critical two-tail	2.144786688	

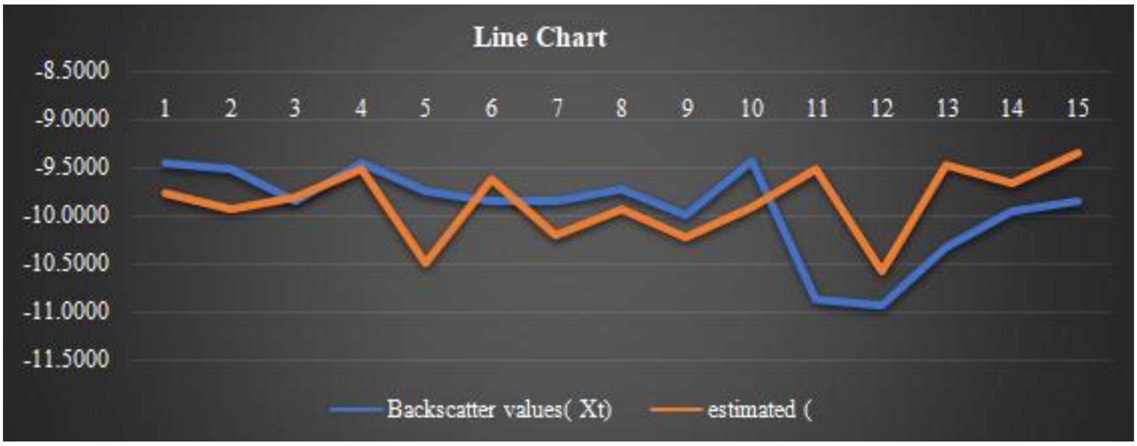

Figure 3.1: Backscatter frequency fluctuated lagged Line curve for Stage II

4. Conclusions

Gaussian probability distribution using Auto Regressive model is perfectly suitable for ScatSat datasets. We suggest that Gaussian probability distribution is exactly fitted for ScatSat data and also data simulation uses to build a auto regressive model for prediction and forecasting of the backscatter values of the ScatSat data. In this chapter we especially focus on regression model approach is also perfecty a novel application to generate, predict and forecasting of the vegetations at identified crop phenological stages in the present study area.

ScatSat datasets are the data given by the scatterometer fixed in satellites and the data called scatterometer data and it gives the backscatter values of the rice crop vegetations in identified study area. Finally, we conclude that the Gaussian probability random generated simulated datasets are perfectly applicable to the Scatterometer data

References

[1] Mehdi Khashei,Mehdi Bijari(2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Applied soft computing,Vol 11,issue 2,pp 2664-2675.

[2] Chin-Yu Lee(2005), Application of Rainfall Frequency Analysis on Studying Rainfall Distribution Characteristics of Chia-Nan PlainArea in Southern Taiwan, Crop, Environment &Bioinformatics,(2), pp 31-38.

[3] Swami Shivprasad, Anandrao Deshmukh, Ganesh Patil, Sagar Kahar (2013), Probability Distribution Analysis of Rainfall Data for Western Maharashtra Region, International Journal of Science and Research, 2319-7064.

[4] Oseni, B. Azeez and Femi J. Ayoola(2012), Fitting the Statistical Distribution for daily Rainfall in IBADAN, based on Chi- Squared and Kolmogrov-Smirnov Goodness of fit tests, European Journal of Business and Management,4(17),pp 62-70.

[5] T. Mayooran and A. Laheetharan (2014), The Statistical Distribution of Annual Maximum Rainfall in Colombo District, The Sri Lankan Journal of Applied Statistics,15(2)

[6] SK. Khadar Babu, Karthikeyan, M.V. Ramanaiah and D. Ramanaiah (2011), Prediction of rain-fall flow time series using auto- regressive models, Pelagia research library, Advances in applied science research, 2(2):128-133.

[7] M.Sunitha, B. Polaiah,S.K. KhadarBabu, K. Pushpanjali, M.V. Ramanaiah (2019), Normal Optimisation Technique for HydrologicalData Sets under Fuzzy Environment, International Journal of Pure and Applied Biosciences,7(6),264-269.

[8] S.R.Bhakar, Mohammed Iqbal, Mukesh Devanda, NeerajChhajed and Anil K. Bansal (2008), Probability Analysis of Rainfall at KOTA, Indian Journal of Agricultural Research,42(3),pp 201-206

[9] Chitrasen Lairenjam, Shivarani Huidrom, Arnab Bandyopadhyay and Aditi Bhadra (2016), Assessment of Probability Distribution of Rainfall of North East Region (NER) of India, Journal of Research in Environmental and Earth Science,2(9),pp12-18.