Normal Statistical Methodology for Scatsat Datasets at EG And WG, Andhra Pradesh, India

B. Polaiah[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 CommunityMedicine,SVIMS,Tirupathi
[3]Department of Mathematics,Sri Venkateswara University,Tirupathi,
[4]Associate Professor,Department of Mathematics,VIT University,Vellore,TN,India.

ABSTRACT
In the agricultural field choosing a probability distribution plays a important role for analyse the data. Actually, every distribution having the standard parameters, that can be useful to estimate the standard measures of the experiments. 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. Using the method of moments and take them to logistic distribution for data analysis.

Keywords: Gaussian distribution, frequency analysis, method of moments, Auto Regressive model etc.,

INTRODUCTION

Gaussian distribution plays an important role on mathematical statistics and probability theory both theoretically and practically. The distribution exists in different fields including industrial production natural phenomenon and high technology, etc., Generally Gaussian distribution is an index effected by many explanatory factors, but everyone has a tiny effect. For instance, the products of the quality indexes such as size of tools, strength of fibers and a single characteristics groups like vital capacity of people in a class, plant strength of stem diameter of rice in one area, the measured data of highest air temperature and average rainfall. All the above indexes are obeying the rules and regulations of the gaussian probability distribution. 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. The major objective is to identify the phenological stages using graphical approach. After that planning to apply Gaussian probability approach and generate 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 a bulky data set. The autoregressive model is a best fitted generator for time series data sets.

REVIEW OF LITERATURE:

T.Mayooran and A laheetharan(2014) is to identify the best fit probability distribution of annual maximum rainfall flow timeseries in Colombo district for each period of study. In this paper for estiation 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 timeseries 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 timeseries, wind speed timeseries, 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.

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

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(x - \xi)^2\right)$ $-\infty < x < \infty$; $-\infty < \mu < \infty$, $\sigma>0$.

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

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

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

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 I statistical Analysis

Actually, for the data, divided into different stages and did analysis for the scatsat data. In this case, first find the different parameters for the taken distributions applied different methods like moments procedure and MLE method for parameter estimation. After simplification of the two values, consider a value for mean and standard deviation are the parameters of the Gaussian probability approach and generated by simulation using R-studio for data analysis. For these data, the distribution is suitable or not using the t and Z statistical tests for testing the fitted probability distribution.

Table:3.1: Auto regressive model building for stage – I 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
-8.87	-9.18527	0.26360	0.06948	-0.03057	0.00093	-0.00806	---------
-8.90	-9.05599	0.22900	0.05244	0.09872	0.00974	0.02261	-8.90100
-8.94	-9.13560	0.19440	0.03779	0.01910	0.00036	0.00371	-8.93560
-8.97	-9.23396	0.15980	0.02554	-0.07926	0.00628	-0.01267	-8.97020
-9.00	-9.42863	0.12520	0.01568	-0.27393	0.07504	-0.03430	-9.00480
-9.04	-9.01484	0.09060	0.00821	0.13986	0.01956	0.01267	-9.03940
-9.07	-9.18427	0.05600	0.00314	-0.02956	0.00087	-0.00166	-9.07400
-9.11	-9.23492	0.02140	0.00046	-0.08022	0.00644	-0.00172	-9.10860
-9.14	-9.05898	-0.01320	0.00017	0.09572	0.00916	-0.00126	-9.14320
-9.18	-9.31826	-0.04780	0.00228	-0.16356	0.02675	0.00782	-9.17780
-9.21	-8.99868	-0.08240	0.00679	0.15602	0.02434	-0.01286	-9.21240
-9.25	-9.30742	-0.11700	0.01369	-0.15272	0.02332	0.01787	-9.24700
-9.28	-9.17719	-0.15160	0.02298	-0.02249	0.00051	0.00341	-9.28160
-9.32	-9.19736	-0.18620	0.03467	-0.04266	0.00182	0.00794	-9.31620
-9.35	-9.11450	-0.22080	0.04875	0.04020	0.00162	-0.00888	-9.35080
-9.39	-8.82865	-0.25540	0.06523	0.32605	0.10631	-0.08327	--------

The first order markovian model is also like first order auto regressive approach, in this the data follows gaussian distribution with mean zero and variance 1. Then the mean and variance of the data calculated using the method of MLE. Then obtained auto correlation coefficient for the taken backscatter data set. By using the formulas given in equations are applied.

Then we get the following parameters for model building.

Mean (Xt) = -9.1259 r_1 = -0.248199218

σ_ϵ^2 = 0.027563388 σ_ϵ = 0.166022251

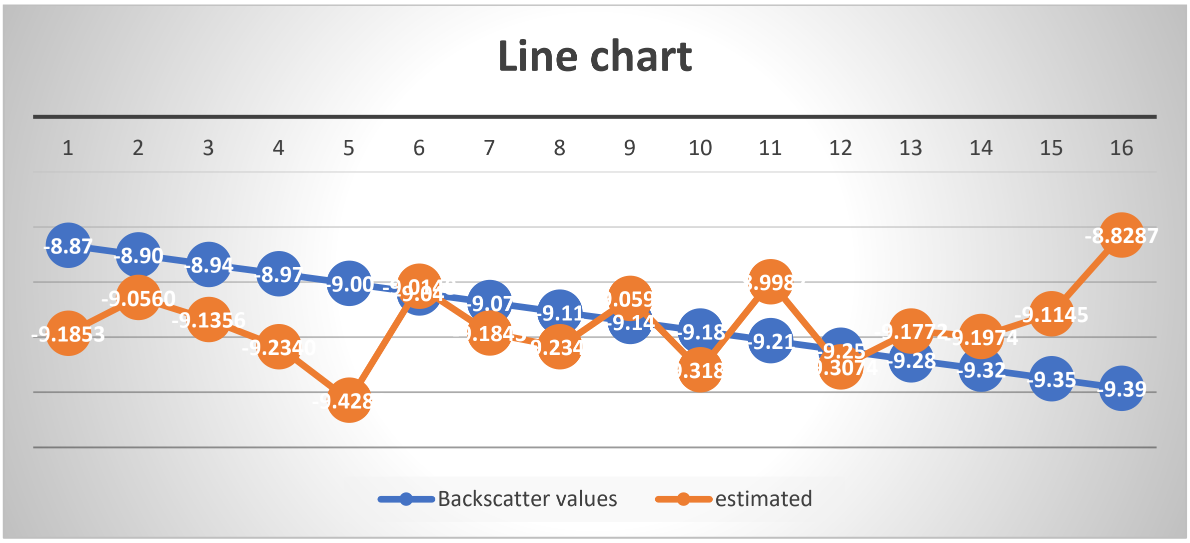
Therefore AR (1) equation for the problem can be written as

Xt - (-9.1259) = -0.248199218(Xt-1+9.1259) +0.66022251 Xt= -0.248199218(-9.39+9.1259) +0.66022251+9.13 X17 = **-8.89547**

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

	Variable 1	*Variable 2*
Mean	-9.1259	-9.154656438
Variance	0.027135627	0.02087115
Observations	16	16
Pearson Correlation	-0.248289225	
Hypothesized Mean Difference	0.03	
Df	15	
t Stat	-0.020337068	
P(T<=t) one-tail	0.492021274	
t Critical one-tail	1.753050356	
P(T<=t) two-tail	0.984042547	
t Critical two-tail	2.131449546	

Fig 3.1: Backscatter frequency fluctuated lagged curve for Stage I



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 perfectly 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. ChitrasenLairenjam, ShivaraniHuidrom, 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.
2. Chin-Yu Lee(2005), Application of Rainfall Frequency Analysis on Studying Rainfall Distribution Characteristics of Chia-Nan Plain Area 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. M.Sunitha, B. Polaiah,S.K. KhadarBabu, K. Pushpanjali, M.V. Ramanaiah (2019), Normal Optimisation Technique for Hydrological Data Sets under Fuzzy Environment, International Journal of Pure and Applied Biosciences,7(6),264-269.
5. 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.
6. 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)
7. S.R. Bhakar, Mohammed Iqbal, Mukesh Devanda, Neeraj Chhajed and Anil K. Bansal(2008), Probability Analysis of Rainfall at KOTA, Indian Journal of Agricultural Research,42(3),pp 201-206