Bayesian Change-Point Modelling of Rainfall Distributions In Nigeria

W. B. Yahya[1]; K. O. Obisesan[2]; T. M. Adegoke[3]

[1]Department of Statistics, University of Ilorin,
Ilorin, Nigeria.
e-mail: wbyahya@unilorin.edu.ng

[2]Department of Statistics, University of Ibadan,
Ibadan, Nigeria.
e-mail :Obidairo@gmail.com

[3]Shopping Internet Services,
Lagos, Nigeria.
e-mail: adegoketaiwom@gmail.com

***Abstract*—A Bayesian framework is developed to detect single change abrupt shift in a time series of the annual amount of rainfall in Nigeria. The annual amount of rainfall is modelled by a Normal probability distribution where the means are codified by a normal probability distribution and inverted gamma probability distribution for the variance. Based on the sampling from an estimated informative prior for the parameters and the posterior distribution of hypotheses, the methodology is applied to the time series of amount of rainfall in six states in Nigeria. Although, the model under study seems quite simple, but no analytic solutions for parameter inference are available, and recourse to approximations is needed. It was shown that the Gibbs sampler is particularly suitable for change-point analysis, and this Markovian updating scheme is used. The result from the analysis showed that, in all the six states considered displayed that indeed a single change point occurred.**

Keywords: *Change point analysis, Bayesian method, change in mean level, inverted gamma distribution Gibbs Sampling, Posterior Distribution*

I. INTRODUCTION

The investigation of the change point in time ordered data started in the 1950s [5]. Change point problems originally arise from quality control or reliability. Nowadays, these problems cover almost all areas including environmental sciences, hydrology, signal processing, biology, climatology, economics, etc. There are considerable amounts of work on change point problems and related topics and these have appeared in the literature. Change point problems have generally been solved by the maximum likelihood methodology and the Bayesian procedure based on the parametric method, the nonparametric method, and the decision theoretic method. Our primary interest is restricted to the parametric Bayesian method applicable to change point problems in a sequence of independent normal observations.

II. MATERIALS AND METHODS

A. Data Description

The data for this research work were collected from the Nigeria Meteorological Agency (NIMET) on rainfall distribution in Nigeria. Rainfall data for six states were collected which spanned different years due to their deferring years of creation. Therefore, for each state, the on-set of the data collected started from the year the state was created.

The rainfall data of 45 years (1971 - 2015) for the three states of Lagos, Zamfara and Taraba, 35 years (1981 – 2015) data for Akwa-Ibom Stateand FCT and 43 years (1973 - 2015) data for Anambra State were used.

B. Methodology

In this research, the change point analysis focuses on the detection of change in the mean level of the time series data. Several classical methods in change point analysis include non-parametric Wilcoxon test, the Student-t test, Maximum likelihood, and the sequential Mann-Kendall test. The alternative for these classical approaches is Bayesian approach which takes into consideration prior information, the model of the shift assumed and observed data into forming posterior distribution to model the data into forming a posterior distribution to model associated analysis.

Bayesian approach in change identification problem for mean level in time series data have been used by previous researchers such as [2, 3, 4, 5, 6, 7 8,910,11, 12,13]. The Bayesian approach in this study is based on single shifting model distributions on the unknown change point.

In contrast to the classical approach, the Bayesian methods take into consideration the parameter of the model as random variables represented by a statistical distribution (prior distribution) rather fixed values. The Bayesian methods allow the integration of statistical analysis through the prior distribution with the most current information based on the observations into a posterior distribution. In other words, the prior to experimentation; the posterior distribution an updated belief about the parameters after sample data is obtained. The analysis involved changes getting the mean value before and after the change, the amount of the change and the variation in observations. In this study, two related problems will be analyzed that is on the detection of the change point and the estimation of the change point.

Suppose that there is a sequence of independent normal random variables, $y_1, y_2, \ldots, y_n$. These are observed along with time. This sequence is said to have a change at a time point γ *often called a change point if* $\mu_1 \neq \mu_2$.

$$Y_i \sim \begin{cases} N(\mu_1, \sigma^2), & i = 1, 2, \cdots \gamma \\ N(\mu_2, \sigma^2), & i = \gamma+1, \gamma+2 \cdots T \end{cases} \qquad (M1)$$

where $N(\mu, \sigma^2)$ *represents a Normal distribution with the density function given in Eq.2:*

$$f(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\sigma}} exp\left\{-\frac{(y-\mu_r)^2}{2\sigma^2}\right\}, x \in \mathbb{R} \qquad (1)$$

The parameters μ_1, μ_2 *and* σ^2 represent the change point, the mean before and after the shift and the variance of the series, respectively. The prior distribution of μ_1 *and* μ_2 assumed to be the same Normal distribution, denoted in Eq. 2:

$$\pi(\theta) = \pi(\sigma^2)\pi(\mu_1|\sigma^2)\pi(\mu_2|\sigma^2)\pi(\gamma|M1)$$
$$= \begin{cases} \mu_1|\sigma^2 \sim N(\psi_1, \sigma^2\kappa_1) \\ \mu_2|\sigma^2 \sim N(\psi_2, \sigma^2\kappa_2) \\ \sigma^2 \sim IG(\frac{v_0}{2}, \frac{v_{0\sigma_0^2}}{2}) \end{cases} \qquad (2)$$

All the studies on change point problems are generally divided into two parts. The first part is to detect the existence of changes, that is, to test the no-change model by Eq. 3

$$Y_i \sim N(\mu_1, \sigma^2), \qquad i = 1,2 \cdots T \qquad (3)$$

against the change model (1) using the Bayes factor or the posterior probability'.

The likelihood function resulting from T observations $y = (y_1, y_2, \cdots, y_T)$ generated model (M1) can be writing by

$$p(y|\mu_1, \mu_2, \sigma^2) = \prod_{i=1}^{\gamma} N(\mu_1, \sigma^2) \prod_{i=\gamma+1}^{T} N(\mu_2, \sigma^2)$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\left(\frac{T}{2}\right)} exp\left\{-\frac{\gamma}{2\sigma^2}\left[s^2{}_1 + (\bar{y}_\gamma - \mu_1)^2\right]\right\}$$
$$\times exp\left\{-\frac{T-\gamma}{2\sigma^2}\left[s^2{}_2 + (\bar{y}_{T-\gamma} - \mu_2)^2\right]\right\} \qquad (4)$$

Where $\bar{y}_\gamma = \sum_{i=1}^{\gamma} \frac{y_i}{\gamma}$, $\quad \bar{y}_\gamma = \sum_{i=\gamma+1}^{T} \frac{y_i}{T-\gamma}$

$$S_1^2 = \sum_{i=1}^{\gamma} \frac{(y_i - \bar{y}_\gamma)^2}{\gamma}, \qquad S_1^2 = \sum_{i=\gamma+1}^{T} \frac{(y_i - \bar{y}_{T-\gamma})^2}{T-\gamma}$$

Assuming for model $M1$ prior independence between γ and the
other parameters $(\mu_1, \mu_2, \sigma^2)$and that $\pi(\gamma|M1)$ is any discrete distribution on the set {1,2,...,T-1}. Because of conjugate properties [1], under $M1$, the conditional joint posterior distribution$p(\mu_1, \mu_2, \sigma^2|\gamma, y, M_1)$ given γ and the observed data y also belongs to the class of normal-inverted gamma distributions, but with updated parameters$\left(\psi_1', \psi_2', \kappa_1', \kappa_2', \frac{v_n}{2}, \frac{v_n\sigma_n^2}{2}\right)$.More precisely,

$$p(\mu_1, \mu_2, \sigma^2|\gamma, y, M_1) = NNIG\left(\psi_1', \psi_2', \kappa_1', \kappa_2', \frac{v_n}{2}, \frac{v_n\sigma_n^2}{2}\right) \qquad (5)$$

Where,

$$\kappa_1' = \frac{\kappa_1}{(1+\gamma\kappa_1)}, \qquad \psi_1' = (1-\kappa_1'\gamma)\psi_1 + \kappa_1'\gamma\bar{y}_\gamma$$

$$\kappa_2' = \frac{\kappa_2}{(1+(T-\gamma)\kappa_2)}, \; \psi_2' = (1-\kappa_2'(T-\gamma))\psi_2 + \kappa_2'(T-\gamma)\bar{y}_{T-\gamma}$$

$$\sigma_n^2 = \frac{1}{v_n}(v_0\sigma_0^2 + (\gamma-1)S_1^2 + (1-\gamma\kappa_1')(\bar{y}_\gamma - \psi_1)^2 + (T-\gamma-1)S_2^2 + (1-(T-\gamma)\kappa_2') + (\bar{y}_{T-\gamma} - \psi_2)^2)$$

$$v_n = v_0 + T$$

The prior predictive density can be expressed as

$$p(y|\gamma, M_1) = \left(\frac{1}{2\pi}\right)^{\frac{T}{2}} \sqrt{\frac{\kappa_1'\kappa_2'}{\kappa_1\kappa_2}} \frac{\left(\frac{v_{0\sigma_0^2}}{2}\right)^{\frac{v_0}{2}} \Gamma\left(\frac{v_n}{2}\right)}{\left(\frac{v_{n\sigma_n^2}}{2}\right)^{\frac{v_n}{2}} \Gamma\left(\frac{v_0}{2}\right)} \qquad (6)$$

In our problem, γ is unknown and its marginal posterior distribution $p(\gamma|y, M_1)$has to be derived. Using Bayes theorem and the prior predictive density Eq. (6), the marginal posterior density of the change point $\gamma = 1, 2, \cdots T-1$ under model M_1 is seen to be

$$p(\gamma|y, M_1) = \frac{p(y|\gamma, M_1)p(\gamma|M_1)}{\sum_{\gamma=1}^{T-1} p(y|\gamma, M_1)p(\gamma|M_1)}$$

$$\propto p(\gamma|M_1)\sqrt{\kappa_1'\kappa_2'}\left(\frac{v_n\sigma_n^2}{2}\right)^{\frac{v_n}{2}} \qquad (7)$$

This distribution is discrete and gives, for each time point, the posterior probability of shift occurrence in the mean level assuming a change occurred with certainty.

The calculation in this procedure may not be expressed in a simple form but it can be estimated by using Monte Carlo Markov Chain approach.

III. Results

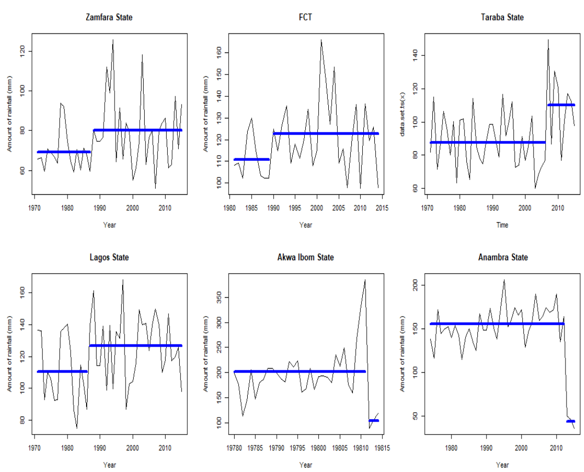

Fig 1: Change Point Detection in amount of rainfall for some states

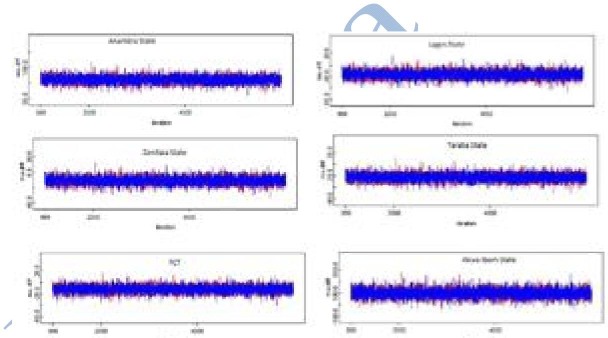

Fig 2:: History plot for change point detection for the selected states.

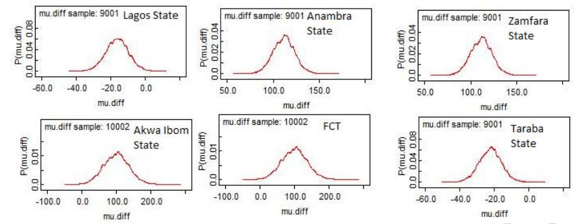

Fig 3: Posterior density distribution for change point detection for the selected states.

Fig 1 presents the change-point plot for six states (Zamfara, FCT, Lagos, Taraba, Akwa Ibom, and Anambra) under study. Table 1 gives the summary statistics of amount of rainfall in six states and Tables 1 and 2 shows the summary statistics of the set of data before and after the change point was detected.

Fig 4 shows a box plot of the amount of rainfall in the six states, before and after the detection of change-point. It was observed that the overall median for Akwa Ibom state and Taraba sate are very close to the overall median.

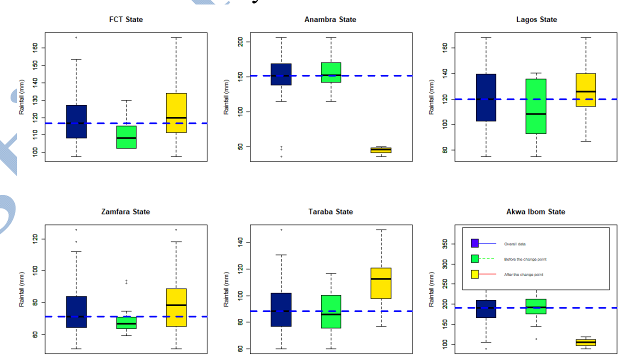

Fig 4: Box plot for all the data set, before and after the change point detection.

Table 1: Summary Statistics of the overall Rainfall data for the Six selected states.

State	Duration	Min	Q1	Q2	Mean	Q3	Max
Zamfara	1971-2015	50.9	64.3	71.3	76.0	83.8	125.6
FCT	1981-2015	102.3	102.4	108.1	110.7	115.3	129.9
Taraba	1971-2015	59.9	76.9	88.5	92.1	102	135.7
Anambra	1973-2015	35.9	138.6	151.	147.	168.5	205.9
Akwa-Ibom	1981-2015	88.7	166.8	191.5	193.8	210.3	385
Lagos	1971-2015	74.6	102.5	119.6	120.8	139.6	168.3

Table 2: Summary Statistics of the Rainfall data for the Six selected states before the change occurred.

State	Duration	Min	Q1	Q2	Mean	Q3	Max
Zamfara	1972–1987	59.1	64.4	67.2	69.7	71.0	93.8
FCT	1981–1988	102.3	102.4	108.1	110.7	115.3	129.9
Taraba	1971–2006	59.9	75.9	86.1	87.6	100.2	116.5
Anambra	1973–2012	114.9	141.9	152.2	155.2	170.2	205.9
Akwa Ibom	1981–2012	113.9	175.4	192.6	202.2	212.1	385.0
Lagos	1971–1986	74.6	92.7	108.1	110.5	135.8	140.4

Table 3: Summary Statistics of the Rainfall data for the Six selected states after the change occurred.

State	Duration	Min	Q1	Q2	Mean	Q3	Max
Zamfara	1998-2015	50.9	64.3	76.7	79.4	86.3	125.6
FCT	1989-2015	97.5	111.4	119.8	122.7	134.1	166.0
Taraba	2007-2015	76.9	97.9	112.4	110.3	120.9	149.5
Anambra	2013-2015	35.9	41.2	46.5	44.2	48.4	50.2
Akwa-Ibom	2013-2015	35.9	41.2	46.5	44.20	48.35	50.20
Lagos	1987-2015	88.7	97.3	105.8	122.5	112.4	138.9

Table 4: Bayesian quantities for Akwa Ibom State

	Posterior Mean	95% Posterior Interval	MC_Error
μ_1	187.3	[165.9, 206.4]	0.04
μ_2	56.95	[-4.019,113.6]	0.17
γ	130.3	[70.84, 111.3]	0.17
σ^2	3359	[1990.0,5594.0]	10.08

Table 5: Bayesian quantities for Anambra State

	Posterior Mean	95% Posterior Interval	MC_Error
μ_1	154.2	[147.9 , 160.4]	0.03
μ_2	40.93	[18.52 , 62.56]	0.106
γ	113.3	[90.9 ,136.2]	0.107
σ^2	393.5	[249.8 , 605.6]	0.96

Table 6: Bayesian quantities for FCT

	Posterior Mean	95% Posterior Interval	MC_Error
μ_1	108.8	[98.43 , 119]	0.05
μ_2	123.0	[116.4, 129.2]	0.03
γ	-14.14	[-26.19,1.958]	0.00
σ^2	108.8	[98.43 , 119]	0.05

Table 7: Bayesian quantities forLagos State

	Posterior Mean	95% Posterior Interval	MC_Error
μ_1	108.2	[76.77, 119.4]	0.06
μ_2	124.5	[113.5, 134.7]	0.04
γ	-16.24	[-31.69,-0897]	0.13
σ^2	531.2	[319.9, 860.4]	0.85

Table 8: Bayesian quantities forTaraba State

	Posterior Mean	95% Posterior Interval	MC_Error
μ_1	87.06	[81.32, 92.79]	0.03
μ_2	108	[96.24, 119.4]	0.06
γ	-20.99	[-33.72,8.016]	0.06
σ^2	306.4	[198.5,464.8]	0.73

Table 9: Bayesian quantities for Zamfara State

	Posterior Mean	95% Posterior Interval	MC_Error
μ_1	68.43	[60.53, 76.26]	0.0413
μ_2	81.21	[73.32, 88.76]	0.036
γ	-12.78	[-23.87,1.752]	0.054
σ^2	271.8	[164.6, 438.7]	0.715

IV. Discussions

The Bayesian method presented in this paper can be viewed as an extension of the normal models and practitioners can perform such change-point analysis routinely using standard statistical toolboxes. This approach can be generalized to other type of normal models and even to other types of p.d.f.s . More precisely, use of probability distributions which belong to the exponential class of p.d.f.s allows for exactly the same line of reasoning based on conjugacy.

From the results obtained from the analysis performed, Figures 1 showed the evidence that indeed a change actually occur in all the six states considered (Zamfara, FCT, Taraba, Anambra, Akwa Ibom, and Lagos) in different years (1987, 1988,1886, 2012, 2012 and 1985 respectively). The Tables 4 – 9 provided the summary of the quantity for each parameters for the six states using R programming language. This includes the means, differences in mean, variance and the MC error. We also observed that there is a decrease in the amount of rainfall in states like Zamfara, Taraba, Lagos and FCT and whereas states like Akwa Ibom and Anambra experienced an increase in the amount of rainfall.

V. Conclusion

Hydrologists can take full advantage of the existence of conjugate distributions when studying a single change-point in a statistical model belonging to the exponential family. If prior independence between the epoch of change and model parameters is assumed, the joint posterior distribution is a finite mixture of conjugate distributions.

This allows for easy computation of posterior odds, as illustrated by the univariate normal model under the configuration of a single change in the mean level.It was showed that more complex change-point problems can be readily addressed by Gibbs sampling, which is easily accessible to the average statistical practitioner.

As a whole, states like Akwa-Ibom and Anambra displays more significant changes in trend compared to other states. The increase in significant trend at Akwa-Ibom and Anambra for extreme cumulative rainfall amount, extreme intensities and extreme frequency need to be viewed with caution.

ACKNOWLEDGMENT

The authors were grateful to the anonymous reviewers whose comments and suggestions were valuable to improve the exposition of the paper.

REFERENCES

[1] Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, Springer, New York.

[2] Booth, N.B. and Smith, A.F.M. (1982). A Bayesian approach to retrospective identification of changepoints. *Journal of Economics,* 19: 7-22. DOI: 10.1016/0304- 4076(82)90048-3

[3] Chen, J. and Gupta, A. (1997). Testing and locating variance change points with application to stock prices. *Journal of American Statistical Association: JASA*, 92(438):719-747.

[4] Chen, J. and Gupta, A. (1999). Change point analysis of a Gaussian mode. *Statistical Paper*, 40:323-333.

[5] Chen, J. and Gupta, A. (2004). Statistical inference of covariance change points in Gaussian model. *Statistics*, 38:17-28.

[6] Chen, J. and Gupta, A. (2007). A bayesian approach to the statistical analysis of a smooth abrupt change point analysis. *Advance and Applications in Statistics*, 7:115-125.

[7] Hinkley, D. and Hinkley, E. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57:1{17.

[8] Kim, C., Suh, M. S. and Hong, K.O. (2009). Bayesian changepoint analysis of the annual maximum of daily and sub-daily precipitation over South Korea. *Journal of Climate*, 22: 6741-6757. DOI:10.1175/2009JCLI2800.1

[9] Lee, A.S.F. and S.M. Heghinian, 1977. A shift of the mean level in a sequence of independent normal random variables: A Bayesian approach. *Technometrics,* 19: 503-506. DOI:10.2307/1267892

[10] Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523-527.

[11] Perreault, L., Bernier, J. Bobee, and Parent, E. (2000a). Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrolology*, 235: 221-241. DOI: 10.1016/S0022-1694(00)00270-5

[12] Perreault, L., Bernier, J. Bobee, B. and Parent, E. (2000b). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrol*ology, 235: 242-263. DOI: 10.1016/S0022-1694(00)00271-7.

[13] Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika,* 62: 407-416. DOI: 10.1093/biomet/62.2.40