DOI: http://doi.org/10.4038/sljastats.v23i1.8041

Sri Lankan Journal of Applied Statistics, Vol (23-1)

Modelling Rare Events in an Adaptive Cluster Sampling Design with Heterogeneity among Networks and within the Network Units

O. A. Wale-Orojo[1*], O. M. Olayiwola[1], F. S. Apantaku[1], I. T. Omoniyi[2], and A. O. Ajayi[3]

[1]Department of Statistics, Federal University of Agriculture Abeokuta, Ogun State, Nigeria.

[2]Department of Aquaculture and Fisheries Management, College of Environmental Resources Management, Federal University of Agriculture Abeokuta, Ogun State, Nigeria

[3]Department of Statistics, Federal University of Agriculture Abeokuta, Ogun State, Nigeria.

*Corresponding author: seunorojo@gmail.com

Received: 24^{th} April 2021/ Revised: 05^{th} March 2022/ Published: 31^{st} August 2022

ABSTRACT

Rare events population (φ) is hard-to-reach, sparsely distributed and clustered; an Adaptive Cluster Sampling (ACS) is the design to collect information from φ. Researchers and Policy Makers have modelled φ in ACS design with homogeneity assumptions. This study modelled φ with heterogeneity among networks and within the network units. Data from the International Institute of Tropical Agriculture on Culcasia Scandens, an understory plant and simulation were used to validate the model. Estimators for total and average number of rare events were derived and their statistical properties were examined. Bayesian Model was embedded in the designed ACS to develop the model for predicting the total number of rare events. Parameters α, β and λ were used in the model to control the expected number of grid cells with rare events, the conditional expected number of sub-network and expected number of rare events in each sub-network respectively. Markov Chain Monte-Carlo Algorithm with R and Winbugs software were used to estimated these parameters. The robustness of the model was examined and its Sensitivity Analysis was

carried out. Diagnostic checks were done and the proposed model was compared with the existing model. The samples converged and represented the target posterior and the total number of rare events estimated lies within the 95% HPD credible interval. The derived estimators were unbiased, consistent and efficient. The model was criterion and inference robust with a good fit. The results revealed that rare event was best modelled under heterogeneity assumptions.

Keywords: Adaptive Cluster Sampling, Markov Chain Monte-Carlo (MCMC), Rare Events

1 Introduction

Rare events are of importance in many branches of science and may occur occasionally but are often of great impact and are therefore of high interest (Bouchet et al., 2019). Probabilities of rare event computation had been proven to be impossible or hard to observe due to the rarity of the event (Tuffin and Ridder, 2012). Kalton (2009) classified population into subgroups using proportion into rare ($< 0.01\%$), mini ($0.1 - 1\%$), minor ($1 - 10\%$) and major ($> 10\%$), while Tourangean (2014) referred to rare event population as hard-to-reach. Sampling from rare populations has boundless arrangement of study into methods (Wagner and Lee, 2014) and techniques used in knowing the status and distribution of the rare event because of limited sampling resources (Thompson, 2013; Breininger *et al.* 2019). The fewer the events of interest, the more the population may be difficult. Conventional sampling design like the Simple Random Sampling (SRS), Systematic Sampling, Stratified Sampling, are inappropriate to accomplish reliable accuracy for such populations (Qureshi *et al.,* 2020). As a result of these difficulties, sequence of recent development of new methods that are non-conventional were considered (Wayner and Lee, 2014). Adaptive survey designs have beautiful features for surveying large locations as long as the anticipated spatial spread for responding to heterogeneity maintains the basic elements of probability-based statistical survey (Brown *et al.*, 2013).

Adaptive cluster sampling technique use best of survey effort when valuable and also reduces survey effort where necessary. Adaptive cluster sampling technique accomplished better for the population when the rare events are sparsely distributed but yet clustered than usual grid cell sampling methods.

Bayesian inference and estimation has merit in statistical modelling as well as data analysis, Congdon (2006) which provides leyway to the improvement of sparse datasets estimation (Stroud, (1994); Richardson and Best (2003)), and permit predetermined sample inferences deprived of demand to big sample influences as in maximum likelihood and other classical procedures.

Bayesian inference has turned out to be strictly connected to sampling-based estimation functions of parameters that concentrate on the whole density. Iter-

ative Monte Carlo methods encompass recurring selection procedure that converged to sample from the posterior distribution (Smith and Gelfand, 1992).

Thompson (1990) presented the perception of adaptive sampling first and used the strategy to maximize survey effort where variable of interest was found. Richardson and Green (1997) fully developed Bayesian mixture analysis and analyze univariate normal mixtures with ordered prior model that stipulates approach that deals with weak prior information using reversible jump Markov Chain Monte Carlo (MCMC) methods. Chaudhuri (2000) and Chaudhuri*, et al.* (2004) provided simple methods for estimating total variable of interest in an unbiased manner and the variance which depends on specific earners with concentrations in scattered and unknown locations. Viallefont, *et al*. (2002) recommended modelling overdispersion of Poisson mixture and estimation using Markov Chain Monte Carlo algorithms, including reversible jump algorithms in a hierarchical Bayesian model. The posterior distributions of the unknown quantities in the mixture permits changing the measurement of the mixture. The performance of various move created for changing dimension was estimated and the model was extended by introducing covariates with homogenous or heterogeneous effect. Rapley and Welsh (2008) modelled adaptive cluster sample data and developed a model-based Bayesian analysis using the prior knowledge about the population to update the sampling design and make inference with homogeneity assumption among the networks.

Goncalves and Moura (2016) proposed for clustered and sparse populations a unit-level mixture model when the data are obtained from an ACS considering heterogeneity among units that belong to different clusters to estimate the population total. Thompson (2017) mentioned that Adaptive link-tracing network sampling methods are significant for sampling hard to reach populations. Also, that the most convenient methods for inference gotten from adaptively selected samples are mostly design-based and Bayesian methods.

Zhu and Fu (2018) developed k-step ACS based on Horvitz-Thompson (H-T) estimator to control final sample sizes and compared the effectiveness with adaptive cluster sampling Hansen-Horvitz (ACS-HH) and adaptive cluster sampling Horvitz-Thompson (ACS-HT) estimators. Generalized ratio estimator was proposed by Qureshi *et al*. (2020) under Stratified Adaptive Cluster Sampling (SACS) design for the estimation of heavily clustered population mean using single auxiliary variable to get maximum benefit under the SACS design. Olayiwola *et al.* (2020) developed a Bayesian based model to estimate total number of COVID-19 carriers in Nigeria and proposed the combination of contact tracing by professionals with the model developed to control the rate of COVID-19 spread in Nigeria.

Nolau *et. al* (2021) proposed a unit-level model that assumed relationship between counts and auxiliary variables in a Bayesian framework with ACS from an African Buffaloes population in a 24,108km^2 area of East Africa.

Lesaffre and Lawson (2012) said without too much emphasis that MCMC techniques permit handling statistical modeling problems that are impossible or even hard to evaluate with the method of maximum likelihood, thus

presenting to the applied Statistician a robust method of statistical modeling techniques.
This study modelled φ with heterogeneity among networks and within the network units. The data used is a secondary data which was obtained from The International Institute of Tropical Agriculture (*IITA*), Nigeria on an understory specie of plant known as Culcasia Scandens and also a simulated data to validate the model.

2 Methodology

Adaptive cluster sampling is applicable in any situation where the variable of interest is distributed sparsely but heavily clustered.

2.1 Adaptive Cluster Sampling Design

Let region φ containing rare events with population of size N be partitioned into T regular grid cells. Then, the rare event in a location represents the average rare events quantity in that location. The population of rare events represent the true posterior probabilities which is not known at any one time.
The networks are linked such that adjacent networks can be accessed easily from one another.

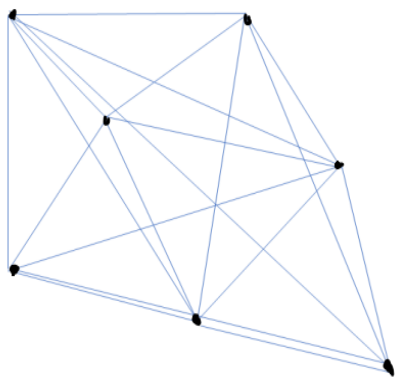


Fig. 1: Network Linked structure

There was need for a general distribution to enable wider accessibility order than the adjacent networks. From the population of rare events $f(x, \theta)$ with unknown parameter θ, samples were taken to estimate the unknown parameter θ. The prior information about the rare event in the population was either present or absent and the distribution obtained before the data set (y) were seen as prior distribution. The probability density function (pdf) of the data

set (y) can be defined as;

$$P(y) = \int_{\theta \epsilon \Theta} p(y, \theta) \tag{1}$$

using Bayes' rule

$$p\left(y \mid \theta\right) = \frac{p\left(y, \theta\right)}{p(\theta)}$$

$$p(y) = \int_{\theta \epsilon \Theta} p\left(y|\theta\right) p(\theta) d\theta \tag{2}$$

Another set of data to be obtained (y') when the experiment was repeated was the posterior distribution of the data set.

$$p(y'|y) = \int_{\theta \epsilon \Theta} p(y') p(\theta|y) d\theta \tag{3}$$

also using Bayes' rule,

$$p(y'|y) = \int_{\theta \epsilon \Theta} p(y'|\theta, y) p(\theta|y) d\theta \tag{4}$$

Where y' and y are independent, gave the product of likelihood and posterior.

For this work, normal distribution with variance σ^2 was suitable as symmetric proposal distribution (McElreath, (2016); Ford, (2018)) then,

$$\theta_{prop} \sim Normal(\theta_{current,}\ \sigma^2)$$

θ is the probability of either there is rare event in the grid cells or not. For each network $f\left(x\right) = \theta^x (1-\theta)^{1-x}$ and repeated over nth times of Bernoulli trials, the likelihood function becomes Binomial such that the $data \sim Binomial\ (n,\ p = \theta)$.

For binomial distribution, $0 < \theta < 1$ and $0 < \pi < 1,\ 0 < a < 1,\ 0 < b < 1$ for beta distribution, therefore, beta (a, b) is valid as conjugate prior for binomial.

$Beta\ (a,\ b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1},\ 0 < \pi < 1,\ 0 < a < 1,\ 0 < b < 1$, a prior for a random variable θ is expressed as

$$g\left(\theta\right) = \frac{\Gamma\left(a+b\right)}{\Gamma\left(a\right)\Gamma\left(b\right)} \theta^{a-1} (1-\theta)^{b-1}$$

The posterior distribution is a beta distribution as well with parameter $(\theta|x) \sim beta\,(a+x, n+b-x)$.

2.2 Development of the Predictive Model to Estimate the Total Number of Rare Events in a Population.

The existing predictive model (Ralph and Welsh 2008) for estimating the total number of rare events $\hat{N}$ was expressed as,

$$\hat{N} = 1_{P_0}^T N_0 + 1_{P_1}^T \hat{N}_1 \tag{5}$$

where $1_{P_0}^T$ - vector of zeros of the edge unit, $1_{P_1}^T$- vector of ones for the non empty network,
N_0 - Number of the rare events in the edge units, N_1 - Number of the rare events within the network. The proposed predictive model for estimating the total number of rare events $\hat{N}$ when partitioned into non-overlapping sub-networks became

$$\hat{N} = 1_{P_0}^T N_0 + 1_{P_{11}}^T \hat{N}_{11} + 1_{P_{12}}^T \hat{N}_{12} + 1_{P_{13}}^T \hat{N}_{13} + \cdots + 1_{P_{1n}}^T \hat{N}_{1n} \tag{6}$$

where
$1_{P_0}^T$ - vector of zeros in the edge units, N_0 - Number of rare events in the edge units
$1_{P_{11}}^T$- vector of ones in sub-network 1, N_{11} - Number of rare events in sub-network 1
$1_{P_{12}}^T$- vector of ones in sub-network 2, N_{12} - Number of rare events in sub-network 2

$\vdots$

$1_{P_{1n}}^T$- vector of ones in sub-network n, N_{13} - Number of rare events in sub-network n
Therefore,

$$\hat{N} = 1_{P_0}^T N_0 + 1_{P_{11}}^T \hat{N}_{11} + 1_{P_{12}}^T \hat{N}_{12}, \cdots, 1_{P_{1n}}^T \hat{N}_{1n}$$

$$\equiv 1_{P_0}^T N_0 + \sum_{i=1}^{n} 1_{P_{1i}}^T \hat{N}_{1i} \tag{7}$$

i.e., $1_{P_{1i}}^T$- vectors of ones from within the sub-networks
$\hat{N}_{1i}$ - Number of rare events from within the sub-networks

2.3 Mean and Variance of the existing Adaptive Cluster Sampling estimator with homogeneous assumption within the network was defined as

$$\widehat{\mu} = 1/n \sum_{i=1}^{n} w_i \tag{8}$$

and

$$w_i = \frac{1}{m_i} \sum_{\sum s} y_i \tag{9}$$

The variance of the initial sample without replacement is

$$var\,(\widehat{\mu}) = \frac{N-n}{Nn(N-1)} \sum_{i=1}^{N} (w_i - \mu)^2 \tag{10}$$

C_i was the network that included unit i, and m_i the number of units contained in the network. w_i was the observation mean for the network that included the i^{th} unit of the initial sample and n the initial sample size, such that a modified unbiased estimator of the initial sample mean was of the form (Turk, (2005); Dryver and Thompson (2005), Thompson (1990)).

2.4 The Proposed Estimator with Heterogeneous Assumption among Networks and within the Network Units

Let C be the number of distinct networks such that each network in C is assumed heterogeneous in nature.

LAYOUT FOR PROPOSED ADAPTIVE SAMPLING DESIGN

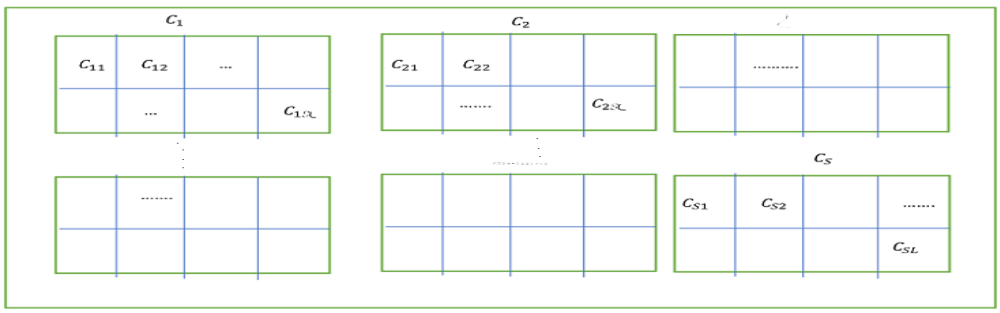


Fig. 2: SxL networks for proposed design

Partitioning each network in $C = (C_1, ..., C_S)'$ into non-overlapping homogeneous sub-networks say L_i of equal sizes.

$$C_1 = C_{11},\ C_{12},\ \cdots\cdots\cdots,\ C_{1L}$$

$$C_2 = C_{21},\ C_{22},\ \ldots\ \cdots\cdots, C_{2L}$$

$$\vdots$$

$$C_S = C_{S1},\ C_{S2},\ \cdots\cdots\cdots,\ C_{SL},$$

then $C = \{(C_{11},\ C_{12},\ \ldots,\ C_{1L}), (C_{21},\ C_{22},\ \ldots,\ C_{2L}),\ \cdots (C_{S1},\ C_{S2},\ \ldots,\ C_{SL})\}$, where the cardinality of C is SxL and it is scalar.
Now extending C((S x L)-vectors) to vector Z = (C', I'_{N-X})$'$ of measurement $N-X+S$, where I'_{N-X} is the vector of edge unit, with measurement $N-X$, then $X = \sum_{i=1}^{S}\sum_{j=1}^{L} C,\ i = 1,\ 2,\ \cdots,\ S\ and\ j = 1,\ 2,\ \cdots,\ L$
The observations from each of the non-overlapping homogenous sub-networks can be represented as

$$C = \left\{ \begin{array}{l} \langle (y_{111},\ \ y_{112}, ..., y_{11m_{11}}), (y_{121},\ \ y_{122}, ..., y_{12m_{12}}), ..., (y_{1l1},\ \ y_{1l2}, ..., y_{1lm_{1l}})\rangle, \\ \langle (y_{211},\ \ y_{212}, ..., y_{21m_{21}}), (y_{221},\ \ y_{222}, ..., y_{22m_{22}}), ..., (y_{2l1},\ \ y_{2l2}, ..., y_{2lm_{2l}})\rangle, \\ \ldots\ \ldots \\ \langle (y_{s11},\ \ y_{s12}, ..., y_{s1m_{sl}}), (y_{s21},\ \ y_{s22}, ..., y_{s2m_{s2}}), ..., (y_{sl1},\ \ y_{sl2}, ..., y_{slm_{s1}})\rangle \end{array} \right\}$$

where $i = 1,\ 2,\ 3,\ \cdots,\ s; j = 1,\ 2,\ 3,\ \cdots,\ l; k = 1,\ 2,\ 3,\ \cdots,\ m$ and $m_{1L} + m_{2L} + m_{3L} + \cdots + m_{SL} = m_{ij}$.

2.5 Proposed Estimator for the Total Number of Rare Events

Defining $Y^* = (Y_1, \ldots, Y_X)'$ to be the vector of rare events with the numbers of observations within each non- empty sub-network; such that, $Y_i \geq 1$. Observations in each sub-network are not the same size i.e., $m_1 \neq m_2 \neq m_3 \neq \cdots \neq m_l$ for at least one l in any of the l sub-networks, therefore it follows a proportional allocation procedure.
The estimator for the total number of rare events in the population will be

$$\hat{N} = \sum_{i=1}^{X}\sum_{j=1}^{L} Y_i^* \tag{11}$$

Let g be the distinct number of sub-networks in C_i and ξ_{h_i} be the weight attached to each sub-network. Put m_h to be the total number of adaptive samples from all sub-networks, $\overline{y}_h$ is the overall mean of samples from all sub-networks, as $h = 1,\ 2,\ \cdots,\ g$**.**

The average of y-units in the sub-network i be defined as

$$w_i^* = \sum_{h=1}^{g} \xi_{h_i} \overline{y}_{h_i}$$

and

$$\overline{y}_{h_i} = \frac{1}{n_i} \sum_{h=1}^{g} y_{h_i} \equiv \frac{1}{n_1} \sum_{i=1}^{h} y_{h_1} + \frac{1}{n_2} \sum_{i=1}^{h} y_{h_2} + \cdots + \frac{1}{n_g} \sum_{i=1}^{h} y_{h_n}$$

The proposed population mean is then defined as

$$\widehat{\mu}^* = 1/n_i \sum_{i=1}^{n_i} w_i^* = 1/n \sum_{i=1}^{n_i} \sum_{h=1}^{g} \xi_{h_i} \overline{y}_{h_i} \tag{12}$$

note that $\xi_{h_i} = \frac{n_{h_i}}{n_i}, \sum_{h=1}^{g} \xi_{h_i} = 1 \; and \; \sum_{h=1}^{g} n_{h_i} = n_i$
The proposed population total becomes

$$\hat{Y} = N\widehat{\mu}^* = N/n_i \sum_{i=1}^{n_i} w_i^* = N/n \sum_{i=1}^{n_i} \sum_{h=1}^{g} \xi_{h_i} \overline{y}_{h_i} \tag{13}$$

2.6 Variance for the Proposed Estimator

The variance estimator can be defined as

$$var\left(\widehat{\mu}^*\right) = \sum_{i=1}^{n} \left\{ \frac{\sum_{h=1}^{g} \xi_{h_i} s_{h_i}^2}{n_i} - \frac{\sum_{h=1}^{g} \xi_{h_i} s_{h_i}^2}{n} \right\} \tag{14}$$

Proof

$$var\left(\widehat{\mu}^*\right) = var\left(1/n_i \sum_{i=1}^{n_i} \sum_{h=1}^{g} \xi_{h_i} \overline{y}_{h_i}\right)$$

$$= 1/n_i^2 var\left(\sum_{i=1}^{n_i} \sum_{h=1}^{g} \xi_{h_i} \overline{y}_{h_i}\right)$$

$$= 1/n_i^2 var\left(\sum_{i=1}^{n_i} (\xi_{1i} + \xi_{2i} + \cdots + \xi_{gi}) \overline{y}_{h_i}\right)$$

$$= 1/n_i^2 var\left(\sum_{i=1}^{n_i}(1+1+\cdots+1)\overline{y}_{h_i}\right)$$

$$= 1/n_i^2 var\left(\sum_{i=1}^{n_i}(n_i)\,\overline{y}_{h_i}\right) = 1/n_i^2 var\left(n_i^2\overline{y}_{h_i}\right)$$

$$\Rightarrow var\left(\overline{y}_{h_i}\right) = var\left(\overline{y}_{1_i}+\overline{y}_{2_i}+\cdots+\overline{y}_{g_i}\right)$$

$$= \left[var\left(\overline{y}_{1_i}\right)+var(\overline{y}_{2_i})+\cdots+var(\overline{y}_{g_i})\right]$$

With the selection of rare event done adaptively from non-overlapping sub-networks and the samples are not equal for at least one stratum, the covariances are zero since the sub-networks are independent.
The variance estimator then becomes;

$$var\left(\hat{\mu}^*\right) = \left\{ \begin{array}{l} \left[\frac{\sum_{h=1}^{g}\xi_{h_1}s_{h_1}^2}{n_1} - \frac{\sum_{h=1}^{g}\xi_{h_1}s_{h_1}^2}{n}\right] + \left[\frac{\sum_{h=1}^{g}\xi_{h_2}s_{h_2}^2}{n_2} - \frac{\sum_{h=1}^{g}\xi_{h_2}s_{h_2}^2}{n}\right] + \cdots \\ \qquad\qquad + \left[\frac{\sum_{h=1}^{g}\xi_{h_n}s_{h_n}^2}{n_n} - \frac{\sum_{h=1}^{g}\xi_{h_n}s_{h_n}^2}{n}\right] \end{array} \right\}$$

$$\therefore var\left(\widehat{\mu}^*\right) = \sum_{i=1}^{n}\left\{\frac{\sum_{h=1}^{g}\xi_{h_i}s_{h_i}^2}{n_i} - \frac{\sum_{h=1}^{g}\xi_{h_i}s_{h_i}^2}{n}\right\}$$

Where $s_{h_i}^2$ is the within-subnetwork variance

$$s_{h_i}^2 = \frac{1}{n-1}\sum_{h=1}^{g}\left(y_i-\overline{y}_{h_i}\right)^2$$

2.7 Properties of Proposed Estimator for the Mean

2.7.1 Unbiasedness

$$E\left(\widehat{\mu}^*\right) = E\left(1/n_i\sum_{i=1}^{n_i}\sum_{h=1}^{g}\xi_{h_i}\overline{y}_{h_i}\right) = E\left(\sum_{h=1}^{g}\xi_{h_i}\overline{y}_{h_i}\right)$$

$$= E\left\{ \begin{array}{l} (\xi_{11}\bar{y}_{1_1}+\xi_{12}\bar{y}_{1_2}+\cdots+\xi_{1g}\bar{y}_{g_1})+(\xi_{21}\bar{y}_{2_1}+\xi_{22}\bar{y}_{2_2}+\cdots+\xi_{2g}\bar{y}_{g_2}) \\ \qquad +\cdots+\xi_{n1}\bar{y}_{n_1}+\xi_{n2}\bar{y}_{2_n}+\cdots+\xi_{ng}\bar{y}_{g_n}) \end{array} \right\}$$

$$= E\left\{ \begin{array}{l} (\xi_{11}+\xi_{11}+\cdots+\xi_{1g})\bar{y}_{h_1}+(\xi_{21}+\xi_{22}+\cdots+\xi_{2g})\bar{y}_{h_2}+\cdots \\ \qquad +\xi_{n1}+\xi_{n2}+\cdots+\xi_{ng})\bar{y}_{h_n} \end{array} \right\}$$

$$= E\left\{(\xi_{1i} + \xi_{2i} + \cdots + \xi_{gi})\overline{y}_{h_i}\right\}$$

$$= E\left\{1.\overline{y}_{h_1} + 1.\overline{y}_{h_2} + \cdots + 1.\overline{y}_{h_n}\right\} = E\left\{\overline{y}_{h_1} + \overline{y}_{h_2} + \cdots + \overline{y}_{h_n}\right\}$$

Recall that $E(\overline{y}_i) = \frac{\overline{Y}_{h_i}}{n_i}$
Then,

$$E\left\{\overline{y}_{h_1} + \overline{y}_{h_2} + \cdots + \overline{y}_{h_n}\right\} = \frac{\overline{Y}_1 + \overline{Y}_2 + \cdots + \overline{Y}_n}{n_i} = n_i * \frac{\overline{Y}}{n_i}$$

$$\therefore E(\overline{y}_i) = \overline{Y}$$

2.7.2 Consistency

$$\lim_{n\to\infty}(var(\widehat{\mu}^*)) = \lim_{n\to\infty}\left(\sum_{i=1}^{n}\left\{\frac{\sum_{h=1}^{g}\xi_{h_i}s_{h_i}^2}{n_i} - \frac{\sum_{h=1}^{g}\xi_{h_{11}}s_{h_i}^2}{n}\right\}\right) = 0.$$

Hence the estimator is consistent.

2.7.3 Efficiency

$$\sum_{i=1}^{n}\left\{\begin{array}{c}\frac{\sum_{h=1}^{g}\xi_{h_i}s_{h_i}^2}{n_i} - \frac{\sum_{h=1}^{g}\xi_{h_{11}}s_{h_i}^2}{n} \\ \text{proposed estimator}\end{array}\right\} < \left\{\begin{array}{c}\frac{1}{n(n-1)}\sum_{i=1}^{n}(w_i - \widehat{\mu}_1)^2 \\ \text{existing estimator}\end{array}\right\}$$

The numerator of the proposed estimator is a quantity multiplied by weight of each network work that is distinctly lesser than 1. Therefore, variance of the proposed estimator is less than the existing variance. Hence, it is efficient.

2.8 Bayesian Modelling

Based on the prior knowledge about the population that the events are rare and clustered, adaptive cluster sampling was adopted and Bayesian approach was used to fit the model. The MCMC approach was used to for estimating the unknown parameters α_{ij}, β, and λ_{ij} in the model, where α_{ij} is the parameter that controls the projected number of non-empty grid cells; β, is the the conditional projected number of non-empty networks, and λ_{ij} is the projected number of observations in each sub-network respectively and the unobserved quantities X_u, S_u, C_u and Y_u,

The proposed predictive model for estimating the total number of rare events cannot be evaluated analytically. Therefore, R and winbugs software was used to fit the model from MCMC output. MCMC was also used for the diagnostic check of the model.

2.9 Generating Heterogeneity Among the Networks

In the proposed model, the use of Coefficient of variation to control level of heterogeneity among the networks was considered. The focus was to ensure heterogeneity among the networks with minimum of 75% coefficient of variation such that, $1.01 \leq d \leq 1.75$ with $75\% \leq C.V \leq 100\%$ respectively and $v > 0$. Aside from these conditions, networks will be homogeneous.
λ_{ij} is the expected number of observations in sub-network i in network j, C is distributed with Gamma (d, v)

2.10 Generating Heterogeneity Among Units Within the Networks

Having β to control the conditional expected number of the non-empty networks,
$\beta \sim beta\,(a_\beta, b_\beta)$, Goncalves and Moura (2016) as

$$mean\,(\beta) = \frac{a}{(a+b)} \tag{15}$$

$$variance\,(\beta) = \frac{ab}{(a+b)^2(a+b+1)} \tag{16}$$

$$C.V = \frac{std}{mean} * 100\% = \left(\frac{\sqrt{ab}}{(a+b)\sqrt{a+b+1}} \div \frac{a}{(a+b)} \right) * 100\%$$

$$C.V = \left(\frac{100\sqrt{b}}{\sqrt{a(a+b+1)}} \right) \% \tag{17}$$

For sparse and clustered population, $a = 1,\ b = 9$, (Rapley and Welsh, 2008). Hence, $C.V = 81.82\%$.

2.11 Generation of α_{ij}, β and λ_{ij}

Knowing that the rare events are sparse and clustered, having α_{ij} controlled the expected number of non-empty sub-network i in network j, β controlled the conditional expected number of non-empty networks and λ_{ij} was fixed arbitrarily in each sub-network i in network j. The estimator performance depends on the basic values of $\alpha_{ij}\ and\ \beta$. For small values of $\alpha_{ij}\ and\ \beta$, the population total was well estimated, Rapley and Welsh (2008).
Consequent upon the prior knowledge of the rare event, the model became a Bayesian-based model, with the prior distributions of $\alpha_{ij} \sim beta\,(a_\alpha, b_\alpha)$ and $\beta \sim beta\,(a_\beta, b_\beta)$.

The conditional distribution of α given every other thing in the model proposed is

$$[\alpha_{ij}|X,\ S,\ \epsilon,\ C,\ Y,\lambda,\beta] \propto \pi\left(\alpha\right) \frac{\alpha^{X}(1-\alpha)^{N-X}}{1-(1-\alpha)^{N}} \tag{18}$$

Taking the prior distribution of α as a beta conjugate $\beta(a_\alpha, b_\alpha)$ with distribution $a_\alpha = 3\ and\ b_\alpha = 15$. The Beta distribution parameters were chosen to reflect that α was necessarily small in a sparse, clustered population, Rapley and Welsh (2008).
The conditional distribution of β given every other thing in the model is

$$[\beta|X,\ S,\ \epsilon,\ C,\ Y,\lambda,\alpha] \propto \pi\left(\beta\right) \frac{\beta^{X}(1-\beta)^{X-S}}{1-(1-\beta)^{X}} \tag{19}$$

The prior distribution of β taking to be a beta (a_β, b_β) distribution with $a_\beta = 1$ and $b_\beta = 9$, is necessary for β to be small, otherwise the population will not be clustered, Rapley and Welsh (2008).
Also given everything in the model, the conditional distribution of λ_{ij} is

$$[\lambda_{ij}|X, S, \epsilon,\ C, Y,\alpha,\beta] \propto \prod_{j=1}^{S} \prod_{i:\epsilon_i=j} \frac{\lambda^{y_i} exp\left(-\lambda_j\right)}{y_i!\left[1-exp\left(-\lambda_j\right)\right]} \pi\left(\lambda\right) \tag{20}$$

The prior distribution of λ_{ij} is the product of two Gamma distributions. $P(\lambda_{ij}\left|\theta\right), i = 1, \cdots, l; j = 1, \cdots, S$ all equal to the density of Gamma (d, v) with $\theta = (d, v)$, v also follows Gamma (e, f) distribution.

2.12 Gibbs sampler Algorithm Procedure

The steps in the Gibbs sampler for the algorithm goes thus;

a. Let the initial values be specified for the unsampled components $X_u,\ S_u,\ C_u,\ and\ Y_u$

b. Let α_{ij} generation *b*e from the conditional distribution $[\alpha_{ij}|X,\ S,\ \epsilon,\ C,\ Y,\lambda,\beta]$

c. Let β generation *b*e from the conditional distribution$[\beta|X,\ S,\ \epsilon,\ C,\ Y\ ,\lambda,\alpha]$

d. Let λ generation *b*e from the conditional distribution $[\lambda_{ij}|X, S, \epsilon,\ C, Y,\alpha,\beta]$

e. update the allocation ϵ such that C are also updated; and

f. Let $(X_u,\ S_u,\ \epsilon_u,\ C_u,\ Y_u)$ from the conditional distribution $[X_u,\ S_u,\ \epsilon_u,\ C_u,\ Y_u|X_o,\ S_o,\ \epsilon_o,\ C_o,\ Y_o, \alpha, \beta, \alpha_i, \lambda]$ be generated

g. Repeat from b.

2.13 The model parameters Estimation.

The model parameters were estimated using MCMC algorithm. The R package and Winbugs software were used to estimate and implement the MCMC algorithm. Provision was made for heterogeneity within network units and among networks in R code by imposing higher Coefficient of Variation (CV=81.89% and 99%) respectively to make them of different kinds.

2.14 The Proposed Predictive Models Evaluation

Fitting models to estimate number of rare events with heterogeneity assumption among networks and within the network units, 16 models were developed. Model 1 to model 16 for simulated data and also for the real-life data as shown in tables 1 and 2 with varying sample sizes n, $\alpha\ and\ \beta$. The influence of the difference prior models on the Poisson parameters and the performance of the Bayesian estimator were examined. Several simulated clustered populations and samples obtained from the posterior distributions of the model parameter and population parameters were sampled. The population estimates were then compared with the true values to evaluate the model's performance. The proposed predictive models generated using the real-life data when fitted was used and converged to true representation of the targeted posterior was shown in figure 3.

2.15 MCMC Diagnostics

Diagnostic checks were carried out on the model with real-life and simulated data to check the behaviour of the posterior model and see how the prior parameters have contributed or behaved in the processes as expected. Trace plot, Density plot, Gelman-Rubin, Geweke, Raftery-Lewis and Heidelberger-Welch were used for the diagnostics check.

3 Results, Discussion and Summary

The samples converged and gave reasonable target posterior representation showed by the fitted model. The total value estimated for real-life data falls in the accepted 95% HPD credible interval was seen in figure 3 below.

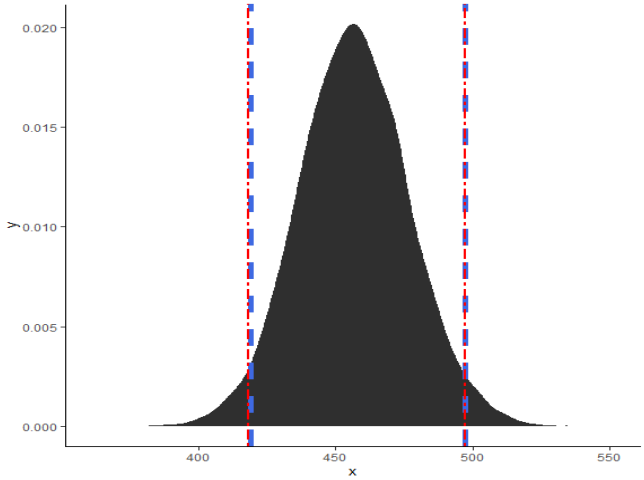


Fig. 3: 95% HPD for the target Posterior and Total Estimation for Real-Life Data

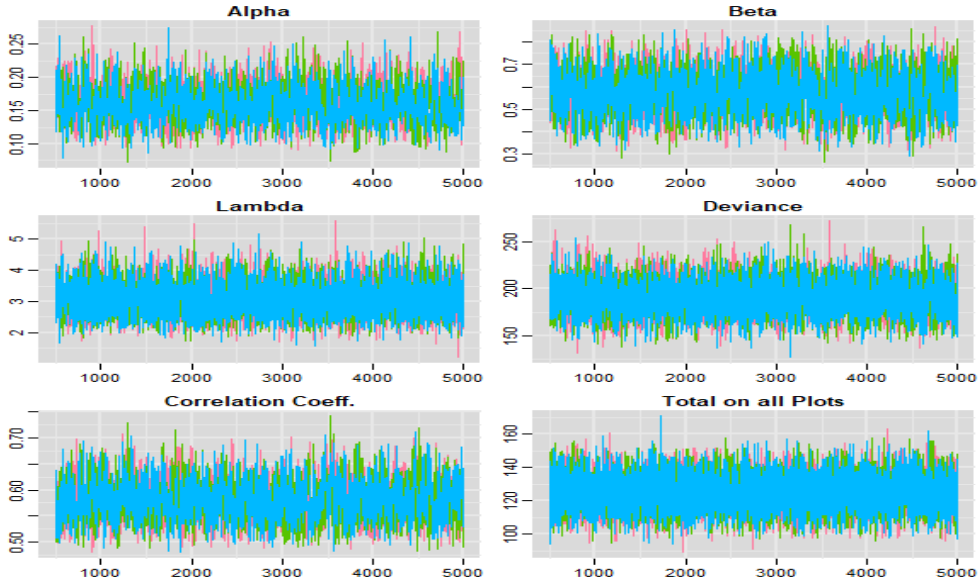


Fig. 4: Trace plot for simulated data

The diagnostics check; Trace plot, Density plot, Gelman-Rubin diagnostic, Geweke diagnostic, Raftery-Lewis diagnostic and Heidelberger-Welch diagnostic showed that fitted model was fit and efficient for estimating total number of rare and clustered event in a given population as in appendix.
For both simulated and real data, the trace plots as shown in figures 4 and 5 respectively gave a good representation of the posterior as all the samples

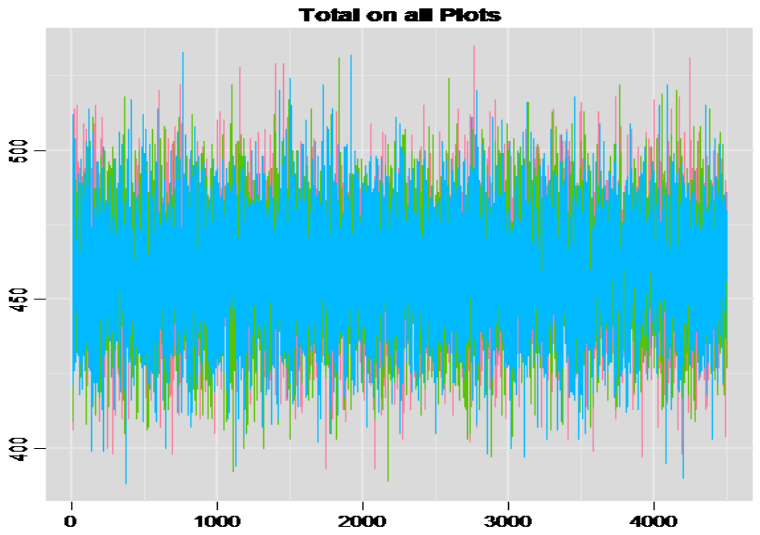


Fig. 5: Trace plot for Real-life data

chains mixed well. This then implied that the sample sizes converged to a reasonable point of the targeted posteriors especially for the total. The density plots for both the simulated and real data as shown in figures 6 and 7 visualized the concentrated around the mode of the posterior distribution.

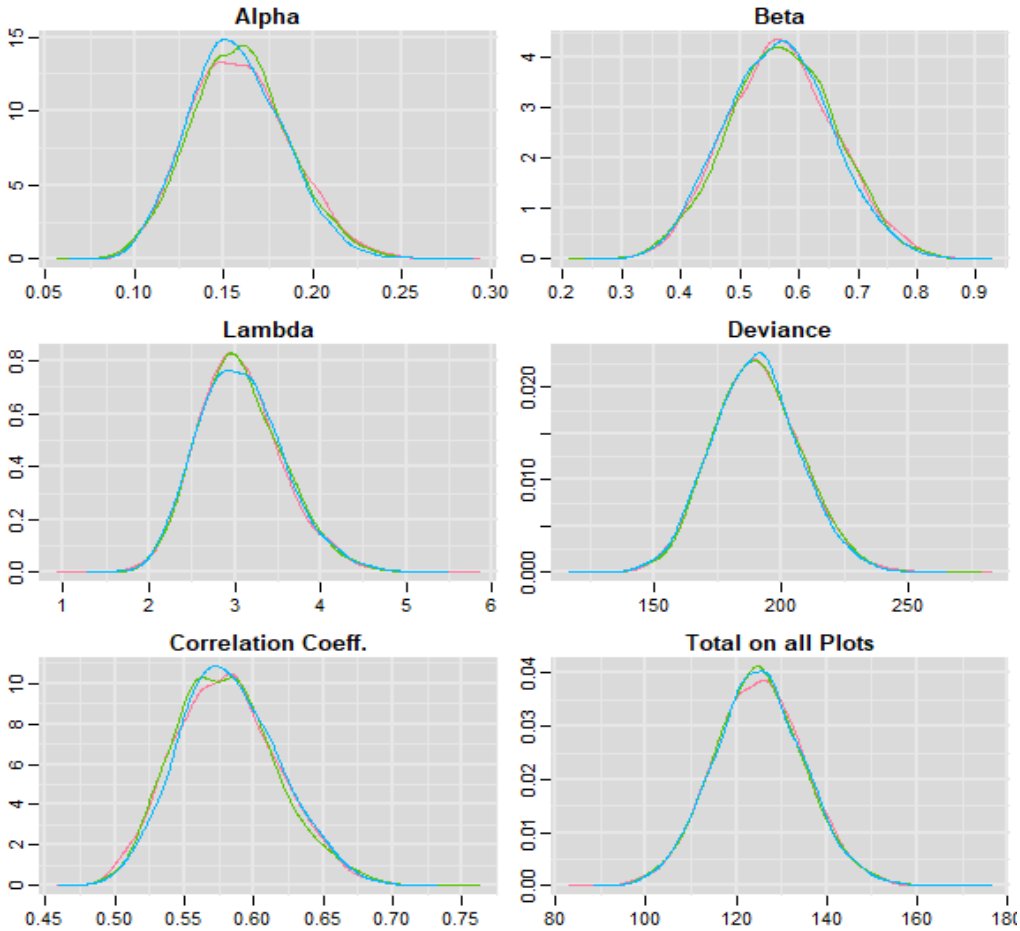


Fig. 6: Density plot for simulated data

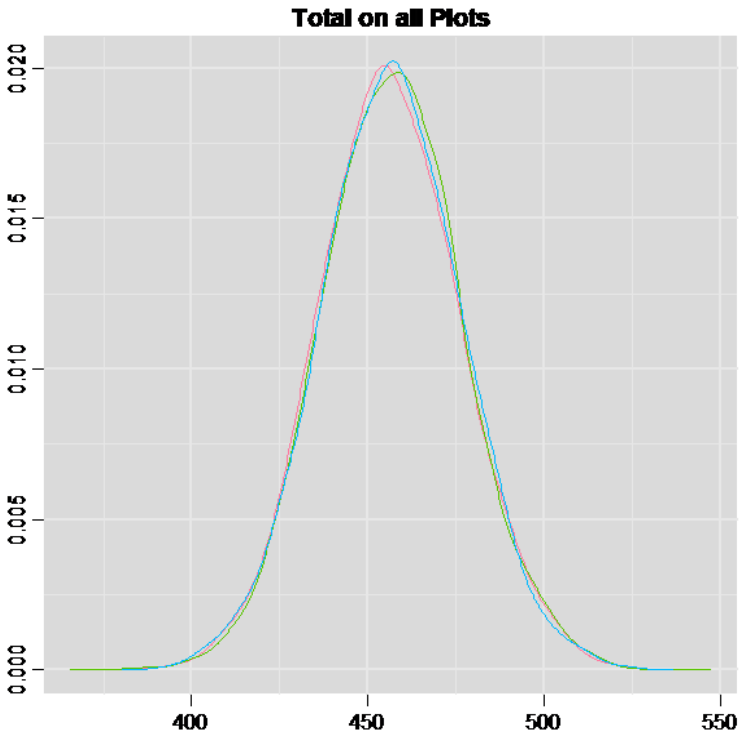


Fig. 7: Density plot for Real-life data

Table 1: Data estimates comparison of the models and different values for population parameters of α, β and N when population is heterogeneous for model 1 to model 16 using simulated data.

	N =200				N = 400				N = 600				N=800			
	0.10.1	0.10.15	0.150.1	0.150.15	0.10.1	0.10.15	0.150.1	0.150.15	0.10.1	0.10.15	0.150.1	0.150.15	0.10.1	0.10.15	0.150.1	0.150.15
set.seed	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
TotalDnew	113.43	125.04	219.04	224.21	326.42	310.96	401.04	357.91	470.26	418.80	699.94	700.80	591.66	489.14	810.44	814.12
B_pvalue	0.55	0.53	0.58	0.55	0.61	0.56	0.63	0.61	0.70	0.55	0.70	0.60	0.67	0.60	0.65	0.66
pD	164.60	149.0	182.90	185.90	298.60	298.30	300.50	308.70	435.00	404.70	479.50	491.70	590.70	556.70	645.10	622.90
DIC	349.90	339.43	483.49	484.55	851.48	788.19	936.61	870.63	1282.78	1053.64	1540.06	1501.09	1633.39	1439.52	1911.61	1897.22
alpha.p	0.23	0.16	0.33	0.28	0.26	0.23	0.33	0.26	0.30	0.19	0.42	0.34	0.27	0.21	0.36	0.32
beta.p	0.54	0.57	0.62	0.62	0.78	0.73	0.81	0.78	0.85	0.77	0.86	0.84	0.85	0.82	0.86	0.88
rmse_	5.02	3.63	3.34	2.89	2.96	3.26	2.81	2.58	3.38	3.04	2.79	2.14	3.37	3.45	3.14	2.40
rae_	2.99	1.18	0.99	0.76	0.79	0.95	0.72	0.63	1.02	0.83	0.72	0.47	1.01	1.06	0.88	0.56
cv_	21.15	16.37	13.51	12.75	12.56	13.02	12.17	11.67	10.84	10.74	9.84	9.10	9.51	9.82	8.97	8.09
rmse_t	2.42	3.04	4.04	4.21	11.42	6.96	13.04	11.91	19.26	8.80	20.94	17.80	18.66	14.14	18.44	26.12
rae_t	0.02	0.02	0.02	0.02	0.03	0.02	0.03	0.03	0.04	0.02	0.03	0.03	0.03	0.03	0.02	0.03
cv_t	7.85	7.97	5.73	5.78	5.29	5.26	4.68	5.00	4.43	4.56	3.52	3.60	3.91	4.36	3.23	3.34
cv_	17.88	17.34	13.35	13.25	9.49	10.87	8.76	9.52	7.34	9.18	6.51	6.96	7.09	7.71	6.27	6.23
cv_b	16.16	18.86	12.48	12.69	7.20	8.71	6.23	7.33	4.90	7.35	4.47	5.06	5.00	5.95	4.42	4.07

Table 2: The Real-Life Data estimates comparison of the models and different values for population parameters of α, β and N when population is heterogeneous for model 1 to model 16

	N =200				N = 400				N = 600				N=800			
	0.10.1	0.10.15	0.150.1	0.150.15	0.10.1	0.10.15	0.150.1	0.150.15	0.10.1	0.10.15	0.150.1	0.150.15	0.10.1	0.10.15	0.150.1	0.150.15
TotalDnew	456.158	456.719	456.386	456.109	458.839	458.358	458.588	458.687	467.024	466.686	466.690	466.860	467.317	467.586	467.814	467.912
alpha.p	0.106	0.106	0.106	0.107	0.079	0.079	0.079	0.079	0.074	0.074	0.074	0.074	0.090	0.090	0.090	0.090
beta.p	0.645	0.638	0.638	0.638	0.701	0.701	0.700	0.702	0.795	0.796	0.795	0.794	0.805	0.806	0.806	0.806
pD	138.700	134.300	137.700	135.000	202.800	202.700	201.200	203.700	265.500	258.400	258.900	258.100	453.700	454.900	443.800	449.400
DIC	437.100	431.700	435.100	432.100	607.300	607.100	605.100	608.200	801.500	794.800	795.100	793.800	1110.600	1111.900	1101.400	1106.900
B_pvalue	0.565	0.574	0.567	0.564	0.571	0.576	0.573	0.575	0.620	0.616	0.620	0.621	0.655	0.658	0.660	0.662
Rel. Bais	0.016	0.017	0.016	0.016	0.022	0.021	0.021	0.022	0.040	0.038	0.039	0.040	0.041	0.041	0.042	0.042
rmsr_	21.211	20.782	20.756	20.713	15.131	15.224	15.029	15.085	14.482	14.661	14.521	14.494	7.857	7.844	7.879	7.890
rae_	0.904	0.903	0.902	0.902	0.931	0.931	0.931	0.931	0.951	0.951	0.951	0.951	0.933	0.933	0.934	0.934
cv_	11.478	10.6740.	10.770	11.164	9.866	10.411	10.089	10.158	10.310	10.237	10.284	10.319	11.137	10.974	10.906	10.938
rmse_t	7.158	7.720	7.386	7.109	9.839	9.358	9.588	9.687	18.024	17.686	17.690	17.860	18.318	18.586	18.814	18.912
rae_t	0.016	0.017	0.016	0.016	0.021	0.020	0.021	0.021	0.039	0.038	0.038	0.038	0.039	0.040	0.040	0.040
cv_t	4.360	4.352	4.318	4.338	4.389	4.323	4.353	4.391	4.660	4.673	4.664	4.641	4.582	4.570	4.521	4.471
cv_	12.753	12.537	12.870	12.546	10.909	10.979	10.937	10.949	9.141	9.133	9.223	9.454	8.701	8.862	8.753	8.973
cv_b	12.782	12.875	12.992	12.959	10.450	10.335	10.371	10.331	7.123	7.023	7.071	7.117	7.098	7.145	7.142	7.047

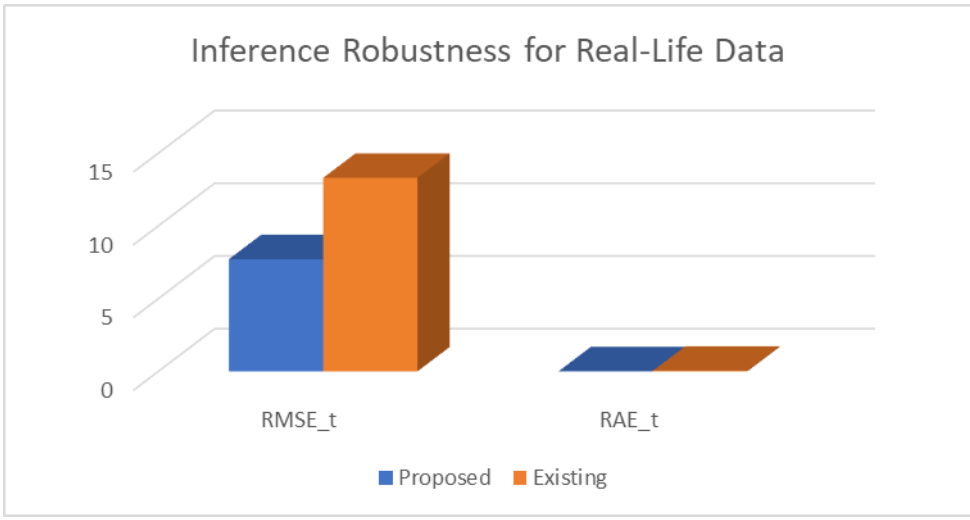


Fig. 8: Inference Robustness with RMSE and RAE

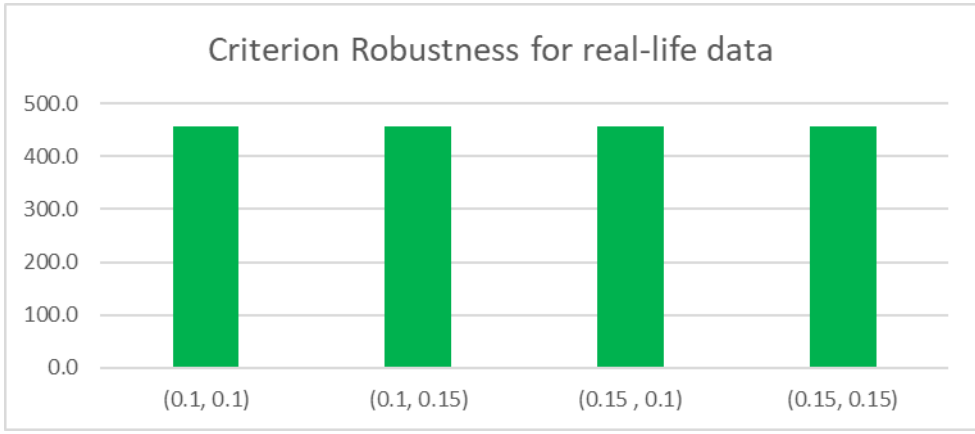


Fig. 9: Criterion Robustness for the fitted model

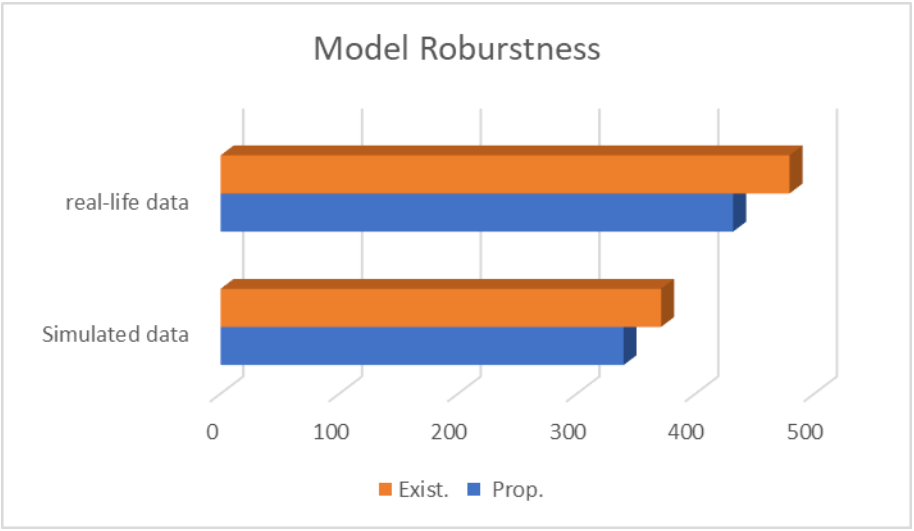


Fig. 10: Model Robustness for the fitted model

Table 3: Comparison between the proposed and existing models DIC

N	α	B	SIM. DIC	Existing DIC	Data DIC	Existing DIC
200	0.1	0.1	349.90	388.4512	437.100	485.98
200	0.1	0.15	339.43	371.0296	431.700	479.19
200	0.15	0.1	483.49	484.5981	435.100	476.95
200	0.15	0.15	484.55	513.5966	432.100	474.97

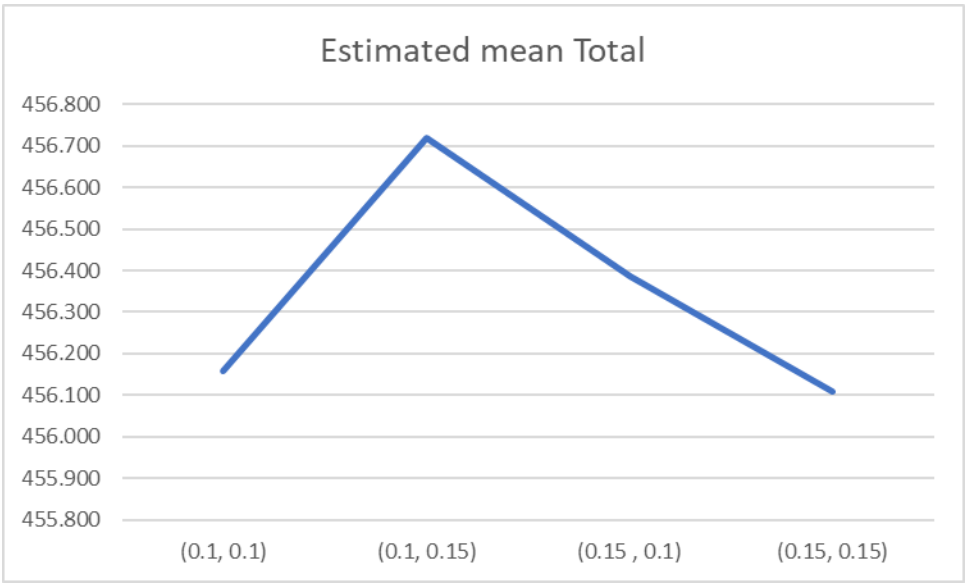


Fig. 11: Sensitivity analysis plot for the fitted model

Table 4: The Real data Summary Statistics of Posterior Estimates (means, Std dev. and quantiles) for the model

mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
totalDnew	456.719	19.875	418.000	443.000	457.000	470.000	497.000	1.00112000
lam	23.027	2.458	18.740	21.280	22.840	24.592	28.230	1.004730
rho.derived	0.575	0.029	0.523	0.554	0.573	0.594	0.637	1.00114000
fit	48.381	7.582	35.149	43.060	47.825	53.140	64.730	1.00114000
fit.new	50.649	9.454	34.560	43.917	49.860	56.490	71.295	1.0018000
alpha.p	0.106	0.013	0.082	0.097	0.106	0.115	0.134	1.0013900
beta.p	0.638	0.082	0.473	0.582	0.639	0.695	0.795	1.00114000
X.gridcells	21.319	5.115	12.000	18.000	21.000	25.000	32.000	1.0018000
R.networks	13.688	4.608	6.000	10.000	13.000	17.000	23.520	1.0016100
C.networks	8.884	4.116	2.000	6.000	8.000	11.000	18.000	1.00114000
deviance	297.392	16.392	266.500	286.200	297.100	308.200	330.500	1.0023200

An approach to monitor MCMC convergence output when m $>$1 chain was run as proposed by Gelman-Rubin (1996) built on the comparison between-chain and within-chain variances which implied that the values of R-hat near 1 suggest convergence and The Heidelberger-Welch diagnostic of all the chains passed the stationarity test and the half width test which implied that no longer MCMC iterations are needed for both the simulated and real-life data for convergence. Geweke (1992) proposed diagnostic for Markov chain based on Z-scored showed that all the parameters value monitored are in the interval of (-1.96, 1.96), then the interaction for each chain is okay as shown in table 5.

4 Conclusion

The results for fitted model with varying values of $N, \alpha \; and \; \beta$ as presented in tables 1 and 2 showed 16 models each for simulated and real data. The proposed model performed best when $\alpha = 0.1 \; and \; \beta = 0.15$ for both simulated and real data with DIC (339.43 and 431.70). The Root Absolute Error (RAE), Coefficient of Variation (CV) and Root mean square error (RMSE) are minimal for parameters estimated which gave the performance and accuracy of a predicted models.

The Estimation of whole number of rare events in a sparse and clustered population was considered. The model proposed gave alternative approach to the existing model of Goncalves and Moura (2016) whose assumption of units which belong to different networks was heterogeneity across the networks. Simulations and real-life data have shown that the proposed adaptive cluster model was efficient in comparison with the existing model.

Table 5: Geweke diagnostic for the real data

totalDnew	lam	rho.derived	fit	fit.new	alpha.p
-0.25999	0.86898	-1.06486	0.14945	0.15314	-0.09247
beta.p	X.gridcells	R.networks	C.networks	deviance	
1.16054	-0.77496	-0.10345	0.21075	0.83817	

References

Bouchet F., Rolland J, and Wouters J. (2019): Rare Event Sampling Methods. Chaos 29, 080402.0 https://doi.org/10.1063/1.5120509.

Breckling, J., Chambers, R., Dorfman, A., Tang, S., and Welsh, A. (1994). Maximum likelihood inference from sample survey data." International Statistical Reviews, 62:349{363. 719, 720, 723.

Bried J. T. **(**2012). Adaptive Cluster Sampling in the Context of Restoration. Society for Ecological Restoration Vol. 21, No. 5, pp. 585–591. doi: 10.1111/j.1526-100X.2012. 00924.x

Brown J. A. and Manly B.J.F. (1998). Restricted adaptive cluster sampling. Environmental and Ecological Statistics 5:49-63.

Chaudhuri, A. (2000). Network and adaptive sampling with unequalprobabilities. Cal. Stat. Assoc. Bull., 50, 238–53.

Congdon P. (2006). Bayesian Statistical Modelling. Second Edition ©John Wiley & Sons, Ltd.

Elzingha, C. L., Salzer D. W., Willoughby J. W., and Gibbs J. P. (2001). Monitoring plant and animal populations. Blackwell Science, Malden, Massachusetts.

Ford, P. (2018). MCMC Algorithms: Beyong Grid and Quadratic Approximattion.
https://www.casact.org/sites/default/files/old/studynotes_mcmc_algorithms_v0.6.pdf

Kalton G. (2009): Methods for oversampling rare subpopulations in social surveys. Survey Methodology, December 2009 125 Vol. 35, No. 2, pp. 125-141 Statistics Canada, Catalogue No. 12-001-X

Gelman, A. and Rubin, D. (1996) Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research*, **5**, 339–355.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian* Statistics *4*, Bernardo, J., Berger, J., Dawid, A. and Smith, A. (eds). Clarendon Press: Oxford.

Goncalves, Kelly C. M., Moura, Fernando A. S. (2016). A Mixture Model for Rare and Clustered Populations Under Adaptive Cluster Sampling. Bayesian Analysis, 11, Number 2, pp. 519–544.

Lesaffre E. and Lawson A. B. (2012). *Bayesian Biostatistics*, First Edition. John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing, 10(4):325–337.

McElreath, R., (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan, CRC Press.

Morrison L.W, Smith D.R, Young C.C, Nichols D.W. (2008). Evaluating sampling designs by computer simulation: a case study with the Missouri bladderpod. Population Ecology 50 (4):417-425.

Pfeffermann, D., Moura, F. A. S., and Silva, P. L. N. (2006). "Multilevel modelling under informative sampling." Biometrika, 93(4): 943–959. MR2285081. doi: http://dx.doi.org/10.1093/biomet/93.4.943.

R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rapley, V. and Welsh, A. (2008). "Model-Based Inferences from Adaptive Cluster Sampling." Bayesian Analysis, 3(4): 717–736. MR2469797. doi: http://dx.doi.org/10.1214/08-BA327. 520, 521, 522, 523, 535, 536, 539, 540

Richardson, S. and Best, N. (2003) Bayesian hierarchical models in ecological studies of health environment effects. *Environmetrics*, 14, 129–147.

Richardson, S. and Green, P. (1997). "On Bayesian analysis of mixtures with an unknown number of components." Journal of the Royal Statistical Society, Series B (Statistical Methodology), 59(4): 731–792. MR1483213. doi: http://dx.doi.org/10.1111/1467-9868.00095. 522, 527, 541

Smith, A. and Gelfand, A. (1992) Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician*, 46 (2), 84–88.

Stroud, T. (1994) Bayesian analysis of binary survey data. *Canadian Journal of Statistics*, 22, 33–45.

Thompson S. K (2017): Adaptive and Network Sampling for Inference and Interventions in Changing Populations. Journal of Survey Statistics and Methodology 5(1): 1-21. 10.1093/jssam/smw035

Thompson, S. (1990). "Adaptive cluster sampling." Journal of the American Statistical Association, 85: 1050–1059.

Thompson, W. L., (2004). Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters. Island Press, Washington, DC.

Tuffin R. and Ridder A. (2012): Probabilistic Bounded Relative Error for Rare Event Simulation Learning Techniques. Proceedings of the 2012 Winter Simulation Conference.

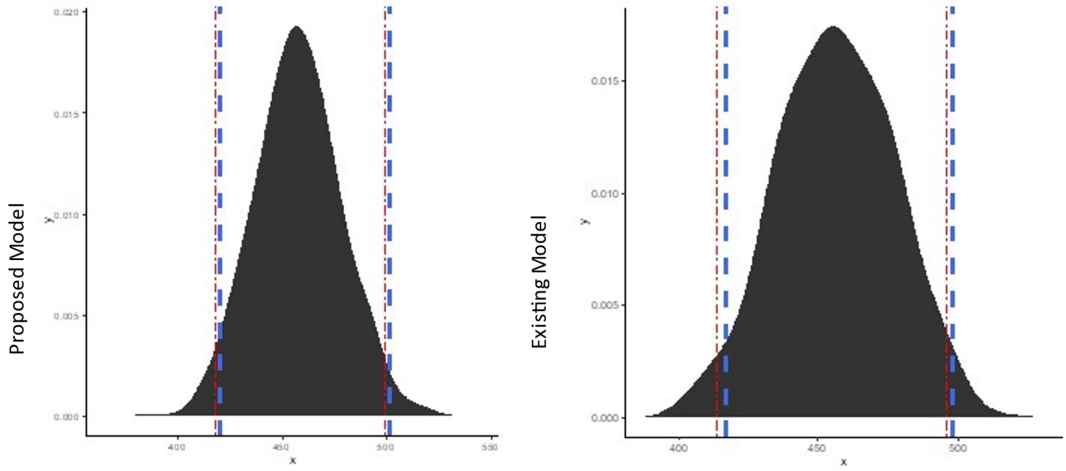
Viallefont, V., Richardson, S., and Green, P. J. (2002). "Bayesian analysis of Poisson mixtures." Journal of Nonparametric Statistics, 14(1–2): 181–202. MR1905593. doi: http://dx.doi.org/10.1080/10485250211383. 522, 526, 528

Zhu G. and Fu L. (2018). k-step adaptive cluster sampling with Horvitz–Thompson estimator. International Journal of Biomathematics, Vol. 11, No. 2 (2018) 1850029 (19 pages). World Scientific Publishing Company. DOI: 10.1142/S1793524518500298.

APPENDIX

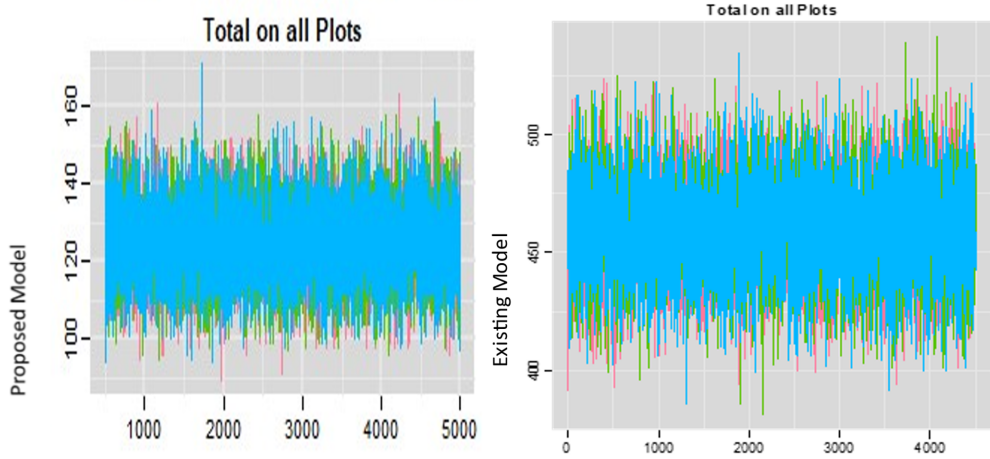
COMPARISON OF THE PROPOSED MODEL

HPD for Proposed Model using Real -Life Data



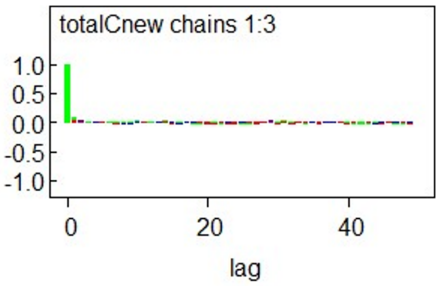
COMPARISON OF THE PROPOSED MODEL

Trace plot for Proposed Model using Simulated Data

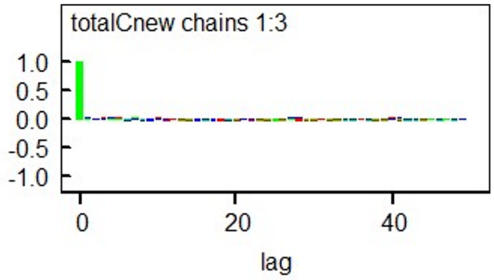


COMPARISON OF THE PROPOSED MODEL

Autocorrelation plot for Proposed Model using Simulated Data



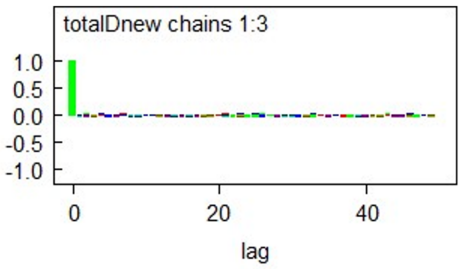
Proposed Model



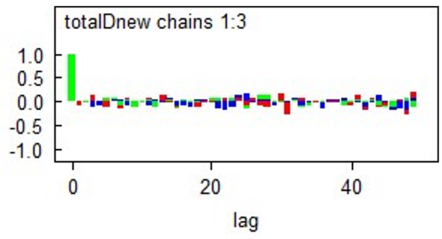
Existing Model

COMPARISON OF THE PROPOSED MODEL

Autocorrelation plot for Proposed Model using Reallife Data



Proposed Model



Existing Model