Supported by tetfund TERTIARY EDUCATION TRUST FUND

USING BAYESIAN MODELLING APPROACH TO ESTIMATE THE PREVALENCE OF MALARIA INFECTION IN CHILDREN UNDER 5 YEARS

Collins Aondona ORTESE[1, a], Kumafan Shadrach DZAAN[2,b], , Samuel Terese ASHEZUA[3,c] Sylvester Abua OGAR [4, d] and Edwin Hart OGWUCHE[5,e]

[1]Department of Mathematics and Computer Sc., Benue State University, Makurdi, Nigeria
[2]Statistics Department, Central Bank of Nigeria, Abuja
[3]Department of Mathematics, Statistics and Computer Science, Kwararafa University, Wukari, Nigeria
[4]Department of Statistics, Joseph Sarwuan University, Makurdi, Nigeria
[5]Department of Mathematics and Computer Science, Federal University of Health Science, Otukpo, Nigeria
Corresponding author: collinzaondona@gmail.com
Received: September 21, 2022 **Accepted:** November 12, 2022

Abstract

Malaria is the leading cause of death of children under 5 years of age in Sub-Saharan Africa and Nigeria in particular, it is an acute febrile illness caused by *Plasmodium* parasites transmitted through bites of infected female *Anopheles* mosquitoes. The aim of this study is to estimate the relative prevalence of the disease for this important age-specific group in the months of the year. The Bayesian Simulation Modeling Approach was employed in estimating the rates (unknown parameter π) and relative rates(R) of children infected with malaria for each month of the year. This was done using the Markov Chain Monte Carlo (MCMC) Algorithm implemented on the Windows Bayesian Inference Using Gibbs Sampling (WINBUGS) platform. The Beta and Gamma prior distributions were assumed and the Binomial likelihood as the preferred distribution to estimate the posterior paramaters using simulation. Retrospective aggregated data on registered Malaria patients under 5 years were retrieved from four major hospitals in Makurdi town and used as training dataset for the Algorithm. Results show that the malaria prevalence rate in children under 5 years is (π_1 =0.1461) 14.61% in January, a synonymous prevalence rate was observed in the following four months (February, March, April and May), June had a highest prevalence rate with an increase of 23.3% in relative to the previous month. Visually inspecting the density plot and history plot, we have confirmed the model convergence and adequacy. Density plot reflect the target distribution which further validates the prior distribution selected. From the findings, Bayesian modeling via simulation is the most suitable for studying disease spread and adjusting public health measures for disease control especially in the face of incomplete data.

Keywords: autocorrelation, Bayesian Models, Box plot, density plot, Likelihood, Posterior distribution, Prior distribution.

Introduction

Malaria remains a global significant public health challenge especially among children under five (5) years. It is one of leading infectious diseases that cause death on this age-specific group. After decades of control efforts, malaria still poses a serious public health threat, with 229 million estimated reported cases in 2019 (WHO, 2020) and 405,000 attributable deaths, of which two-thirds (272,000) occurred in children under 5 years of age (WHO, 2019). According to 2013 World Health Organization report, globally, an estimated 3.3 billion people are at risk of being infected with malaria and developing the disease (WHO, 2013). African Region accounted for heaviest global burden of Malaria (88%) and children less than 5 years of age accounts for 78% of all deaths (WHO, 2014).

Malaria is an acute febrile illness caused by *Plasmodium* parasites, which is spread to people through the bites of infected female *Anopheles* mosquitoes. There are five (5) parasite species that cause Malaria in humans, among them, *P.falciparum* is the deadliest parasite and the most prevalent in Sub-Saharan Africa.

According to World Health Organization (2021), Nigeria alongside DR Congo, Tanzania and Mozambique account for over half of all malaria cases and deaths worldwide of which Nigeria has the highest number of global malaria cases (27% of global malaria cases) and the highest number of deaths (32% of global malaria deaths) in 2020 (Houmsou *et al*., 2009)

Malaria is transmitted throughout Nigeria with 97% of the population at risk of malaria (NPC, 2015). It has a high mortality which has caused nearly a half million deaths in Nigeria, most of which were in children under 5 years of age (WHO, 2017). According to WHO, malaria has killed no fewer than 200, 000 Nigerians and afflicted 61 million others in 2021. Compared to the highest socioeconomic group, prevalence among children in the lowest socioeconomic group is seven times higher (38% vs 6%)(Okwa *et al*., 2000).

A systematic review of assessing patterns of malaria variation by age with respect to severity, transmission intensity and seasonality in Sub-Saharan Africa found clinical Malaria burden to be higher in younger age groups. Hospital admissions were also higher in children with higher levels of mortality among infants (Carneiro *et al*., 2010).

Amin *et al*. (2020) also developed a Bayesian model for malaria infection in the south of Iran. Results show that children under 6 years of age have an incremental impact on the incidence ratio in the study area. Oyibo *et al.* (2021) studied the geographical and temporal variation in reduction of Malaria infection among children under 5 years of age in Nigeria, their analysis showed a substantial sub-national variation in the extent and timing of reductions

in malaria infection in young children, with an apparent dichotomy before and after 2015.

The obvious research gab from previous analysis on the subject matter was the limited use of Bayesian modeling approach which is more efficient for aggregated data, incorporation of prior distribution and uncertainty measurement. In Bayesian modeling, a prior belief about a distribution is combined with likelihood from the available limited data to obtain a posterior belief. Bayesian approach gives a better option when you have a quantifiable prior belief, limited data and uncertainty in estimates.

Materials and Methods

We consider a set of Malaria patients y_i that expresses the number of Malaria Patients under 5 years over an aggregated daily registered cases per month (for i=1,2,…, k) for one year (k=1:12)

Hence $y_i \sim binomial\ (\pi, N_i)$, resulting to a likelihood given by

$$f(y/\pi) = \prod_{i=1}^{n}\{\binom{N_i}{y_i}\pi^{y_i}(1-\pi)^{N_i-y_i}\} \quad (1)$$

$$= \prod_{i=1}^{n}\{\binom{N_i}{y_i}\pi^{n\bar{y}}(1-\pi)^{N-n\bar{y}}\}$$

Where N = $\sum_{i=1}^{n} N_i$ is the Total number of the Bernoulli experiments in the sample. We consider a beta prior distribution with parameters $\pi = (a, b)^T$, denoted by beta(a, b) and density function

$$f(\pi) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\pi^{a-1}(1-\pi)^{b-1} \quad (2)$$

the resulting posterior is given as $f(\pi/y) \propto f(y/\pi)f(\pi)$ (3)

$$\propto \prod_{i=1}^{n}\{\binom{N_i}{y_i}\pi^{n\bar{y}}(1-\pi)^{N-n\bar{y}}\} \times \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\pi^{a-1}(1-\pi)^{b-1} \quad (4)$$

$$\propto \pi^{n\bar{y}+a-1}(1-\pi)^{N-n\bar{y}+b-1}$$

$\pi/y \sim beta(n\bar{y}+a, N-n\bar{y}+b)$

with posterior parameter $\bar{\alpha} = (n\bar{y}+a, N-n\bar{y}+b)^T$ and (5)

Posterior mean $E(\pi/y) = \widetilde{\mu_\pi} = \frac{n\bar{y}+a}{N+a+b}$ (6)

And posterior variance $V(\pi/y) = \tilde{\sigma}_\pi{}^2 = \frac{(n\bar{y}+a)(N-n\bar{y}+b)}{(N+a+b)^2(N+a+b+1)}$ (7)

A beta prior with low parameter values (a=b=0.01)

Given the likelihood and prior distribution, we have sample model parameters for the posterior distribution. After the parameters have been sampled for many iterations, parameter estimates are obtained and inference made.

The Model

Consider the following expression of probabilities of prevalence of Malaria infection in children under 5 years

$\pi_k = \pi_{k-1} \times R_k$ for k = 2,3, . . . , 12(corresponding to the 12 months of the year 2021), while π_1 is the estimate for the prevalence in the first month (January). Quantities (R_k) denote the prevalence of the malaria in children under five years in comparison to the total confirmed malaria patients.

We use the beta prior for the success probability of the first month and gamma priors for the relative success measures R_k. We use

$\pi_1 \sim beta(0.01, 0.01)$ and $R_k \sim gamma(0.01, 0.01)$ for k = 2,…, 12.

In the WINBUG syntax, we fit the Bayesian model.

The simulation was run for 1000 burn-ins after which samples were collected for 2000 iterations. A thinning of 32 would be maintained throughout the simulations and the overlay check box in WINBUG checked to reduce autocorrelation. Other modeling requirements are as stated by the WINBUG Software documentation. All analysis was done on the WINBUG software version 1.4.3

Method of estimating the Posterior Distribution

We employed the simulation approach to estimate the posterior distribution. The posterior distribution is described using the descriptive measures and density plots.

Model Convergence and Diagnostic Check

Model convergence diagnostics was done using history plots, trace plots and autocorrelation plots. The plots were produced while the model parameters and measures were monitored.

Results and Discussion

Data: A Retrospective data was retrieved from the database of four (4) renowned hospitals in Makurdi Metropolis (Benue State University Teaching Hospital, Makurdi, Federal Medical Centre, Makurdi, Bishop Murray Hospital, Makurdi and Foundation Hospital, Wurukum, Makurdi) from January 2021 to December, 2021 which was summed up and used for as training set for this study.

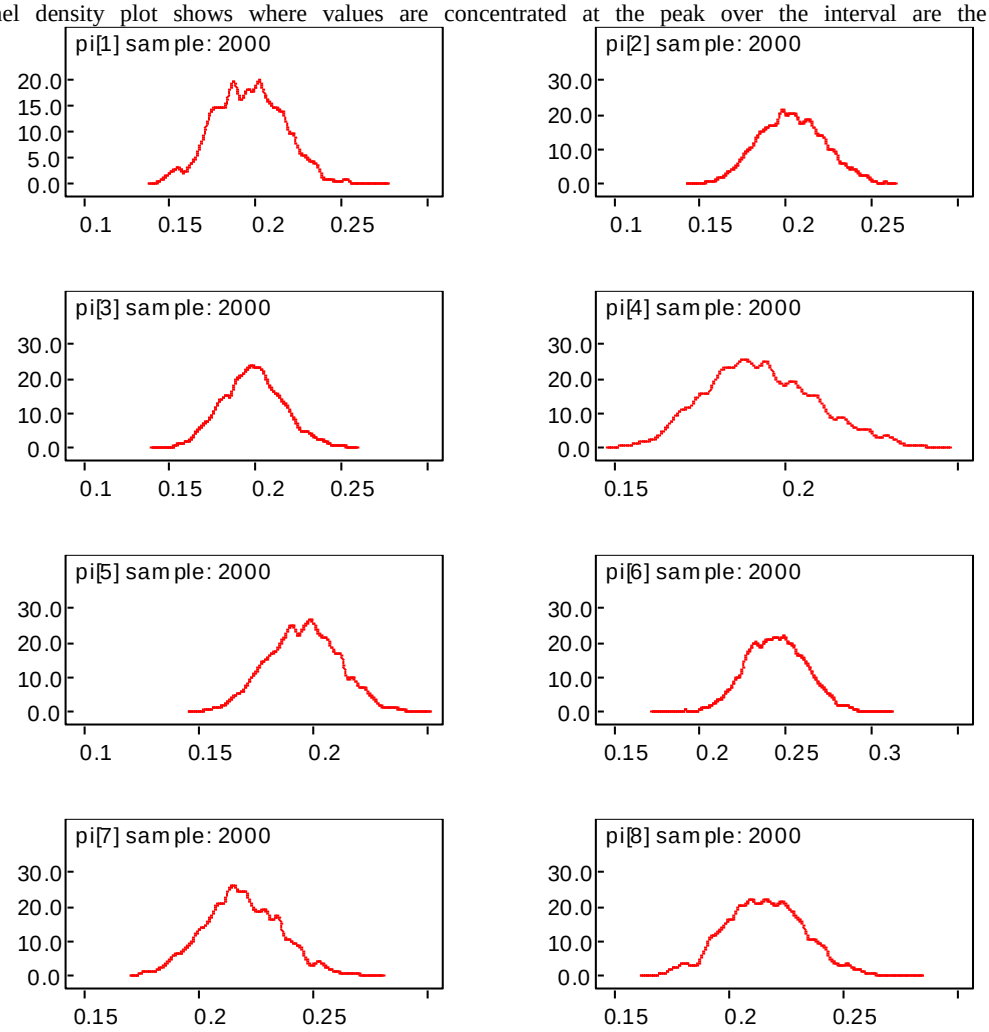
Calculation of Posterior Summaries: Table 2 below shows the estimates of the posterior mean of the node (unknown quantity), standard deviation of the node, computational accuracy of the mean (MC error), lower and upper endpoints (2.5% and 97.5% respectively) of the credible intervals and quantiles (including median) for the generated sample. The generated sample size (iterations) used to approximate the posterior distribution is 2000 and the burn-in period is 1000.

From the results, it can be inferred that the malaria prevalence rate in children under five (5) years was equal to 14.61% for January. For the next four months (February, March, April and May), the prevalence rate was about the same (posterior means for R_2=1.047, R_3=0.9842, R_4=0.9803 and R_5=1.022). In June, the prevalence rate increased 23.3% (R_6=1.255) and had the highest rate of prevalence in the year. There was a slight decline in the prevalence rate the following two months (R_7=0.9027 and R_8=0.9899). The month of September witnessed an increase in the prevalence rate 6.21% more than the previous month. Then decreased in October (R_{10}=0.8991) before increasing again in November and December (11.79% and 49.8% respectively).

Table 2: Posterior summaries

node	mean	Sd	MC error	2.5%	median	97.5%	start	sample
pi[1]	0.1961	0.02018	0.002151	0.1558	0.196	0.2353	1001	2000
pi[2]	0.2032	0.01904	0.001708	0.1674	0.2028	0.2414	1001	2000
pi[3]	0.1983	0.01784	0.001605	0.1646	0.198	0.2354	1001	2000
pi[4]	0.1928	0.01623	0.001267	0.1643	0.1918	0.2276	1001	2000
pi[5]	0.1958	0.01567	0.00109	0.1655	0.196	0.2262	1001	2000
pi[6]	0.2441	0.01772	0.001117	0.2099	0.244	0.278	1001	2000
pi[7]	0.219	0.01694	9.989E-4	0.1873	0.2183	0.2538	1001	2000
pi[8]	0.2156	0.01741	0.001104	0.1809	0.2155	0.2512	1001	2000
pi[9]	0.2253	0.01968	9.659E-4	0.1881	0.2252	0.2644	1001	2000
pi[10]	0.2012	0.01885	9.214E-4	0.1665	0.2006	0.2401	1001	2000
pi[11]	0.2027	0.02145	9.37E-4	0.1605	0.2029	0.2434	1001	2000
pi[12]	0.3036	0.02812	6.678E-4	0.2497	0.3036	0.3606	1001	2000

Posterior Density: Figure 1 below shows the graphical representation of the posterior density estimate for each unknown parameter. It plots a smoothed kernel density estimate which shows the distribution of values for our dataset. Visually inspecting the kernel density plot shows where values are concentrated at the peak over the interval are the point estimates.



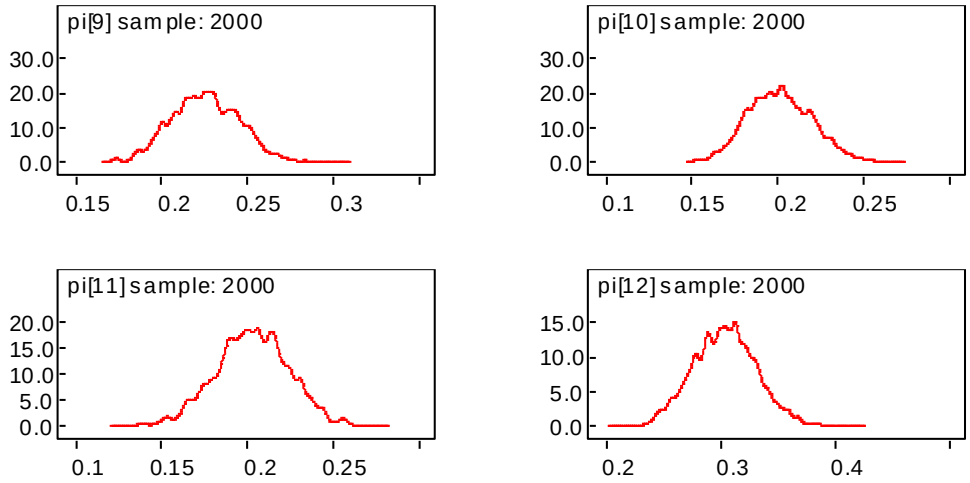
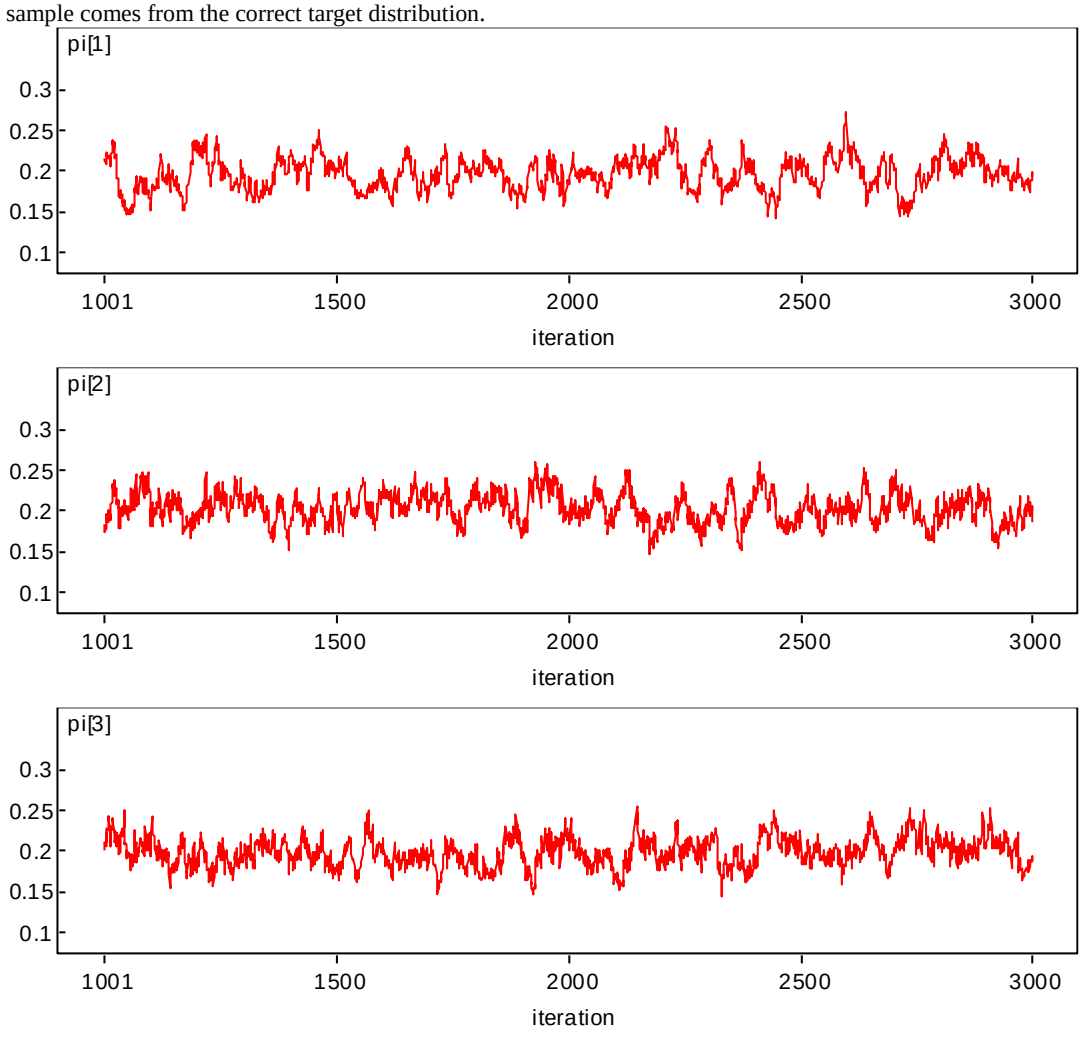
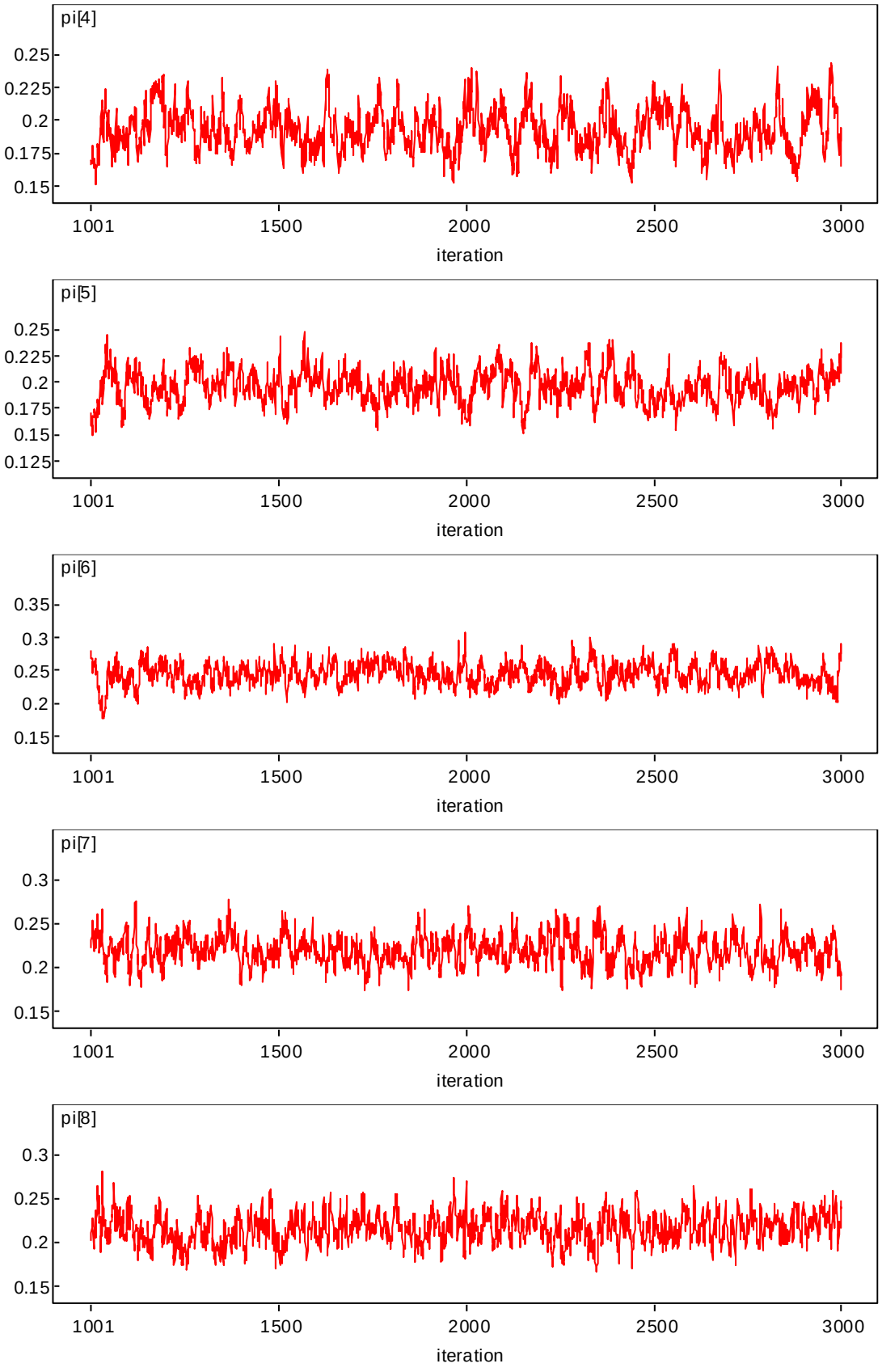


Figure 1: Density plot for model parameters $\pi and R$ for January to December.

***Checking convergence using History Plot*:** Figure 2 below shows the history plot for parameters π *and* R. Monitoring the patterns for the parameters shows that no patterns or irregularities exist and therefore convergence can be assumed. Plotting the trace plot also shows similar results. We therefore conclude that the algorithm has reached its equilibrium implying that the generated sample comes from the correct target distribution.




pi[4]
0.25
0.225
0.2
0.175
0.15
1001
1500
2000
2500
3000
iteration
pi[5]
0.25
0.225
0.2
0.175
0.15
0.125
1001
1500
2000
2500
3000
iteration
pi[6]
0.35
0.3
0.25
0.2
0.15
1001
1500
2000
2500
3000
iteration
pi[7]
0.3
0.25
0.2
0.15
1001
1500
2000
2500
3000
iteration
pi[8]
0.3
0.25
0.2
0.15
1001
1500
2000
2500
3000
iteration

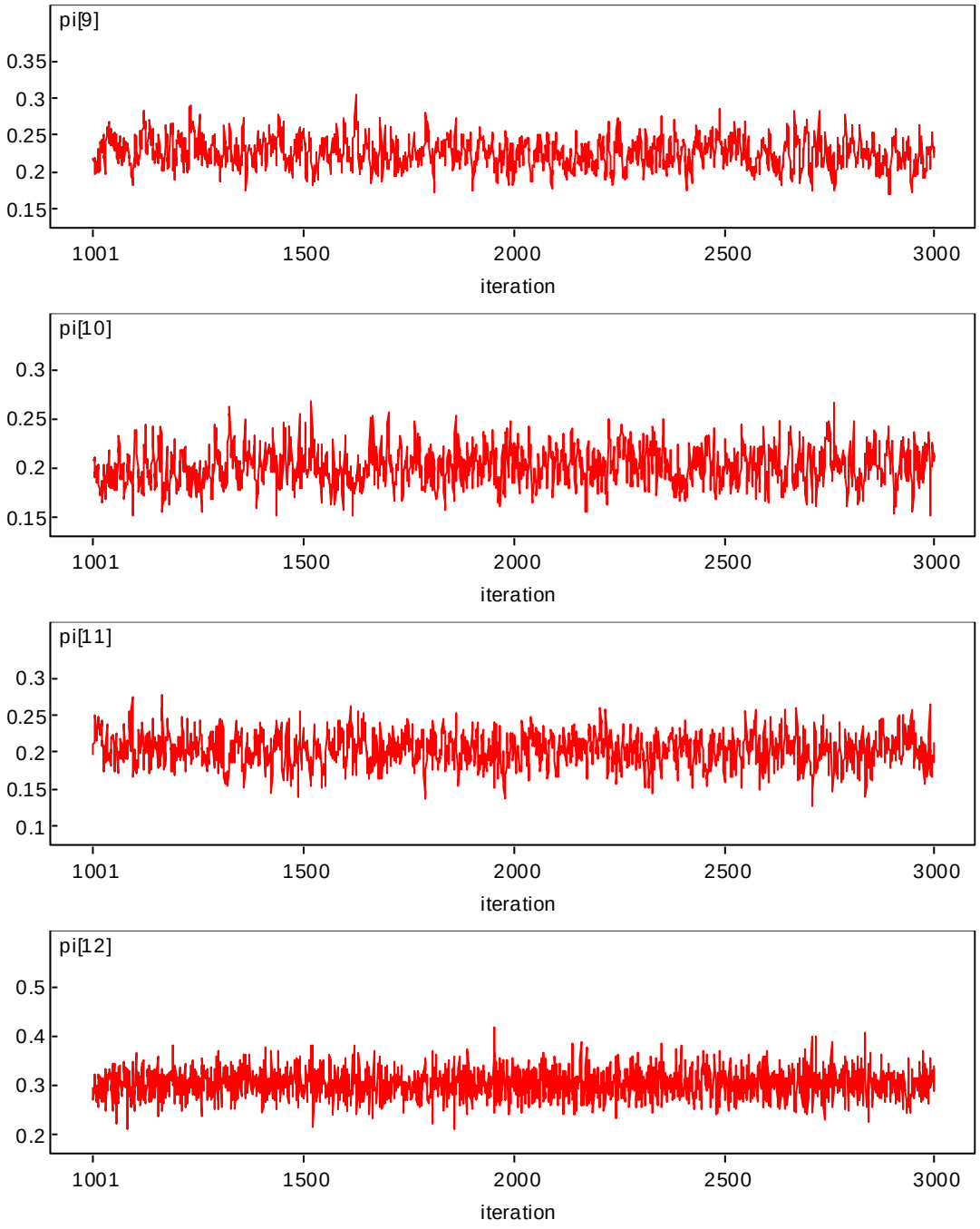


Figure 2: **History plot for model parameters π *and* R for January to December.**

Autocorrelation plot: Figure 4 shows the autocorrelation plot. We can observe that for all parameters (π and R) become low only after considering a lag equal to 20. This implies that an independent sample can be obtained by running the algorithm within *thin* set equal to 20 at the update tool. The autocorrelation plots of each parameter and measure depict the independence of the samples generated. This is because the autocorrelations become negligible fairly quickly, after a few lags

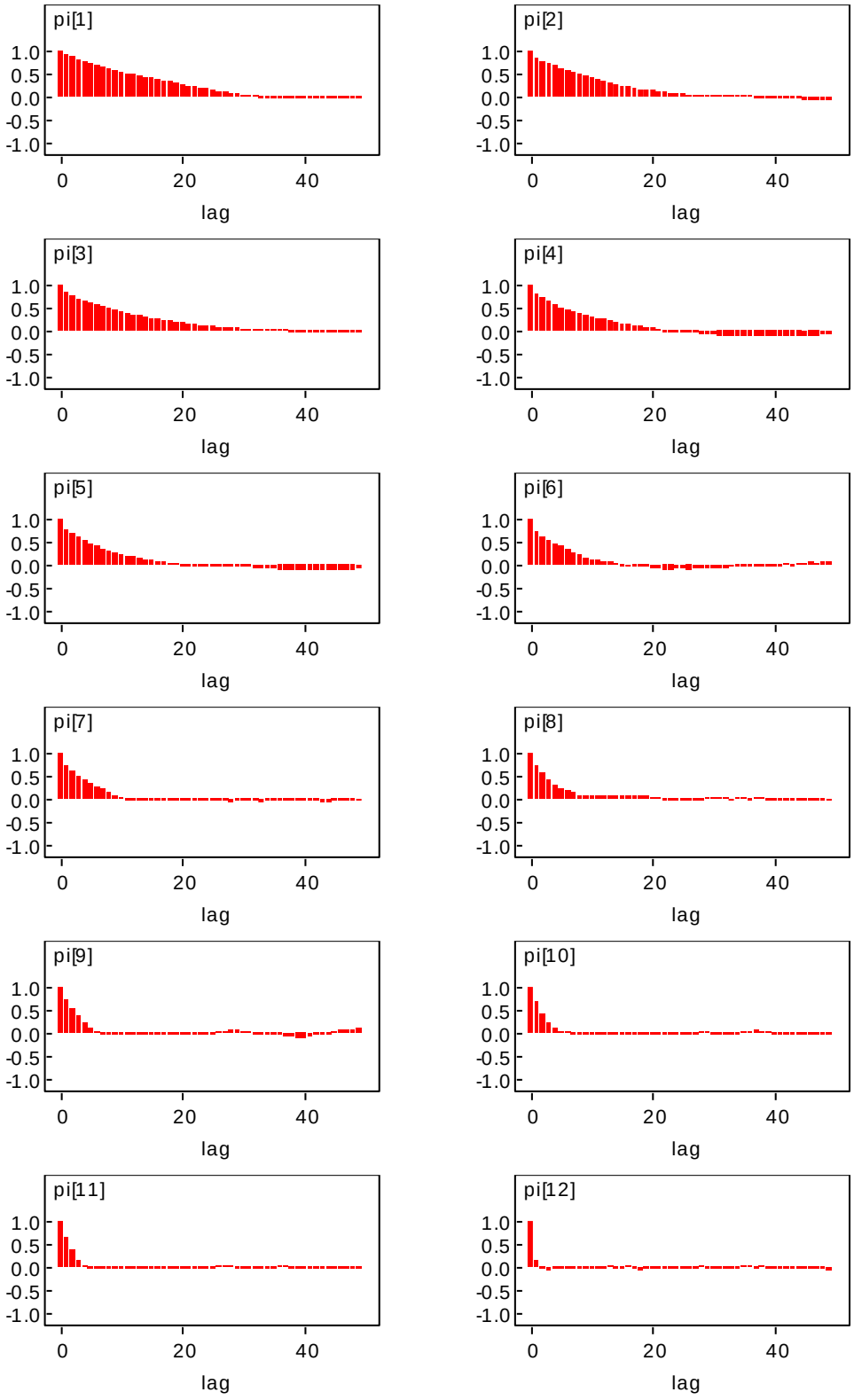


Figure 4: **Autocorrelation plot for model parameter π for January to December.**

Box Plot of the Posterior Distribution*:* This displays the posterior median, the quartiles, the range of values covered by the dataset and any outlier which may be present. Figure 6a and 6b below shows the box plot for parameters π and R.

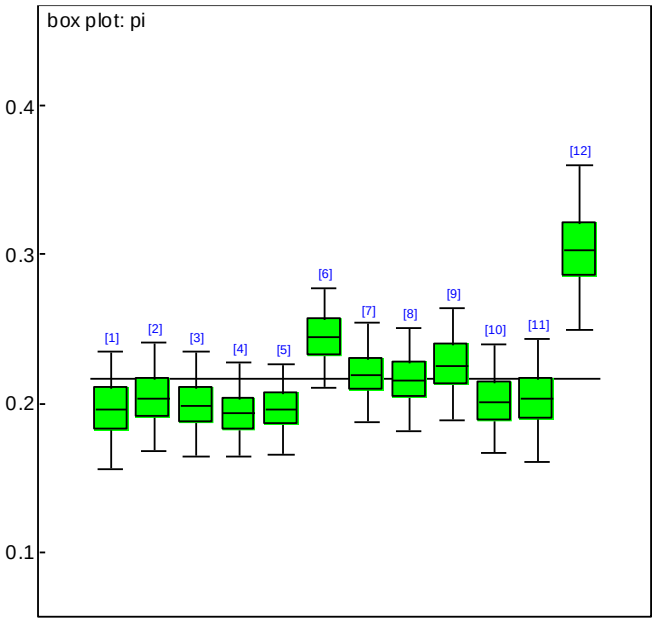


Figure 6: Boxplot for model parameters π for January to December.

Model fit and adequacy: Figure below (7a and 7b) shows the model fit for the parameter π and R. from the illustration, we can observe that that model is equivalent to a piecewise linear model

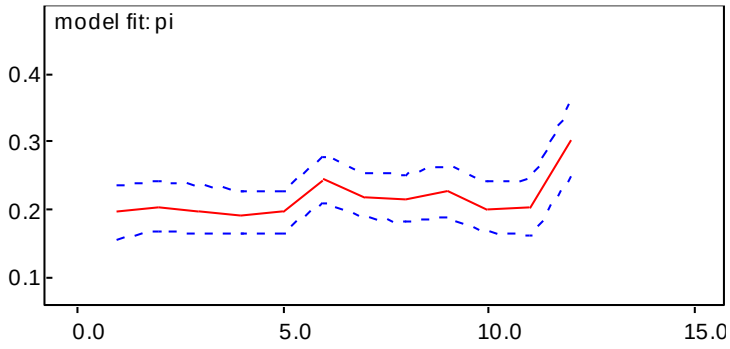


Figure 7: Model fit for parameter π

The deviance information Criterion was used to check for model adequacy. Output from the WINBUG syntax produced the result below

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

Table 3: Deviance Information Criterion (DIC)

	Dbar	Dhat	pD	DIC
y	86.678	77.904	8.774	95.452
total	86.678	77.904	8.774	95.452

The DIC value of 86.678 indicates a good-fitted model. It interprets therefore that the observed data is sufficient to construct the posterior distribution and to evaluate the estimated model

Conclusion

After developing the model, convergence diagnostic checks were conducted for each model parameter in order to ascertain model adequacy. The history plot, density plots and autocorrelation plots were used for this purpose. Though, all these plots were made for each model parameter πand R, sample plots were presented in Figure 2 observe that the history plots shows that the model parameters are well – mixed. This is because they traverse the posterior domain rapidly with nearly constant mean and variance. The model prior distribution for parameters π and R are beta and gamma (0.01, 0.01) respectively

The density plots of these priors reflect this distribution which further validates the model.The density plots of the model parameters were checked against their actual

probability distributions to see whether the right distribution is simulated. This was done for the π and R distribution for each month i. The autocorrelation plots of each parameter and measure depict the independence of the samples generated. This is because the autocorrelations become negligible fairly quickly, after a few lags.

Also Monte Carlo errors (MC error) is lower in comparison to the standard deviation (sd), then we can conclude that the estimated posterior mean was estimated with high precision

Our approach for investigating convergence issues is by inspecting the mixing and time trends within the chains of individual parameters. The history plots are the most accessible convergence diagnostics and are easy to inspect visually. It plots the simulated values for the parameter against the iteration number. The history plot of a well-mixing parameter should traverse the posterior domain rapidly and should have nearly constant mean and variance.

Competing Interest: The authors declare that they have no competing interest

Reference

Amin G.,, Habibollah T.,, Mehdi Y.,, Ahmad R.,, Abbas F., (2020). "A Bayesian Modelling of Malaria in the South of Iran: The Effect of Socio-demographic Factors*", International Journal of Pharmaceutical and Phytopharmacological Research*, 10(**6**), pp.230-236.

Carneiro I., Roca-Feltrer A., Griffin JT., Smith L., Tanner M., Schellenberg A., (2010) Age-patterns of malaria vary with severity, transmission intensity and seasonality in sub-Saharan Africa: a systematic review and pooled analysis. *PLoS ONE*. 2010;**5**: e8988.

Houmsou, C. A., Hay, S.I., Lucioparedes, L. S., Gikandi, P. W., Tatem, A. J., Noor, A. M. and Snow, R. W. (2009). Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malaria Journal* 6 (**1**):17.

National Population Commission, (NPC) (2015). National malaria control programme (NMP) and ICP International, Nigeria Malaria Indicator Survey 2015: Final Report. 2015

Ogbuefi E.O., Aribodor D.N., Joe-Ikechebelu N.N., Okechukwu U.C., Orjichukwu O.N. (2022).Epidemiological Studies of Malaria in Gboko Metropolis, Gboko Local Government Area of Benue State, Nigeria. DOI: https://doi.org/10.21203/rs.3.rs-1405677/v2

Okwa O.O., Akinmolayan F.I., Carter V., Hurd H., (2009) Transmission dynamics of malaria in four selected ecological zones of Nigeria in the rainy season. *Ann Afr Med*. 2009;**8**:1–9. doi:10.4103/1596-3519.55756.

Oyibo W., Ntadom G., Uhomoibhi P., Oresanya O., Ogbulafor N., Ajumobi O., Okoh F., Kolawole M., Ezeiru S., Nwokolo E.,Amajoh C., Ezeigwe N., Audu M., Conway D.(2020) Geographical and temporal variation in reduction of malaria infection among children under 5 years of age throughout Nigeria. *BMJ Global Health* 2021;**6**:e004250. DOI:10.1136/bmjgh-2020-004250

WHO World Malaria Report 2020, 2020. Available: https://www.who.int/ publications/i/ item/9789240015791

World Health Organization (2013). Guideline: Daily Iron and Folic Acid Supplementation in Pregnant Women. Geneva: Accessed July 15, 2020.

World Health Organization (2015).Global Malaria Programme, Avenue Appia CH –1211. Geneva 27. Web: www.who.int/malaria. (Accessed. April 2021).

World Health Organization (2021).World Malaria Report 2021.

World Health Organization.(2012). World Malaria Report. Geneva: World Health Organization; (Accessed. December, 2021).

World Health Organization.(2014). A Strategic Framework for Malaria Prevention and Control during pregnancy in the Africa region. Brazzaville: World malaria report. Geneva: World Health Organization Regional Office for Africa; 16:21. (Accessed January, 2022).

World Health Organization.(2017). Guidelines for the Treatment of Malaria. 4th edition: World Malaria Report. Geneva: World Health Organization; (Accessed. April, 2022).

World Health Organization. World malaria report 2018. 2018. http:// apps.who. int/ iris/ bitst ream/ handle/ 10665/ 275867/ 97892 41565 653-eng.pdf. Accessed 22 May 2021.

World Health Organization. World malaria report 2019. 2019. https:// apps.who.int/ iris/ bitst ream/ handle/ 10665/ 330011/ 97892 41565 721-eng.pdf. Accessed 20 Feb 2022.

World Health Organization.World malaria report 2020: 20 years of global progress and challenges.2020. https:// www.who. int/ docs/default-source/ malar ia/ world-malar ia-reports/ 97892 40015 791-doublepage-view.pdf? sfvrsn= 2c243 49d_5. Accessed 19 Feb 2022.