

## Comparing SARIMA and Dynamic Seasonal Model: Application to Acute Encephalitis Syndrome (AES) for Gorakhpur, India

Pooja Kushwaha\* and Richa Vatsa

Department of Statistics,

Central University of South Bihar-Gaya, Patna (Bihar), India.

(Corresponding author: Pooja Kushwaha\* [poojakushwaha82@gmail.com](mailto:poojakushwaha82@gmail.com))

(Received: 26 January 2023; Revised: 16 February 2023; Accepted: 20 February 2023; Published: 22 March 2023)

(Published by Research Trend)

**ABSTRACT:** In this paper, we applied a Bayesian dynamical seasonal modelling of count data. Their usefulness is illustrated by their application to Acute Encephalitis Syndrome (AES) cases from Gorakhpur regions and by comparing them with the widely used seasonal autoregressive integrated moving average (SARIMA) models for seasonal modelling. The outbreak of encephalitis causes many deaths and long-term disabilities among children and young adults. We considered the AES case data of Gorakhpur from (Jan-12 to Nov-17). We focus on the case of response variables following a Poisson distribution, concentrating on the dynamical seasonal harmonic model. The study helps the policy maker, future disease spread and a better understanding of high-risk months, which may be associated with AES cases. Prior knowledge of the disease outbreak is a main and essential step for policymakers to minimise the disease risk and mortality of children, and enhance health services, vaccination programmes, and other public health initiatives.

**Keywords:** AES, Bayesian estimation, dynamic seasonal harmonic model, cross-validation, SARIMA, WAIC.

### INTRODUCTION

Acute Encephalitis Syndrome (AES), also known as 'Chamki Fever' or Litchi Virus in India, is a unified term used for infections that cause inflammation and cause irritation or swelling in the brain. It represents (Rajnish *et al.*, 2012) as an acute onset of fever and clinical neurological manifestations that include mental confusion, disorientation, delirium, or coma. This syndrome is very complex. It can happen by viruses, bacteria, fungi, and many more factors responsible for this disease. Japanese encephalitis (JE) virus is the most common cause of AES in India, with a union health ministry estimate attributing 5-35 per cent of cases due to JE. However, the syndrome is also caused by scrub typhus, dengue, mumps, measles, and even Nipah or Zika virus, etc. (Julia and Natasha 2007). In several cases, though, the cause of AES remains clinically unidentified. AES occurs every year, from July to November. The outbreak of this disease causes many deaths and long-term disabilities among children and young adults.

In the southern State of Madras, which is now Tamil Nadu, the disease was first reported in India in 1955 (Webb and Sheila 1956). According to the National Vector Borne Diseases Control Programme (NVBDCP), 14,995 AES cases were diagnosed in 2019, with 710 deaths across 23 states NVBDCP (2019). India records a fatality rate of 6 per cent in AES, and the fatality rises to 25 per cent amongst children (NVBDCP, 2017). Assam, Bihar, Jharkhand, Uttar Pradesh, Manipur, Meghalaya, Tamil Nadu, Karnataka, and Tripura are the worst affected states by this disease. The AES is endemic in as many as 171 districts in 19 states (Jai and Shiv 2014). AES disease includes illnesses caused by many

infectious as well as non-infectious causes. Most AES cases are due to viral encephalitis. JE has been considered to be the most important cause of AES in India (Rashmi, 1999), in Asia (Howard and Yuri 2005; Pam 2004). Many studies were executed in the past for AES status and their scenario in India (Rajnish *et al.*, 2012; Nagabhushana, 2012; Sourish and Anirban 2016; Jai *et al.*, 2017). AES and its association with different diseases and possible factors are responsible for the AES cases investigated in several studies. A descriptive case study was carried out to determine the proportion of JE and dengue among AES cases (Sneha and Bellara 2016). In a case-control study, Girish *et al.* (2016) investigated AES cases and associated factors such as socio-demographic and behavioural practices, the presence of pigs, chickens, birds, and other cattle, and the lychee orchard in the vicinity of households. The study summarised that literacy status, the occupational status of parents, travel using public transport modes, and the presence of a lychee orchard near the vicinity of households are factors in the occurrence of AES cases. Other factors identified in studies as responsible for AES cases include heat stroke, pesticides or hypoglycin A (Bandyopadhyay *et al.*, 2015), very hot and humid temperatures, and undernutrition (Sen *et al.*, 2014). A clinical study found that scrub typhus was also a cause of AES (Mahima *et al.*, 2017; Mahima *et al.*, 2018). All these studies are at the district level and survey- or questionnaire-based.

The risk model for encephalitis was also designed at the micro-level, including factors such as villages, land use, pig population distribution, the occurrence of encephalitis cases, and the population under the age of 6 (Nutan and Santanu 2019; Manoj *et al.*, 2018). A Bayesian generalized additive model for AES cases was

used in association with meteorological variables by Praveen *et al.* (2021). Therefore, the literature survey shows that only survey-based studies on AES were reported by Praveen *et al.* (2021). Thus, there are a few studies on the statistical modelling of AES cases.

However, dynamic seasonal modelling of time-series models has been applied in literature and in many fields. The seasonality behaviour of time series data appears very slowly in a non-stationary manner. But with dynamic seasonal models, seasonal behaviour is effectively treated by Christopher (1974). It also ensures that the seasonal component, which is non-stationary, is not confounded with the trend and that the seasonal pattern is predicted into the future over seasons that are consecutive time periods (Andrew and Andrew 1994). Later, a dynamic seasonal model was used in different fields where a seasonal pattern must be taken into account (Andrew, 1997; Giovanni *et al.*, 2009; Serdar *et al.*, 2021).

The AES cases show seasonal behaviour as they appear to be high in the summer season. Moreover, with the analysis of SARIMA models, the observed series must be stationary (the mean and variance do not change with time). In addition, in real-life circumstances, it is impossible to get the data in a stationary form. As a result, some transformation is required, and the original series' information is thus lost.

Dynamic models are a straightforward way to deal with linear and nonlinear data without the need to transform or alter the response variable, accounting for covariates with naturally time-varying behaviour. Thus, in this study, we attempt to apply the Bayesian dynamic seasonal model to model AES cases. The aim of this study is as follows:

- To establish a simple Bayesian dynamic seasonal model for AES cases and compare its model adequacy with the Bayesian seasonal ARIMA model via the WAIC technique.
- To investigate prediction ability of the Bayesian dynamic seasonal model with the help of an appropriate cross-validation technique.

The Bayesian approach offers a basis for projecting future AES with uncertainty measures. The current study shall help the policymakers know the disease spread for the AES cases. Prior knowledge of the disease outbreak is a primary and essential step in minimizing the disease risk and mortality and enhancing health services, vaccination programs, and other public health initiatives. The rest of the paper is structured as follows. Section 2 presents the SARIMA, dynamic Poisson and seasonal harmonic model for modelling AES; it also introduces the particle filtering which is applied for AES estimation in the paper. Data and software used in the paper are also discussed in the same section. Section 3 provides the Bayesian analysis of AES and its results, respectively. Section 4 concludes the paper and discusses its strengths and weaknesses.

## DATA AND MODELS

The Gorakhpur division has seen a seasonal outbreak of AES with high fatalities since 1978 (Mahima and Komal 2014), with 875 cases and 278 deaths reported in 1988

(Rathi *et al.*, 1993). According to the NVBDCP report of 2019, a total of 2,871 cases of AES were reported, out of which 2,389 were in the Gorakhpur region (NVBDCP, 2019). The monthly AES cases are taken for the study from 2012 to 2017 from the NVBDCP, India, reports. The Gorakhpur region is near rivers, lakes, irrigation canals, and rice fields. This flood-prone area is more favourable for AES cases. Therefore, we chose AES cases from Gorakhpur for further analysis in this study. Before an appropriate model is considered for AES data, the stationary behaviour of AES data is checked with a unit-root test. Through the ADF (Augmented Dickey-Fuller) test of unit-root, we found that the p-value is 0.01, which is less than the critical value of 0.05. The null hypothesis is rejected against the alternative hypothesis (stationary time series). Thus, we found that the AES case data is stationary in trend. Further, Fig. 1 presents the time series plot of the AES data.

Fig. 2, which shows the ACF and PCF of AES data, supports the fact that AES case data is seasonal. ACF and PACF both have cyclic behaviour; ACF has significant lag effects at 4,5,6,7,8,11,12,13,17,18, and so on. PACF decays gradually. Even if the data shows stationary behaviour, the seasonal or cyclic behaviour of time-series data is not predictable without transformation, such as in trend non-stationary behaviour analysis (Rob and George 2018). Thus, we may treat AES data as non-stationary, with no trends or seasonal behaviour. Therefore, for the purposes of analysing AES case data, we have tried dynamic seasonal models. As per the literature, there has been little work done on modelling the AES data. We applied seasonal ARIMA models for comparison purposes.

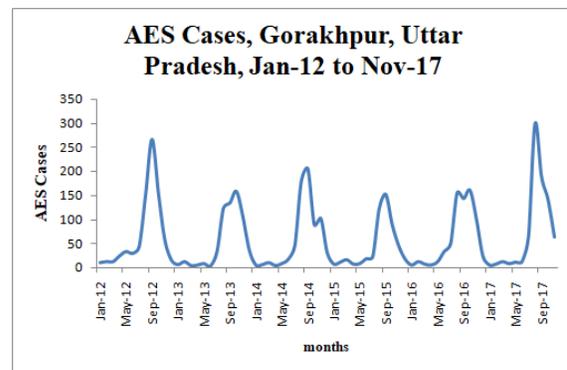


Fig. 1. AES cases data of Gorakhpur, Uttar Pradesh, India during Jan-12 to Nov-17.

**SARIMA Models:** ARIMA models are also used to analyse seasonal data. The ARIMA model was altered to model seasonality by incorporating additional seasonal terms. The seasonal ARIMA or SARIMA model followed by Hyndman *et al.*, (2018) is ARIMA (p,d,q) (P,D,Q) s.

Where, the terms p, d, and q are the non-seasonal order of auto-regressive, difference and moving-average respectively. The terms P, D, and Q are the seasonal order of auto-regressive, difference and moving-average respectively. The term s is the number of seasonal cycles. Given a dependent time series,  $Y_t: 1 < t <$

mathematically the ARIMA seasonal model is written as:

$$(1 - B)^d(1 - B^s)^p Y_t = \mu + \frac{\theta(B)\theta_s(B^s)}{\varphi(B)\varphi_s(B^s)} a_t$$

The Box-Jenkins approach for AR, MA, ARMA, ARIMA and SARIMA models with their identification

of order describe in book (Burnham *et al.*, 2004). The detail with their orders such as AR(1), AR(2), MA(1), MA(2), etc. in univariate and multivariate respect given by William (2006). The time-series analysis with introduction and applications in different fields such as econometric and biology etc (Peter and Richard 2002; Chris, 2003; Rob and George 2018).

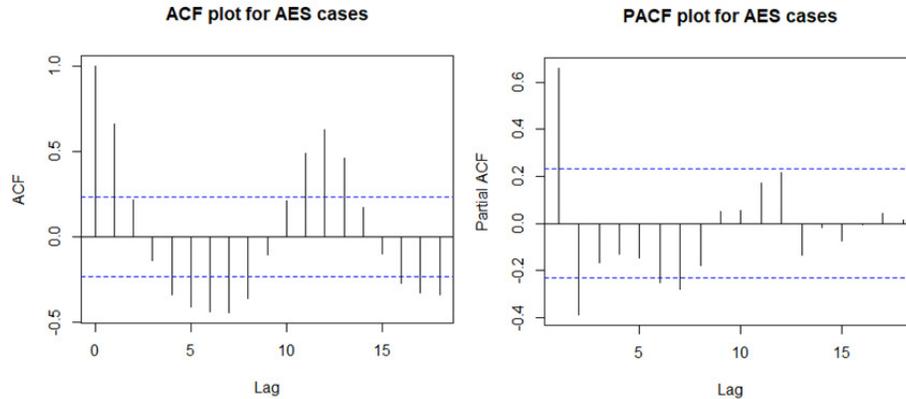


Fig. 2. ACF (left) and PACF (right) plots of AES cases.

**Dynamic Poisson Model:** The Poisson distribution is a very simple model for modelling count data that assumes that a small number of events occur randomly in a given time interval or location. The Poisson distribution may be used to model AES count data. There are several studies that used the Poisson model for modelling count data Mike (2013); Alexander *et al.*, (2016), etc. For the dynamic Poisson model, the natural logarithm is often chosen as the link function; that is,  $g(\lambda_t) = \log(\lambda_t)$ . A dynamic Poisson model in a dynamic form consisting of observation and state equations is as follows:

$$P(y_t | \lambda_t) = \frac{\exp(-\lambda_t) \lambda_t^{y_t}}{y_t!}, \quad (1)$$

$$\log(\lambda_t) = \mu_t + \gamma_t, \quad (2)$$

where, the term  $y_t$  is the observed AES cases. The term  $\lambda_t$  is the mean and variance of the

**Poisson model.** The term  $\gamma_t$  is the seasonal component.

**Dynamic Harmonic Seasonal Models:** There has been different forms of harmonic or trigonometric representation of seasonal component,  $\gamma_t$ . Here for the harmonic seasonal model we follow the from given by Claus and Soren (2006) with  $g(\cdot)$  as a link function:

$$g(E[y_t]) = \mu_t + \gamma_t, \quad (3)$$

$$\mu_{t+1} = \mu_t + \eta_t, \eta_t \sim N(0, Q_{trend,t}) \quad (4)$$

$$\gamma_{t+1} = a_t \cos\left(\frac{2\pi}{s} t\right) + b_t \sin\left(\frac{2\pi}{s} t\right), \quad (5)$$

$$a_t = a_{t-1} + w_{a,t}, w_{a,t} \sim N(0, W_{a,t}) \quad (6)$$

$$b_t = b_{t-1} + w_{b,t}, w_{b,t} \sim N(0, W_{b,t}) \quad (7)$$

Where, the term  $\mu_t$  is the level or trend component of a time series, the terms  $\gamma_t$ , and  $\eta_t$  are seasonal and random components, respectively. The terms  $a_t$  and  $b_t$  are harmonic coefficients. The terms  $w_{a,t}$  and  $w_{b,t}$  are the errors or white-noise components in the process  $a_t$  and  $b_t$  underlying the data, respectively. Also assumed to be Gaussian distribution with mean zero and variances  $W_{a,t}$

and  $W_{b,t}$ , respectively. The term  $s$  is the number of seasons, for monthly time series it is 12, for weekly, 7, etc. AES cases are count data. Hence Poisson distribution may be used to model the AES count data. There were several studies uses Poisson model for count data, Mike (2013); Alexander *et al.* (2016), etc.

**Bayesian Computation Problem:** We know that the computation of posterior distribution and related Bayesian problems are often computationally difficult and technically costly. Hence the MCMC method may be applied for true solutions to Bayesian computation problems in long run.

Under this study, we utilized particle Markov Chain Monte Carlo (PMCMC) methods (Christophe *et al.*, 2010) for inference purpose. A variety of SMC methods currently exist, including the bootstrap filter (Neil *et al.*, 1993), auxiliary particle filter (Michael and Neil 1999), Liu and West filter (Jane and Mike 2001); Storvik filter (Geir, 2002), particle learning algorithm (Carlos *et al.*, 2010) and others. In addition, algorithms such as Particle MCMC (PMCMC) have been developed that place SMC methods within a broader MCMC framework. Particle filters may be known as an alternative of the Kalman filter when analytic solution is not possible for generalized linear models and for unknown parameters such as error variances are unknown or they are time-variant. Christophe *et al.* (2010) gave the three types of PMCMC algorithm, namely, particle independent Metropolis-Hastings (PIMH), particle marginal Metropolis-Hastings (PMMH) and particle Gibbs sampler. For the present study we used PMMH algorithm. The PMMH algorithm is an MCMC algorithm that provides a good approximation in state estimation problems. It uses the marginal Metropolis-Hastings update, along with particle filters to draw samples. More details on SMC and PMMH can be found in Arnaud *et al.* (2000); Sanjeev *et al.* (2002); Arnaud and Adam (2009); Andrieu *et al.* (2010); Murray (2010); Lisa and Chris (2013), etc.

But the execution of MCMC or particle MCMC is time-consuming. To the letter, quick software and packages that deal with MCMC simulations are available. The NIMBLE interface with R (R Core Team 2018) is one such piece of software. It is free to download from the internet at [www.r-nimble.org](http://www.r-nimble.org). The programming language of the software is quite easy to handle, and a direct arrangement of many Bayesian models is thinkable. The advantage of using NIMBLE (Nimble, 2019) is that it converts BUGS code into the model object; after that, we may use any algorithms of our interest, such as the bootstrap filter, particle filtering, Monte Carlo expectation maximisation, and some other tools. In this study, an application of the NIMBLE 0.9.0 version is used to carry out particle MCMC simulation for analysis purposes.

In the present study, model selection is made using WAIC (Sumio and Manfred 2010). The advantage of WAIC is that it can be calculated easily without having information on the true distribution. The selection rule for choosing the best-performing model is that the lower the value of the criterion, the better the fit of the model. The Widely Applicable Information Criterion (WAIC), as given by Andrew *et al.* (2014), is as follows:

$$WAIC = -2(lppd - pWAIC)$$

Where, the term *lppd* is log point-wise predictive density or log-likelihood. The term *pWAIC* is the effective number of parameters.

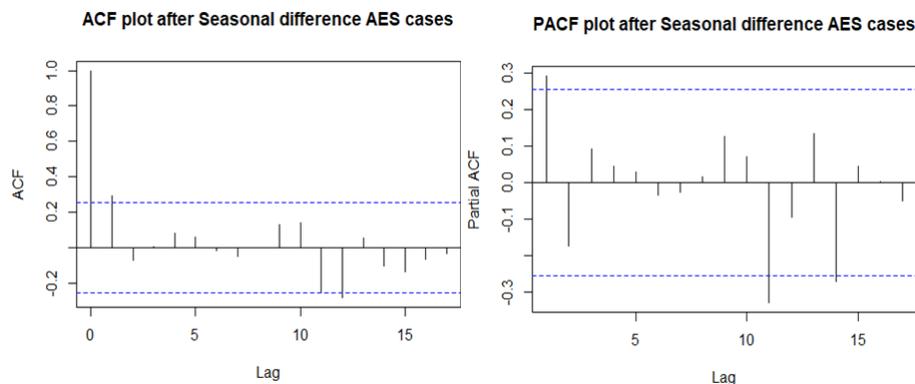
For checking the predictive ability of above models a cross validation technique is applied. The usual split/train test cross validation (cv) technique, because of its randomness, is not suitable for time series data models. For time series data, their chronological order or temporal structure affects the analysis. To overcome all these issues, a new cross validation technique, namely, 'evaluation on rolling forecast origin' is proposed by Rob

and George (2018). As per the methodology described in Rob and George (2018), initially, the training data size is considered 50, with 21 sets of training data preceding each observation (51, 52), up to 70. The one-step-ahead forecast is calculated based on this data. For the accuracy measurements of a predicted model with the CV technique, the mean absolute percentage error is calculated here used by Pooja and Richa (2022). For the convergence of chains, we used Gelmen. diagnostic.

Following section presents analysis and results obtained from Bayesian time series SARIMA and harmonic-seasonal models with the dynamic Poisson model.

## DATA ANALYSIS AND RESULT

We use AES case data from Gorakhpur, India from Jan-2012 to Nov-2017 as observed data,  $y_t$ . The Bayesian SARIMA time-series modelling for AES cases is performed with R via the bayes forecast package (Asael *et al.*, 2021), version 1.0.1. We found that there is no trend in the data, so the non-seasonal difference is taken to be 0. For the non-seasonal order part, the seasonal difference is taken from the previous season. The ACF and PACF plots are shown in Fig. 3 after adjusting for seasonal differences. The PACF plot clearly shows the spike at lag 1 and lag 11, so the AR(1) term may be included in the non-seasonal term. Thus, we choose ARIMA (1, 0, 0) (0, 1, 0) for modelling AES time series data. Also, the ACF plot shows a seasonal spike at lag 12 with a negative cluster, and the PACF plot has spikes at lags 12 and 14, which shows some seasonal MA(1) may be used for the model. Therefore, another model that we choose is ARIMA (1,0,0) (0,1,1) for the AES cases. After discarding 500 initial iterations of the two chains and in-built prior values N(0,0.5) for AR and MA nonseasonal and seasonal coefficients, 30000 samples were considered for the time series analysis.



**Fig. 3.** ACF(left) and PACF(right) plots after seasonal difference of AES cases.

Fig. 4 presents the AES cases as a trend, seasonal, remainder, or irregular components.

We can easily see all three time-series components present in the Fig. 4. The trend component changes at the local level, but the overall trend looks constant.

For the analysis the terms  $\eta_t$ ,  $w_{a,t}$  and  $w_{b,t}$  were the error variables in the state equation, following normal density with mean zero and unknown covariance matrices

$Q_{trend,t} = \sigma_{trend}$ ,  $W_{a,t} = \sigma_a$  and,  $W_{b,t} = \sigma_b$  respectively for each t.

We considered the random walk model as the state equation for AES cases. The terms  $\mu_t$  and  $\gamma_t$  refer to the local level or trend, as well as the seasonal component of AES cases. Since the error variances of the model are unknown, we consider particle filtering for estimation purposes. For simplicity of calculation, we consider precision parameters as the inverses of variance

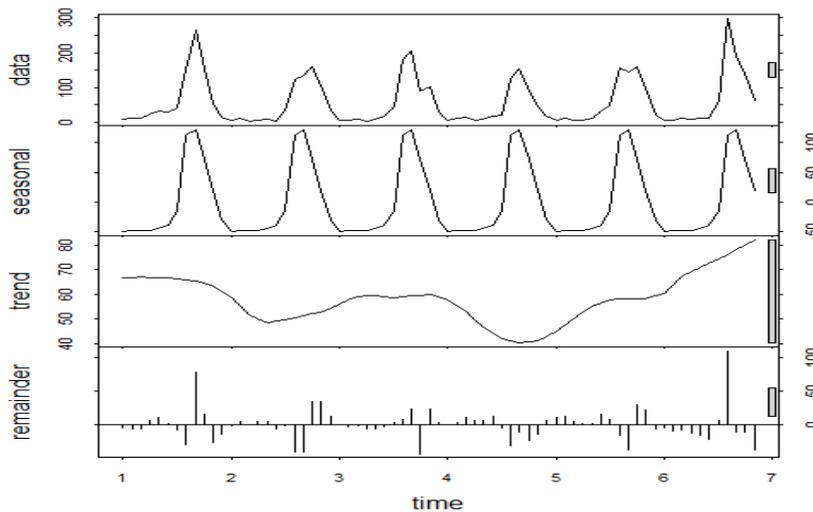
parameters in the analysis, with assumed prior densities for these parameters as gamma densities with hyperparameters of 0.01 and 0.01, respectively. The coefficients 'a' and 'b' of the harmonic model for state variables were chosen to have a Gaussian distribution (0, 0.01). We considered 600,000 samples and discarded 10,000 observations as a burn-in period. Furthermore, 15 particles were used to ensure MCMC chain convergence; two chains with the thin value 15 are considered.

**Results:** The WAIC criterion is used for checking the best fitted model. The Bayesian time series SARIMA models and dynamic harmonic models WAIC values are shown in Table 1. For both Bayesian SARIMA(1,0,0)(0,1,0)[12] and SARIMA(1,0,0)(0,1,1)[12] models WAIC are 597.2 and 591.6, respectively. The seasonal harmonic Poisson model is the least WAIC value 490.38.

Table 2 shows the posterior estimates of parameters  $\eta$ ,  $\sigma_a$  and  $\sigma_b$  respectively for the seasonal harmonic Poisson model. It can be seen that the posterior means of  $\sigma$ ,  $\sigma_a$  and  $\sigma_b$  are 0.03, 0.32, and 0.04, respectively, which is desirable for a good prediction. But the 95% credible intervals are wide bound. As a result, prediction uncertainty may be high.

**Table 1: WAIC values of Bayesian time-series ARIMA, dynamic Poisson seasonal harmonic models.**

Models	WAIC values
SARIMA(1,0,0)(0,1,0)[12]	597.2
SARIMA(1,0,0)(0,1,1)[12]	591.6
Dynamic Poisson	490.38



**Fig. 4.** AES cases decompose into time-series components.

**Table 2: Posterior estimates of parameters for dynamic Poisson seasonal harmonic model.**

Variables					
	Mean	Median	St. Dev.	95%CI-lower	95%CI-upper
$\sigma = 1/\eta$	0.0300	0.0101	0.0612	0.0026	0.247
$\sigma_a = 1/\phi.a$	0.3284	0.3497	0.1761	0.0064	0.6476
$\sigma_b = 1/\phi.b$	0.0404	0.0076	0.0971	0.0020	0.3724
$\eta$	122.7869	98.9431	101.3146	4.0494	385.7594
$\phi.a$	17.2771	2.8596	46.7097	1.5442	156.7568
$\phi.b$	160.1579	131.9295	132.0945	2.6850	490.4584

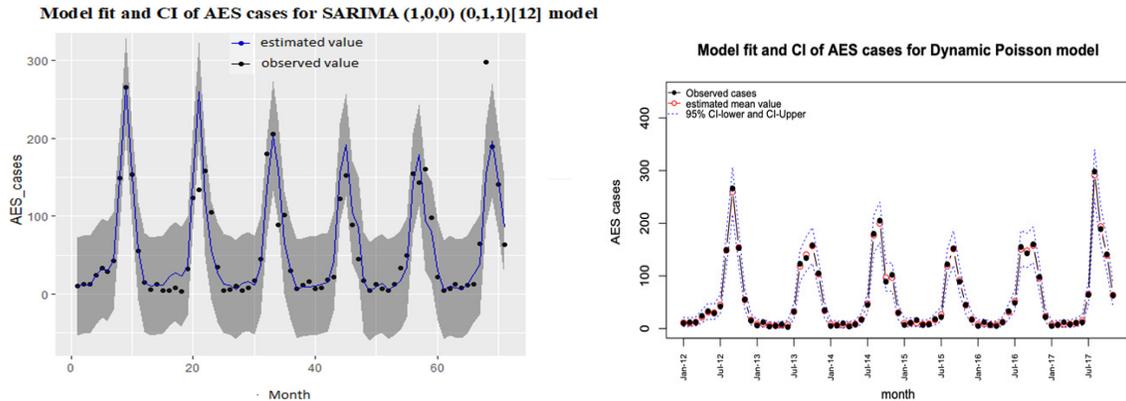
Fig. 5 shows the estimation and observed values of AES cases for the years Jan. 2012–Nov. 2017 for SARIMA (1, 0, 0) (0, 1, 1) [12] and the dynamic Poisson seasonal harmonic model. The Poisson model fits the AES cases better than SARIMA (1, 0, 0) (0, 1, 1) [12]. For SARIMA (1, 0, 0) (0, 1, 1) [12] 95% values in the estimated CI, the observed values within the CI interval are calculated for the estimation period. Whereas, for the dynamic Poisson model, all (100%) values are in the estimated CI. The prediction accuracy of the model is explored via the CV technique (rolling forecast origin). Fig. 6 shows the comparison of observed and predicted AES cases for SARIMA (1, 0, 0) and (0, 1, 1) [12] dynamic Poisson

models. The dynamic Poisson model's CI variability is high, but it contains all observed values. The observed values are within the CI interval for SARIMA (1, 0, 0) (0, 1, 1) [12], with a predictive CI of 66%. The mean absolute percentage errors (MAPE) are calculated based on the particle MCMC simulations. For the dynamic Poisson model, MAPE is 63%, and for SARIMA (1, 0, 0) (0, 1, 1) [12], it is approximately 111%. This shows that the dynamic Poisson model is better for prediction in comparison to SARIMA.

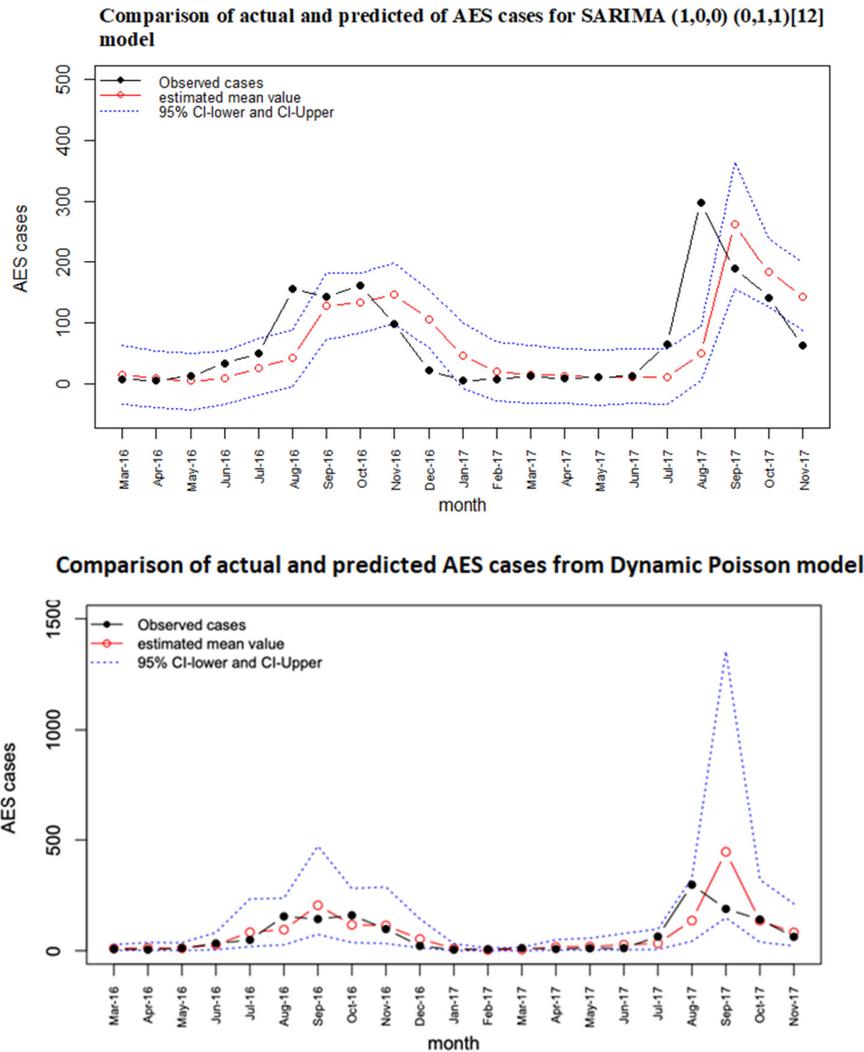
The posterior estimates of trend plots for AES cases are presented in Fig. 7. The trend plot shows that from January 2012 to September 2012, the trend component

was greater than or equal to 3.5. For the year Non-2012 to May-2013, the trend is greater than 3.5. After that, it is approximately equal to 3.5. Fig. 8 presents the posterior estimates of the seasonal component plot. The seasonal component is high in September. As a result, there is a high likelihood of AES cases being reported in

September and in the future. Also, the seasonal component is negatively associated with AES cases from January to June. Hence, AES cases might occur in these months with no seasonal effect. From July to November, the seasonal component is positive. Thus, these seasons may be positively associated with the disease.



**Fig. 5.** Estimation and fitting of observed AES cases for SARIMA(1,0,0)(0,1,1)[12] (top) and dynamic Poisson seasonal harmonic (bottom) model.



**Fig. 6.** Comparison of predicted and observed AES cases for SARIMA(1,0,0)(0,1,1)[12] (top) dynamic Poisson seasonal harmonic (bottom) model.

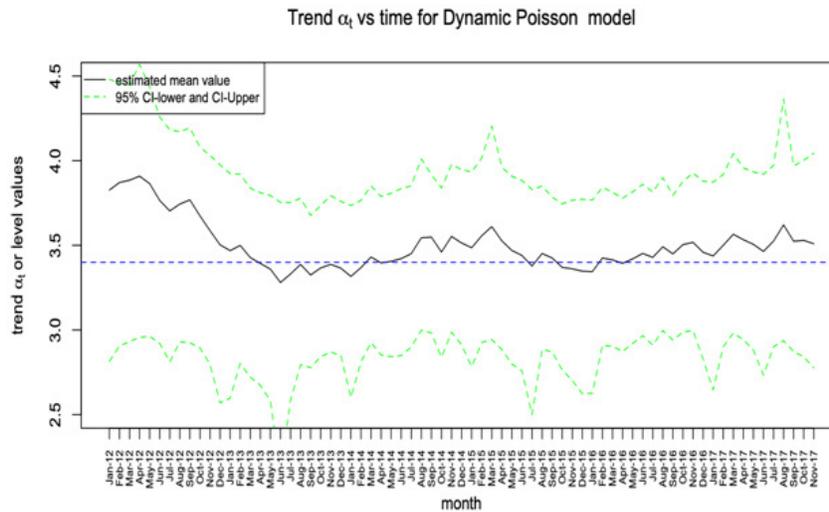


Fig. 7. Trend plot of AES cases for dynamic Poisson seasonal harmonic model.

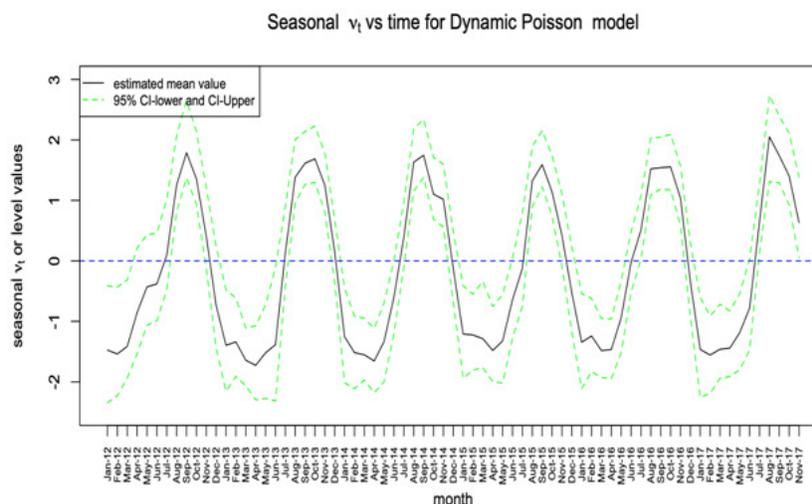


Fig. 8. Seasonal plot of AES cases for dynamic Poisson seasonal harmonic model.

## DISCUSSION

We have applied the dynamic seasonal Poisson model for AES cases. For comparison purposes, we used SARIMA models. The best-fit model is chosen based on the WAIC value. The dynamic model is found to be the best, with a lower WAIC than the Bayesian SARIMA models. The dynamic Poisson model is found to be superior to SARIMA models in the estimation and prediction of observed AES cases. With the help of dynamic seasonal models, the AES case data revealed some trends that the SARIMA models were unable to capture. We find that the predictive ability of the MAPE dynamic Poisson model is better than SARIMA. But it is high for prediction accuracy in general, which shows that covariates may be needed for better accuracy in prediction.

Serdar *et al.* (2021) compared the different models under the state space model (SSM) via the Kalman filter: a hybrid model integrating the logistic regression and SSM models; the seasonal autoregressive integrated moving average (SARIMA); exponential smoothing with the state space model (ETS); and exponential smoothing the

state space model with the Box-Cox transformation (ARMA errors, trend, and seasonal components) (TBATS). They discovered that the SSM model outperformed the SARIMA model. In AES cases, our study found similar results.

Rajnish *et al.* (2012); Nagabhushana (2012); Sourish and Anirban (2016), and others investigated AES cases using questionnaires and surveys. Statistical modelling may be more useful to gain a better understanding of the data. Praveen *et al.* (2021) applied GAM models for AES cases in Gorakhpur with predictors as meteorological variables. They explored the monthly relationship between AES cases with the help of log transformation. So, all results are not based on the original AES cases. Because of the transformation, the observed series is not the original series any longer. In theory, a series can be differenced an arbitrary number of times. But recovering the original series requires the inverse operation, which may not be possible after analysing the series Joseph (1993). In our study without transformation, AES cases are analysed in a dynamic framework, and we find that most AES incidences occur from June to September.

## CONCLUSIONS

In our analysis, we found a clear seasonal peak in September. August and October months are also positively associated with the disease. The dynamic seasonal harmonic Poisson model modelled non-stationarity in seasonal data better than SARIMA models. However, the limit of our study is that it is based only on trends and seasonal effects, with no other explanatory variables included in the study. But this may be seen as an opportunity. In all circumstances, one wants to know the disease pattern and its association with different months. The dynamic seasonal model may provide a better understanding than the SARIMA models.

## FUTURE SCOPE

Some explanatory variables, e.g. literacy status, occupational status of parents, very hot and humid, temperature and under-nutrition, etc. may be included into the dynamic model for better accuracy in prediction of AES data.

**Author contributions.** Pooja Kushwaha contributed in analysis code, performed the analysis and wrote the paper Richa Vatsa conceived and designed the analysis.

**Acknowledgement.** The authors are thankful to National Vector Borne Disease Control Program for these type of data collection and available for the study.

**Conflict of interest.** None.

## REFERENCES

- Alexander, K. M., Oscar, O. N. and Anthony, G. W. (2016). Statistical models for count data. *Science Journal of Applied Mathematics and Statistic*, 4(6), 256-262.
- Andrew, G., Johan, C. C., Hal, S. S., David, B. D., Aki, V. and Donald B. R. (2014). Bayesian inference. Bayesian Data Analysis. 3<sup>rd</sup> ed. *Boca Raton: CRC*, 6–7.
- Andrew, H. (1997). Trends, cycles and autoregressions. *The Economic Journal*, 107(440), 192–201.
- Andrew, H. and Andrew, S. (1994). Seasonality in dynamic regression models. *The Economic Journal*, 104(427), 1324–1345.
- Arnaud, D. and Adam, J. M. (2009). A Tutorial on Particle Filtering and Smoothing: Fifteen years later, *Handbook of nonlinear filtering*, 12(656-704), 3.
- Arnuad, D., Simon, G. S. and Christophe, A. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3), 197–208.
- Asael, M. A., Cristian, C. T., Andres, D., Rob, H. and Mitchel, O. W. (2021). bayesforecast: Bayesian Time Series Modeling with Stan. *ARXIV Preprint*.
- Bandyopadhyay, B., Chakraborty, D., Ghosh, S., Mishra, R., Rahman, M., Bhattacharya, N., Alam, S., Mandal, A., Das, A., Mishra, A., Mishra, A. K., Arvind, K., Haldar, S., Pathak, T., Mahapatra, N., Mondal, D., Maji, D. and Basu, N. (2015). Epidemiological investigation of an outbreak of acute encephalitis syndrome (AES) in Malda district of West Bengal, India. *Clinical Microbiology*, 4(1), 181–187.
- Burnham, P. K. and Anderson, D. R. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63(2020), 10.
- Carlos, M. C., Michael, S. J., Hedibert, F. L. and Nicholas, G. P. (2010). Particle learning and smoothing. *Statistical Science*, 25(1), 88–106.
- Chris, C. (2003). The analysis of time series: an introduction. *Chapman and hall/CRC*.
- Christophe, A., Arnaud, D. and Roman, H. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3), 269–342.
- Christopher, A. S. (1974). Seasonality in regression. *Journal of the American Statistical Association*, 69(347), 618–626.
- Claus, D. and Soren L. C. (2006). Formulating state space models in R with focus on longitudinal regression models. *Journal of Statistical Software*, 16, 1–15.
- Geir, S. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on signal Processing*, 50(2), 281–289.
- Giovanni, P., Sonia, P. and Patrizia, C. (2009). Dynamic linear models. *Springer*.
- Girish, K. S., Chandra, M. S., Alok, R., Neeraj, A., Sanjay, P., Pragya, K. and Ghanshyam, S. (2016). Determinants of Acute Encephalitis Syndrome (AES) in Muzaffarpur district of Bihar, India: A case-control study. *Clinical Epidemiology and Global Health*, 4(4), 181–187.
- Howard, E. G. and Yuri, P. (2005). Infections of the nervous system. *The Lancet Neurology*, 4(1), 12–13.
- Jai, P. N. and Shiv, L. (2014). Responding to the challenge of acute encephalitis syndrome/JE in India. *Journal Communication Disease*, 46, 1–3.
- Jai, P. N., Dhariwal, A. C. and Raina, C. M. (2017). Acute encephalitis in India: An unfolding tragedy. *The Indian journal of medical research*, 145(5), 584.
- Jane, L. and Mike, W. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, 197–223. Springer.
- Joseph, L. H. (1993). An introduction to modeling dynamic behaviour with time series analysis. In *Performance Evaluation of Computer and Communication Systems*. Springer, 203–223.
- Julia, G. and Natasha, S. C. (2007). The epidemiology of acute encephalitis. *Neuropsychological Rehabilitation*, 17(4-5), 406–428.
- Lisa, T. and Chris, S. (2013). An introduction to particle filtering. *Lancaster University Lancaster*.
- Mahima, M. and Komal, P. K. (2014). AES: Clinical presentation and dilemmas in critical care management. *The Journal of communicable diseases*, 46(1), 50–65.
- Mahima, M., Jeromie, W. V. T., Winsely, R., Valsan, P. V., Girish, C. P. K., Sabarinathan, R., Vijay, B., Nivedita, G. and Manoj, V. M. (2017). Scrub typhus as a cause of acute encephalitis syndrome, Gorakhpur, Uttar Pradesh, India. *Emerging infectious diseases*, 23(8), 1414.
- Mahima, M., Vijay, B., Manoj, M., Hirawati, D., Winsley, R., Valsan, V. P., Mahim, M., Patil, G., Ramsamy, S., Jeromie, T. V., Kaliaperumal, K., John Antony, J. P., Gupta, N., Gupte, M. M. and Gupte, M. D. (2018). Acute encephalitis syndrome in Gorakhpur, Uttar Pradesh, 2016: clinical and laboratory findings. *The Pediatric infectious disease journal*, 37(11), 1101–1106.
- Manoj, M., Jeromie, W. V. T., Mahima, M. and Nivedita, G. (2018). Acute encephalitis syndrome in Eastern Uttar Pradesh, India: changing etiological understanding. *Journal of medical entomology*, 55(3), 523–526.
- Michael, P. K. and Neil, S. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446), 590–599.
- Mike, W. (2013). Bayesian dynamic modelling. Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian FM Smith, 145, 166.

- Murray, P. (2010). Algorithms & Computationally Intensive Inference Reading Group Introduction to Particle Filtering Discussion-Notes. *Journal of the American Statistical Association*, 1–31.
- Nagabhushana, R. P. (2012). Incidence rate of acute encephalitis syndrome without specific treatment in India and Nepal. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, 37(4), 240.
- Neil, G. J., David, J. S. and Adrian, F. S. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE proceedings F (radar and signal processing)*, 2(140), 107-113. IET Digital Library.
- Nimble, D. T. (2019). NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling. *R package version*, (9).
- Nutan, T. and Santanu, S. (2019). Geospatial disease risk modeling for the identification of potential areas of encephalitis in a subtropical region of India: a micro-level case study of Gorakhpur tehsil. *Applied Geomatics*, 1–15.
- NVBDCP (2017). National Vector Borne Disease Control Programme.
- NVBDCP (2019). National Vector Borne Disease Control Programme.
- Pam, D. (2004). Infectious disease surveillance update. *The Lancet Infectious Diseases*, 4(1), 7.
- Peter, J. B. and Richard A. D. (2002). Introduction to time series and forecasting. *Springer*.
- Pooja, K. and Richa, V. (2022). Modeling and predicting total fertility rate of India with Bayesian dynamic linear modelling. *Stochastic Modeling & Applications*, 26(3).
- Praveen, K., Pradhan, P. S. and Bharat, B. (2021). Meteorological association for prevalence dynamics of Acute Encephalitis syndrome: a modeling study. *Modeling Earth Systems and Environment*, 1–11.
- R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna Austria.
- Rajnish, J., Kalantri, S. P., Arthur, R. and Jhon, M. C. (2012). Changing landscape of acute encephalitis syndrome in India: a systematic review. *National Medical Journal of India*, 25(4), 212-220.
- Rashmi, K. (1999). Viral encephalitis of public health significance in India: current status. *The Indian Journal of Pediatrics*, 66(1), 73–83.
- Rathi, A. K., Kushwaha, K. P., Singh, Y. D., Singh, J., Sirohi, R., Singh, R. K. and Singh, U. K. (1993). JE virus encephalitis: 1988 epidemic at Gorakhpur. *Indian pediatrics*, 30(3), 325–325.
- Rob, J. H. and George, A. (2018). Forecasting: principles and practice. *OTexts*.
- Sanjeev, M. A., Simon M., Neil, G. and Tim, C. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, 50(2), 174–188.
- Sen, P. K., Dhariwal, A. C., Jaiswal, R. K., Lal, S., Raina, V. K. and Rastogi, A. (2014). Epidemiology of acute encephalitis syndrome in India: changing paradigm and implication for control. *The Journal of communicable diseases*, 46(1), 4–11.
- Serdar, N., Ecem, Ü. and Ceylan, Y. (2021). Performance comparison of filtering methods on modelling and forecasting the total precipitation amount: a case study for Muğlaın Turkey. *Journal of Water and Climate Change*, 12(4), 1071–1085.
- Sneha, K. and Bellara, R. (2016). A clinico-epidemiological profile of acute encephalitis syndrome in children of Bellary, Karnataka, India. *International Journal of Community Medicine and Public Health*, 3(11), 2997–3002.
- Sourish, G. and Anirban, B. (2016). Acute encephalitis syndrome in India: the changing scenario. *Annals of neurosciences*, 23(3), 131.
- Sumio, W. and Manfred, O. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 3571–3594.
- Webb, J. K. G., and Sheila, P. (1956). Clinical Diagnosis of an Arthropod Borne Type of Virus Encephalitis in Children of North Arcot District, Madras State, India. *Indian Journal of Medical Sciences*, 10(8), 573–581.
- William, W.W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods (Second edition ed.). *Pearson Addison Wesley*.

**How to cite this article:** Pooja Kushwaha and Richa Vatsa (2023). Comparing SARIMA and Dynamic Seasonal Model: Application to Acute Encephalitis Syndrome (AES) for Gorakhpur, India. *Biological Forum – An International Journal*, 15(3): 534-542.